AJSLP

Research Article

Cohesive Referencing Errors During Narrative Production as Clinical Evidence of Central Nervous System Abnormality in School-Aged Children With Fetal Alcohol Spectrum Disorders

John C. Thorne[a] and Truman E. Coggins[a]

Purpose: Previous evidence suggests that cohesive referencing errors made during narratives may be a behavior that is revealing of underlying central nervous system abnormality in children with fetal alcohol spectrum disorders (FASD). The current research extends this evidence.
Method: Retrospective analysis of narrative and clinical data from 152 children (ages 6 to 14), 72 of whom had confirmed FASD, was used. Narrative analysis was conducted blind to diagnostic status, age, or gender. Group performance was compared. The associations between measures of cohesive referencing and clinically gathered indices of the degree of central nervous system abnormality were examined.
Results: Results show clear associations between elevated rates of cohesive referencing errors and central nervous system abnormality. Elevated error rates were more common in children with FASD than those without, and prevalence increased predictably across groups with more severe central nervous system abnormality. Risk is particularly elevated for those with microcephaly or a diagnosis of fetal alcohol syndrome.
Conclusion: Cohesive referencing errors during narrative are a viable behavioral marker of the kinds of central nervous system abnormality associated with prenatal alcohol exposure, having significant potential to become a valuable diagnostic and research tool.

In the field of speech-language pathology, sampling and analysis of extended discourse has long been touted as a clinical tool with the potential to provide "important keys to understanding the communication needs of older language-disordered children" because "larger linguistic units have their own structured rules and guiding principles" (Johnston, 1982, p. 144). Because successful production of extended discourse requires the integration of a large number of linguistic and other cognitive skills, it provides the opportunity to explore the nature of and integrity of these cognitive systems. This gives discourse-level tasks the potential for revealing not only language impairment (Gillam & Pearson, 2004; Heilman, Miller, Nockerts, & Dunaway, 2010; Schneider & Hayward, 2010) but also impairment of the central nervous system (CNS) more generally.

The ability to leverage language-based tasks to explore the integrity of neurobehavioral systems makes speech-language pathologists important members of interdisciplinary teams involved in making diagnostic decisions for a range of neurodevelopmental and neurocognitive disorders. The development of valid discourse-level tasks for this purpose has the potential to increase the range and type of information that speech-language pathologists can bring to the table in these contexts. The challenge, of course, is to identify which aspects of an extended discourse task are vulnerable to which kinds of CNS impairment from the myriad choices available in an analysis of discourse behavior. Finding the answer to this question has important clinical and research implications—making it worth the effort. For clinicians, knowing which aspects of discourse are most vulnerable to underlying CNS abnormality can inform clinical decision making and improve the efficiency of diagnostic assessment practices; for researchers,

[a]University of Washington, Seattle
Correspondence to John C. Thorne: jct6@uw.edu
Editor: Krista Wilkinson
Associate Editor: Amanda Van Horne
Received August 26, 2015
Revision received February 1, 2016
Accepted March 7, 2016
DOI: 10.1044/2016_AJSLP-15-0124

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

this same information has important scientific value, potentially facilitating understanding of the complex connections between various underlying CNS impairments and observable communicative behaviors.

Our recent research has focused on cohesive referencing behaviors in the narrative discourse of school-aged children with CNS abnormality associated with prenatal alcohol exposure (i.e., children with fetal alcohol spectrum disorders [FASD]). It has been estimated that FASD affects at least 1% and perhaps as many as 4% or more of children (see e.g., May et al., 2014; Sampson et al., 1997), with prevalence estimates depending heavily on geography and quality of diagnosis (Roozen et al., 2016). Even using the lower prevalence estimates, this makes individuals with FASD a population that will be encountered by most clinicians.

The impact of prenatal alcohol exposure on any individual depends on the timing and dosage of teratogenic exposure (Astley, 2013). As a result, the CNS impairment that results from prenatal alcohol exposure is highly variable, and a wide range of diagnoses are associated with that exposure including attention-deficit/hyperactivity disorder, language disorder, intellectual disability, sensory processing difficulties, memory impairment, and motor impairment (Astley, 2010b; Popova et al., 2016). Because these associated diagnoses are not specific to prenatal alcohol exposure, this population is well suited to serve as a proxy for the population with CNS abnormality more generally, with the advantage that these individuals have a confirmed risk factor that may explain in whole or in part why they have CNS abnormalities.

By investigating a narrow set of discourse-level behaviors in this heterogeneous population with verifiable CNS abnormalities, we have been able to provide preliminary evidence that cohesive referencing errors provide a useful and reliable clinical signal of underlying CNS abnormality, a signal that is identifiable even when children fail to screen positive for language disorder in clinical assessment (Thorne & Coggins, 2008a, 2008b; see also Coggins, Timler, & Olswang, 2007). It is important that this signal is identifiable in this population despite the wide variety of CNS abnormalities present across FASD. This suggests that this aspect of narrative discourse may be particularly sensitive to the types of CNS impairment caused by prenatal alcohol exposure, increasing the utility of this particular behavioral signal for use with this important clinical population. It is possible, of course, that this behavioral signal may instead be sensitive to the presence of CNS abnormalities more broadly, suggesting that it may have potential utility for use with other clinical populations with CNS abnormalities in the absence of prenatal alcohol exposure. If that is the case, confirming its utility for use with suspected FASD is a logical first step in validating it for broader use because there is no guarantee that its utility will transfer to populations without prenatal alcohol exposure.

The current research extends our previous work through retrospective analysis of two large existing samples of children: a sample of children with prenatal alcohol exposure and previously identified CNS abnormality and a control sample of children without concerns related to neurocognitive development. To set the stage for this study, we will first provide a succinct description of FASD, followed by a discussion of the methodology used for quantifying the spectrum of disabilities present among children with prenatal alcohol exposure. We conclude with a summary of the research leading up to the current work.

Fetal Alcohol Spectrum Disorders

Fetal alcohol syndrome (FAS) is the most readily recognized phenotype of the FASD. FAS is a permanent birth defect syndrome resulting from prenatal alcohol exposure that is characterized by growth deficiency, a unique cluster of three minor facial anomalies (i.e., flat philtrum, thin upper lip, and small eyes), and evidence of CNS abnormalities (see e.g., Astley, 2004; Bertrand, Floyd, & Weber, 2005). Disorders on the fetal alcohol spectrum that lack the facial features of FAS are more prevalent than FAS but share a similar range and severity of impairments (Astley, 2004; Sampson, Streissguth, Bookstein, & Barr, 2000; Streissguth et al., 2004). As already mentioned, CNS abnormalities that have been associated with FASD occur across a wide range of functional domains and brain structures (Astley, 2013; Donald et al., 2015). It is the presence of any CNS abnormality across this range in the context of confirmed prenatal alcohol exposure that defines the fetal alcohol spectrum and determines the specific FASD diagnosis that applies to an individual.

Characterizing the range of impairment associated with FASD has proved to be a challenging enterprise, and a number of diagnostic approaches have emerged (e.g., Bertrand et al., 2005; Chudley et al., 2005; Cook et al., 2016; Hoyme et al., 2005). The challenges come primarily from the heterogeneous nature of the physical and neurological impairments associated with prenatal alcohol exposure. Whereas the facial features associated with FAS are highly specific to alcohol exposure (Astley, 2013), most children with prenatal alcohol exposure will not manifest the distinct cluster of features needed to receive a diagnosis of FAS. In addition, as of yet, no clear profile of CNS impairment has emerged that is both specific to and sensitive to prenatal alcohol exposure, with most research suggesting that the population presents a range of cognitive and behavioral profiles that overlaps with other neurodevelopmental disorders (see e.g., Kodituwakku & Kodituwakku, 2014; also Bakoyiannis et al., 2014; Popova et al., 2016).

The most widely used diagnostic system for FASD, the 4-Digit Code (Astley, 2004), is a well-validated (Astley, 2013) case-defined diagnostic approach that uses Likert rank scales to report outcomes for the full spectrum of disabilities associated with prenatal alcohol exposure. It includes scales for the domains of growth deficiency (i.e., growth rank), facial features associated with FAS (i.e., face rank), CNS abnormality (i.e., CNS rank), and prenatal

alcohol exposure (i.e., prenatal alcohol exposure rank). The scale for each of these domains ranges from a rank of 1, indicating *typical/unremarkable presentation of the feature*, to a rank of 4, indicating a *significant presentation consistent with FAS*.

The CNS rank is a nested ranking with a rank of 4 indicating *direct structural or neurological evidence of significant CNS abnormality* (e.g., microcephaly, seizures), whereas Ranks 1 through 3 indicate *increasingly severe functional impairment*. *Significant functional impairment* is defined in the system as performance equivalent to 2 *SD*s from the mean on a standardized, norm-referenced task. A CNS rank of 3 requires significant functional impairment in at least three domains of neurocognitive functioning. A CNS rank of 3 or 4 is required for an FAS diagnosis and is described as "static encephalopathy," whereas a CNS rank of 2 is described as "neurobehavioral disorder" and indicates mild to moderate CNS abnormality. In the 4-Digit Code, the four domain-specific Likert ranks are combined from left to right in the order of growth, face, CNS, and prenatal alcohol exposure to assign an individual the appropriate FASD diagnosis. The 4-Digit Code renders 22 possible diagnoses, eight of which fall under the umbrella of FASD, with FAS representing the most severe presentation. Having a more severe CNS rank or a more severe FASD diagnosis under the 4-Digit Code are both associated with more severe underlying CNS abnormality (Astley, 2013), making the system ideal for quantifying and comparing the degree of CNS abnormality across this clinical population.

Previous Research on FASD and Cohesive Referencing

Coggins et al. (2007) found a minority of older school-aged children with FASD to have markedly deficient performance on traditional standardized language testing, whereas significant majorities failed to produce age-appropriate narrative discourse during a common clinical task: structured narrative generation on the basis of the wordless picture book, *Frog, Where Are You?* (Mayer, 1969). This finding reinforced the notion that narrative discourse provides a meaningful alternative by which to examine the communicative processes of school-age children with FASD and that narrative performance may indeed be sensitive to the type of CNS abnormalities commonly reported in these children.

To better understand what narrative components might be at the root of these findings, Thorne, Coggins, Carmichael Olson, and Astley (2007) examined the diagnostic utility of a variety of linguistic markers and cohesive elements in the narratives of a sample of 32 school-age children, 16 with FASD. Results indicated that errors in referential cohesion (i.e., failure to use an unambiguous term to refer to or introduce a concept) provided the most diagnostically salient signal of CNS abnormality in this population. In the study, most of the narratives were properly attributed to storytellers either with or without FASD on the sole basis of the rate at which they made cohesive referencing errors using a global metric of this aspect of discourse (81.25% overall accuracy; area under the receiver operating curve [AUC] of .86, 95% CI [0.74, 0.99]). This motivated further research into referential cohesion in this population.

Identifying Referencing Errors

Measuring cohesive referencing can be approached in a variety of ways (see e.g., Heilman et al., 2010; Liles, 1993; Schneider & Hayward, 2010; Strong, 1998). Our approach (Thorne, 2006), *Tallying Reference Errors in Narratives* (TREIN), focuses on a child's ability to establish and manage a common ground of story elements using referential terms.

Building on theoretical work by Halliday and Hasan (1976), Van Hoek (1997), and Ariel (1994, 2004; see also Ariel, 2009), our methodological approach categorizes the nominal phrases (i.e., noun phrase or pronoun) in a narrative into one of nine categories: five categories for use of appropriate/cohesive reference strategies; and four categories for "reference errors" (two for pronouns and two for noun phrases). In the system, a nominal phrase is considered to lack referential cohesion when it fails to unambiguously refer to available referents in the narrative or when it fails to appropriately mark the information status ("new" or "known") of referents when concepts are being introduced into the narrative. See Appendix A for more information related to the TREIN and Table A1 for the specific coding scheme used in the TREIN.

Our previous work (Thorne & Coggins, 2008a, 2008b) found that a specific metric, the rate of nominal reference errors (rNRE), could accurately predict which storytellers in small groups of school-age children were children with FASD and which were not (overall accuracy 88%, AUC of .90, 95% CI [0.73, 0.97]). It is important to note that half of the children with an FASD in the sample performed in the average range on a standardized language test (Wiig & Secord, 1989) that required a sentence-level response; the standardized language test would not have revealed CNS abnormality in these children.

Aims of the Current Study

These promising findings motivated further research with a larger group of children. The current research addressed two primary research aims. Our first research aim was to corroborate findings from previous work with the TREIN (Thorne & Coggins, 2008a, 2008b) by examining the cohesive referencing of children with CNS abnormality associated with prenatal alcohol exposure (i.e., FASD) and comparing them with children without concerns related to neurocognitive development. Two specific hypotheses were tested to further this aim.

Hypothesis 1: Children with CNS abnormality associated with FASD will exhibit more cohesive reference errors than children without concerns related to neurocognitive development.

Hypothesis 2: The proportion of children in the FASD group who generate a narrative with a rNRE in the clinically impaired range will be equivalent to or greater than 50%.

All children with FASD, by definition, have CNS abnormality to one degree or another. If any of these CNS abnormalities are associated with reduced capacities for cohesive referencing, one would expect that the FASD group would make more cohesive referencing errors on average than their peers without impairment. Our previous research suggested this, and we found the largest difference in nominal reference errors (NRE). On the basis of these previous findings, we predicted a proportion of children with FASD equivalent to or greater than 50% would have an rNRE that was 2 *SD*s above the mean of their peers without identified CNS impairment.

Our second research aim was to extend our previous research by examining the association between elevated rNRE and the severity of CNS abnormality in children with FASD. Two hypotheses were tested to examine this association.

*Hypothesis 3: As severity of 4-Digit Code CNS rank increases from Rank 1 (*no concern for CNS abnormality*) to Rank 4 (*definite CNS abnormality*), the proportion of children having an elevated rNRE will increase for each rank.*

Hypothesis 4: As severity of FASD diagnosis increases from no-FASD to FAS, the proportion of children exhibiting significantly impaired cohesive referencing on the basis of rNRE will increase.

The 4-Digit Code is designed to capture increasing manifestation of impairment associated with prenatal alcohol exposure. If cohesive referencing provides a salient signal of underlying CNS abnormality associated with prenatal alcohol exposure, we would predict that increasingly severe impacts from prenatal alcohol exposure would come with increased risk of significantly impaired cohesive referencing. We therefore examined the association between significantly elevated rNRE and two specific proxies for severity of CNS impairment derived from the 4-Digit Code: (a) CNS rank, which reflects degree of CNS impairment on the basis of an interdisciplinary clinical assessment of CNS structure and function; and (b) severity of FASD diagnosis, which depends on the CNS rank as well as the degree of impairment of other body structures (e.g., FAS facial morphology). Although these two proxies will be highly correlated, as a group, those with FAS facial morphology have a different pattern of CNS abnormality than that found in children with FASD who lack those facial features, including increased prevalence of frontal lobe hypoplasia, reduced choline levels, and increased overall severity of CNS abnormality (Astley, 2013). These differences cannot be captured by use of the CNS rank alone; therefore, both proxies were examined.

Method

Study Participants

Participants included 152 school-aged children who ranged in age from 6;0 (years;months) to 14;0: a clinical sample of 72 children with CNS abnormalities previously diagnosed during assessments for suspected FAS—the FASD group; and a control group of 80 children with no indication of frank CNS abnormality suggested by a school-based screening—the CG participants. The evidence presented in this study was taken from existing clinical and research databases gathered from the years 1995 through 2010.

FASD Group

The FASD group (n = 72) is a subset of patients evaluated for suspected FASD by an interdisciplinary team using the FASD 4-Digit Code at the University of Washington Fetal Alcohol Syndrome Diagnostic & Prevention Network (FASDPN). To be included in the FASD group, children had to have participated in a narrative assessment during an FASDPN evaluation. There needed to be an audio recording of the narrative assessment available, and their legal guardian had to have provided written consent to use the clinical diagnostic data for research purposes. Only children with normal hearing between ages 6;0 and 14;0 at time of narrative assessment and who had English as the primary home language were included. All children had confirmed alcohol exposure and/or the unique cluster of minor facial anomalies of FAS. Following Astley et al. (2009), each had received one of the following (increasingly severe) diagnoses from the FASDPN clinic evaluation in accordance with the 4-Digit Code:

1. Neurobehavioral disorder
   (a) mild to moderate CNS abnormality (CNS Rank 2)
2. Static encephalopathy
   (a) significant CNS abnormality (CNS Rank 3 or 4)
   (b) FAS facial phenotype absent
3. Full or partial FAS
   (a) significant CNS abnormality (CNS Rank 3 or 4)
   (b) FAS facial phenotype present

Seventy-two children met the inclusion criteria. One had the full face of FAS without independently confirmed alcohol exposure (i.e., smooth philtrum, thin upper lip, and small eyes: 4-Digit Code 3422; see Appendix B). Given that the face of FAS is considered a reliable and valid proxy for alcohol exposure (see Astley, 2013), this participant was included in our analysis. A clinical appraisal of language performance completed at the time of the initial assessment by the clinical team (which included a certified speech-language pathologist) indicated language impairment in 44 of these 72 children (61%), with 21 rated as having mild to moderate language impairment (i.e., equivalent to between 1 and 2 *SD*s from the mean on a standardized test) and 23 rated as having significant impairment (i.e., equivalent to 2 or more *SD*s from the mean on a standardized test).

Control Group

The CG (n = 80) comprised a broad sample of children recruited from public schools in the greater metropolitan Seattle area. These schools had median family incomes and sociodemographic characteristics that were similar across school districts and representative of the area from which our clinical FASD sample was taken. No intelligence or standardized language measures were generated for control participants, and prenatal alcohol exposure was not directly assessed. However, experienced school psychologists familiar with these children and knowledgeable about FASD screened the respective school records for classroom performance and general behavior. On the basis of these reviews, each child was (a) considered at low risk of having an undiagnosed FASD and (b) deemed to be following a developmental course representative of the general population due to unremarkable social behavior and satisfactory, yet unexceptional, school achievement.

This screening protocol provided converging ecological evidence of generally appropriate development in relevant areas including oral language ability. This selection process mirrored standard practice for identification of language impairment and neurodevelopmental disability in school-aged populations, whereby only those children with identified concerns receive an in-depth assessment of abilities; the absence of such concerns serving as a proxy for typical and functionally appropriate skill development.

To be included, each CG participant participated in a narrative assessment conducted through the University of Washington and was between 6 and 14 years of age. The children had English as their primary home language and no concerns with respect to hearing sensitivity. Eighty control participants were identified.

Group Matching on Demographic Variables

Although the CG participants had a wider range of ages and were slightly older on average, a Welch test revealed no significant difference in the age distribution across groups: CG mean = 10.2 years, range = 7;2 to 14;5; FASD mean = 9.7 years, range = 6;4 to 12;8; difference in means 0.5 years, test statistic $t(d) = -1.601$, two-tailed probability $p = .112$. Similarly, Fisher's exact test of proportions indicated no significant difference in the gender distribution between groups (CG: 46 boys, 34 girls; FASD: 34 boys, 36 girls, two unknown; Fisher's exact test: $p = .326$). The control group used here was gathered from public schools in the community with median family incomes and sociodemographic characteristics similar across school districts and representative of the area (Coggins, 1995). As is true of any clinically referred group, the clinical population from which the FASD group was pulled differs from the general population in terms of income and other demographic variables (see Astley, 2010b). However, because specific details of race, income, and other socioeconomic variables were unavailable for members of the CG group, no direct between-groups comparison on these variables was conducted.

Narrative Data

Narratives were collected as part of clinical assessments at the FASDPN or as part of research at the University of Washington Child Language Lab between the years 1995 and 2010. Each FASD and CG participant had an audio recording of a story elicited with the wordless picture book, *Frog, Where Are You* (Mayer, 1969). Stories were told to a naïve listener. Specifically, the children were given the storybook to review, so they could become familiar with the story. After the children previewed the storybook, the listener/examiner asked the participants to tell "the best story possible" while using the picture book as a visual prompt. The listeners/examiners were always seated so as to make it clear that they were unable to see the storybook pictures. The elicitation protocol is detailed in Appendix C.

Each narrative was audio-recorded. Although approximately a third of the narratives included in our sample had been previously examined (Grittner, Coggins, Thorne, & Olswang, 2009), new transcripts for all narratives were prepared for the current study from the source audio using a systematic transcription process designed to ensure the fidelity and uniformity of the transcripts. Two groups of trained transcribers independently transcribed all narratives from audio sources. The research team, including these transcribers, two additional trained research assistants, and the second author, identified differences between the two sets of independent transcripts and came to agreement as to how these differences should be resolved on the basis of additional review of the audiotapes. All subsequent analyses were conducted on these consensus transcripts.

The consensus transcriptions were segmented, coded, and formatted according to conventions from Systematic Analysis of Language Transcripts (Miller, 2004). In addition, 20% of transcripts were randomly selected, segmented, coded, and formatted independently by another trained coder. Interrater reliability between these coders was 90% or greater for all segmentation, formatting, and coding decisions. A licensed speech-language pathologist (the first author) who was blind to age, gender, and diagnostic status of each storyteller analyzed narrative transcripts utilizing the TREIN protocol.

The primary outcome measures generated by a TREIN analysis are

1. Nominal reference errors (NRE) = ambiguous introductions + ambiguous nominal ties (i.e., "[ambigintro] + [ambignties]");
2. Pronominal reference errors (PRE) = pronoun introductions + ambiguous pronoun ties (i.e.,"[pnintro] + [ambigpntie]");
3. Total reference errors (ALL) = [ambigintro] + [ambigntie] + [ambigpntie] + [pnintro]; and
4. Reference opportunities (opp) = NRE + PRE + all other coded introductions and ties.

These counts allow for the calculation of six cohesive referencing error rates that control for narrative length:

1. Rate of nominal reference errors (rNRE) = NRE/NTW;
2. Rate of nominal reference errors by opportunity (rNREopp) = NRE/opp;
3. Rate of pronominal reference errors rPRE = PRE/NTW; and
4. Rate of pronominal reference errors by opportunity (rPREopp) = PRE/opp;
5. Rate of all reference errors (rALL) = (NRE + PRE)/NTW; and
6. Rate of all reference errors by opportunity (rALLopp) = (NRE + PRE)/opp.

Each of the rates defined above is converted to a percentage for reporting.

TREIN coding reliability was established through independent coding of 20% of the corpus by a trained research assistant also blind to age, gender, and diagnostic status of each storyteller. Point-by-point interrater comparison across 8,085 words and 2,409 assigned TREIN codes found agreement on 2,276 coding decisions (95% overall point-by-point agreement; range across 30 transcripts = 100% to 88%; median = 96%; mode = 98%). For the one transcript with agreement falling below 90% agreement (eight disagreements across 69 coding decisions), seven of eight disagreements involved coding of the pronouns *he/him* when the boy and the dog in the story were involved in joint actions. Across the 30 transcripts, pronoun coding accounted for nearly two thirds of 133 disagreements, with only 45 involving noun phrases.

Analysis

Hypothesis 1: Children with CNS abnormality associated with FASD will exhibit more reference errors than children without these impairments.

Design. All 152 children were included in this analysis.

Primary analysis. For each of the six TREIN error rates detailed above, a *t* test for independent groups (CG vs. FASD) was completed (Welch test for unequal variance; two-tailed $p < .05$). A Bonferroni correction for multiple comparisons was used, and alpha was set at .007. AUC was used as an effect-size metric; it scales the ability of a measure to predict group membership, with a value greater than .70 considered important. Although an AUC at this level would not necessarily be sufficient to guarantee clinical utility for a tool that was used in isolation to diagnosis the presence of an FASD, any single measure with an AUC of this magnitude would be predicted to contribute meaningfully to overall diagnostic utility when combined with other measures, as would always be the case in the diagnosis of a complex spectrum of disorders such as that found with FASD.

Supplementary analyses. Performance was compared by gender using a Welch test for unequal variance, with two-tailed $p < .05$. Correlation of each measure to age was examined using Pearson's r ($p < .05$).

Hypothesis 2: The proportion of children in the FASD group who generate a narrative with a rNRE in the clinically impaired range will be equivalent to or greater than 50%.

Design. All 152 children were included in this analysis. The *clinically impaired range* was defined as performance greater than 2 *SD*s above the mean of the control group for the rate of NRE.

Primary analysis. A simple comparison of proportion was made to determine whether the number of children in the FASD group exceeded 50%, with a one-sample test of proportions (two-tailed $\alpha = .05$) to determine equivalency between the observed proportion of children in the defined impairment range for rNRE and a prediction of 50% for any value falling below 50%. This was done for both rNRE (which controls for length using number of total words) and rNREopp (which controls for length using number of referential opportunities).

Supplementary analysis. Language impairment status of those falling within and outside of the clinically impaired range supplied additional descriptive information related to the clinical utility of both rNRE and rNREopp.

*Hypothesis 3: As severity of the 4-Digit Code CNS rank increases from Rank 1 (*no concern for CNS abnormality*) to Rank 4 (*definite CNS abnormality*), the proportion of children who generate a narrative with a rNRE in the clinically impaired range will increase for each rank.*

Design. All 152 children were included in this analysis. Children from the control group were given a default CNS Rank 1 indicating no concern for CNS impairment, whereas FASD cases ranged from CNS Rank 2 to 4. All children with a CNS Rank 4 received that rank due to microcephaly (i.e., an occipital frontal circumference more than 2 *SD*s below the mean for age). The clinically impaired range was defined as performance greater than 2 *SD*s above the mean of the control group for rNRE.

Analysis. Along with visual inspection of the distribution of children with impairment versus those without across categories, a chi-square test for trend (with a two-tailed p value of .05) was used to compare the proportion of children having an elevated rNRE (+2 *SD*s above control group mean) for each categorical rank. Given the high correlation between rNRE and rNREopp (.99, $p < .0001$), this analysis was not carried out with rNREopp.

Hypothesis 4. As severity of diagnosis increases from no-FASD to neurobehavioral disorder to static encephalopathy to FAS, the proportion of children who generate a narrative with a rNRE in the clinically impaired range will increase.

Design. This analysis examined the association between the rNRE and severity of diagnosis, taking advantage of the gradation in severity of diagnosis available across our sample, including both the CG and the FASD groups. Four categories were defined: *No concern for alcohol-related CNS abnormality*, *neurobehavioral disorder*, *static encephalopathy*, and *FAS* (following Astley et al., 2009). The clinically impaired range was defined as performance greater than 2 *SD*s above the mean of the control group for rNRE.

Analysis. Along with visual inspection of the distribution of children with impairment versus those without

across categories, a chi-square test for trend (with a two-tailed p value of .05) was used to compare the proportion of children performing in the impaired range for each categorical rank. Given the high correlation between rNRE and rNREopp (.99, $p < .0001$), this analysis was not carried out with rNREopp.

Results

Hypothesis 1: Children with CNS abnormality associated with FASD will exhibit more reference errors than children without these impairments.

The mean performance for each TREIN error rate was significantly different between the CG and FASD groups for all measures that included NRE (rNRE, rNREopp, rALL, rALLopp), with FASD group means almost twice that of the control group for NRE. As can be seen in Table 1, those measures that were based on pronominal errors were not significantly different despite being approximately 50% higher in the FASD group.

Consistent with previous studies (Thorne & Coggins, 2008a, 2008b), AUC greater than .70 was found for both measures on the basis of the rNRE (rNRE% and rNREopp%). The AUC for both measures that were based on the rPRE (rPRE% and rPREopp%) were below .70, indicating that PRE has less promise as a behavioral marker of underlying CNS abnormality in this age range. This is consistent with previous research on the TREIN. As can be seen in Table 1, when performance was compared by gender, no significant differences were found for any TREIN measure. Age, however, was significantly correlated with performance for all TREIN measures in the control group but not in the FASD group. On the basis of visual inspection of the distribution of performance across the control group, this correlation appears to be driven largely by a decrease in the variation of scores as age increases, with younger children in the control group, particularly those 9 years and under having a wider range of variation than older children in the control group. Because of the correlation between age and performance in the control group, post hoc analyses of this impact are included later in this article as appropriate.

Hypothesis 2: The proportion of children in the FASD group who generate a narrative with an rNRE in the clinically impaired range will be equivalent to or greater than 50%.

Table 1. Group performance on each tallying reference errors in narratives (TREIN) outcome measure.

Parameter	CG	FASD	CG	FASD	CG	FASD
Based on number of total words (NTW)	**rNRE = NRE/NTW**		**rPRE = PRE/NTW**		**rAll = All/NTW**	
N	80	72	80	72	80	72
M (%)	1.52	2.88	1.22	1.83	2.74	4.71
Variance	0.748	2.611	1.720	3.882	3.111	9.520
SD	0.865	1.616	1.312	1.970	1.764	3.086
Minimum (%)	0.0	0.29	0.0	0.0	0.0	1.36
Maximum (%)	3.89	8.89	7.14	11.46	8.04	17.71
Correlation to age (*r*)	–0.446	–0.0334	–0.237	–0.170	–0.395	–0.126
p	< .0001	.780	.034	.154	.0003	.292
Welch test						
Difference	1.361		0.607		1.969	
Test statistic *t*(*d*)	6.374		2.211		4.759	
p	< .0001*		.0289: NS		< .0001*	
Effect size, AUC [95% CI]	.78** [0.703, 0.841]		.60 [0.522, 0.683]		.73 [0.647, 0.794]	
Based on number of reference opportunities (opp)	**rNREopp = NRE/opp**		**rPREopp = PRE/opp**		**rAllopp = All/opp**	
N	80	72	80	72	80	72
M (%)	5.57	10.43	4.41	6.51	9.98	16.94
Variance	10.325	33.648	21.731	46.295	40.386	115.737
SD	3.213	5.801	4.662	6.804	6.355	10.758
Minimum (%)	0.0	1.19	0.0	0.0	0.0	5.00
Maximum (%)	14.58	30.77	23.88	37.93	26.87	58.62
Correlation with age (*r*)	–0.431	–0.0626	–0.230	–0.173	–0.386	–0.143
p	.0001	.601	.0406	.146	.0004	.230
Welch test						
Difference	4.857		2.102		6.960	
Test statistic *t*(*d*)	6.290		2.198		4.789	
p	< .0001*		.0298: NS		< .0001*	
Effect size, AUC [95% CI]	.78** [0.700, 0.838]		AUC = .60 [0.521, 0.681]		AUC = .73 [0.648, 0.795]	

Note. No significant difference for gender was found for any TREIN metric ($p < .05$). CG = control group; FASD = fetal alcohol spectrum disorders; rNRE = rate of nominal reference errors; rPRE = rate of pronominal reference errors; AUC = area under the curve; CI = confidence interval.

*Significant difference at $p < .007$; NS = nonsignificant at $p < .007$ after Bonferroni correction. **AUC with the lower bound of the 95% CI ≥ 0.70 are considered important.

For the metric rNRE, the +2 *SD*s cut-point for impairment was rNRE > 3.25%. For the metric rNREopp, the +2 *SD*s cut-point for impairment was rNREopp >12.00%. For both metrics, the percentage of cases within the FASD group falling in the clinically impaired range was compared with a prediction of 50%. Of the 72 cases, 24 (33.33%) fell in the clinically impaired range for rNRE. A total of 18 cases (29.17%) fell in the clinically impaired range for rNREopp. Although a substantial fraction of children with FASD had elevated rates of both rNRE and rNREopp, these results were contrary to our hypothesis with proportions both below and significantly different from 50% (see Table 2).

Language impairment status. Of the 24 children with elevated rNRE, 14 had clinically identified language impairment, eight in the significant range. Of the 18 children with elevated rNREopp, 12 had clinically identified language impairment, seven in the significant range. Ten children who were not identified as having language impairment during their initial clinical visit had an elevated rNRE. Six children who were not identified as having language impairment during their initial clinical visit had an elevated rNREopp.

Post hoc analysis of Hypothesis 2, adjusting for age. The prediction that a proportion equivalent to 50% or more of children in our sample would be in the clinically impaired range was based on previous research that examined performance in a group of children approximately 9 to 12 years old (Thorne & Coggins, 2008a). Given that all TREIN error measures correlated significantly with age in the control group (with younger children making more errors), and given that our sample included many children significantly younger than 9 years old, it is possible that results for the older children in our sample would be different from those for the whole group. For this reason, Hypothesis 2 was examined for a subset of the older children in our sample: those 9 to 12 years old. All children older than 12 or below the age of 9 were removed from the sample. The means and standard deviations for rNRE and rNREopp were recalculated with the remaining 34 children in the control group.

The resulting mean rNRE of 1.37%, with a standard deviation of 0.767, provided a new +2 *SD* cutoff score for rNRE of 2.90. The rNRE for the 45 children between the ages of 9 and 12 years from the FASD group were then compared with this cutoff. With 19 of these 45 children exceeding this cutoff, 42% would be considered in the clinically impaired range, a proportion below but not statistically different from the prediction of 50% in a sample this size (see Table 3). Results were similar when using the metric rNREopp, with 18 cases (40%) exceeding a +2 *SD* cutoff of 10.5% on the basis of a mean of 5.00% and standard deviation of 2.744. This is again below but not statistically different than 50% (see Table 3). These results suggest that the rNRE has more clinical potential for identifying significant CNS abnormality associated with prenatal alcohol exposure in older children.

Language impairment status. Only six of the older children with significantly elevated NRE (about one third using either metric) had language impairment identified at their initial clinical visit, whereas approximately two thirds did not.

*Hypothesis 3: As severity of 4-Digit Code CNS rank increases in severity from Rank 1 (*no concern for CNS abnormality*) to Rank 4 (*definite CNS abnormality*), the proportion of children who generate a narrative with an rNRE in the clinically impaired range will increase for each rank.*

On the basis of a chi-square test for trend, the proportion of children meeting criteria for impairment based on their rNRE (+2 *SD* rNRE) was significantly different

Table 2. Percentage of FASD cases falling in the impaired range for nominal reference errors compared with prediction of 50% (test for one proportion).

Parameter	Value
Observed proportion of 72 children with FASD in the clinically impaired range (i.e., +2 *SD* of CG mean)	
rNRE (%)	> 3.25 = 33.33 (24 children)
rNREopp (%)	> 12 = 29.17 (18 children)
95% CI of observed proportion	
rNRE	22.86, 45.16
rNREopp	19.05, 41.07
z statistic	
rNRE	2.887
rNREopp	3.535
Significance level (*p*)	
rNRE	.0039
rNREopp	.0004

Note. FASD = fetal alcohol spectrum disorders; CG = control group; rNRE = rate of nominal reference errors; opp = opportunity; CI = confidence interval.

Table 3. Percentage of FASD cases ages 9 through 12 years falling in the impaired range for nominal reference errors compared with prediction of 50% (test for one proportion).

Parameter	Value
Observed proportion of 45 children with FASD in the clinically impaired range (i.e., +2 *SD* of CG group)	
rNRE (%)	> 2.90 = 42.22 (19 children)
rNREopp (%)	> 10.50 = 40.00 (18 children)
95% CI of observed proportion	
rNRE	27.65, 57.85
rNREopp	25.7, 55.67
z statistic	
rNRE	1.044
rNREopp	1.342
Significance level (*p*)	
rNRE	.297
rNREopp	.180

Note. Cutoff scores are based on control group (CG) participants aged 9 through 12 years (n = 34). FASD = fetal alcohol spectrum disorders; rNRE = rate of nominal reference errors; CI = confidence interval.

between categories defined by 4-Digit Code CNS rank (χ^2 trend: 24.057; $p < .0001$). There was a linear trend (see Figure 1) indicating increasing proportions of impairment in groups moving from CNS Rank 1 (5%) to CNS Rank 2 (27%) to CNS Rank 3 (31%) to CNS Rank 4 (53%). We found it interesting that the group with CNS Rank 4 (i.e., those with microcephaly) was nearly twice as likely to exhibit significant impairment on the basis of rNRE as the children with CNS Ranks 2 and 3. Two thirds of the 9- to 12-year-old children with a CNS Rank 4 had rNRE above the cut-point.

Hypothesis 4. As severity of diagnosis increases from typical development to neurobehavioral disorder to static encephalopathy to FAS, the proportion of children who generate a narrative with an rNRE in the clinically impaired range will increase.

On the basis of a chi-square test for trend, the proportion of children meeting criteria for impairment on the basis of elevated rNRE was significantly different between diagnostic categories (χ^2 trend: 22.939; $p < .0001$) with a linear trend (see Figure 2) indicating larger proportions in groups with more severe diagnoses moving from *no concern for CNS abnormality* (5%) to *neurobehavioral disorder* (28%) to *static encephalopathy* (36%) to FAS (57%). All children over the age of 9 years with FAS had rNRE above the cut-point.

Discussion

This research examined the potential for cohesive referencing errors made during narrative discourse to provide a behavioral marker of underlying risk of CNS abnormality in school-age children. A retrospective comparison was conducted between two groups: a clinical population consisting of children who had a previously diagnosed CNS abnormality found during an interdisciplinary diagnostic evaluation for suspected FASD (the FASD group, who served as a proxy for the larger population with CNS abnormality) and a group of children with no concerns related to their neurocognitive development, general behavior, or academic achievement (the CG children). It was predicted that the presence of a CNS abnormality would increase the likelihood that children would have difficulty maintaining cohesive referencing. It was certainly not expected that all children in the clinical group would demonstrate difficulty with cohesive referencing given the remarkable variability in their language performance, type and degree of CNS abnormality, and the severity of FASD diagnosis; however, given that more severely involved children are, by definition, affected across a wider range of functional domains, it was predicted that groups of children with more severe presentations would include higher proportions of children who had difficulty maintaining cohesive reference.

The accumulated evidence presented a clear association between a clinically significant elevation in the rNRE, a metric of cohesive referencing, and previously diagnosed CNS abnormality related to prenatal alcohol exposure. As predicted, the existence of a significantly elevated rNRE is more common in children in our FASD group

Figure 1. Proportion of children above and below cutoff of +2 *SD* rNRE for each 4-Digit Code CNS rank ($N = 152$). SI = severe impairment; rNRE = nominal reference errors; CNS = central nervous system.

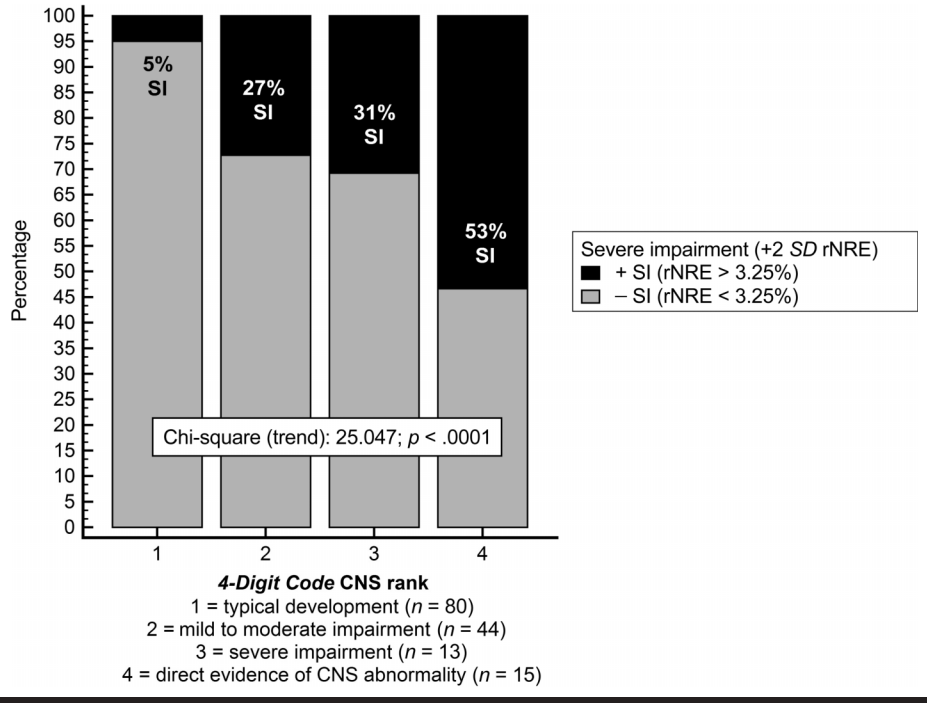
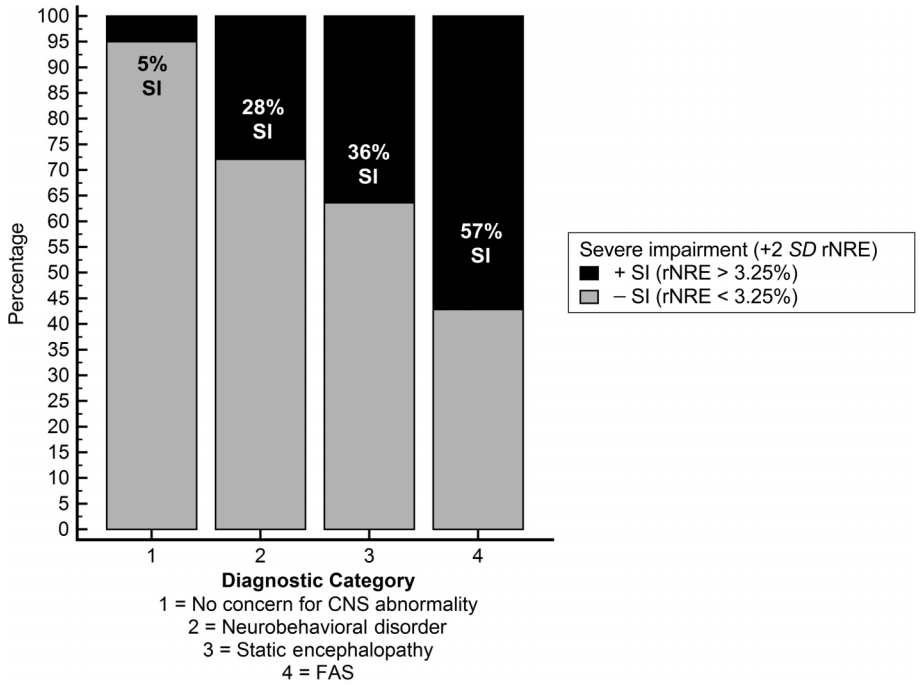


Figure 2. Proportion of children above and below impairment cutoff of +2 *SD* in 4 diagnostic categories (*N* = 152). SI = severe impairment; rNRE = nominal reference errors; CNS = central nervous system; FAS = fetal alcohol syndrome.



than in their peers without FASD and that incidence increases predictably in association with more severe manifestation of the underlying FASD, whether that severity is based on CNS rank (Hypothesis 3) or diagnostic category (Hypothesis 4). Risk was particularly elevated for those with microcephaly (53% in the impaired range) or a diagnosis of FAS (57% in the impaired range). This relationship was most clearly apparent in older children, with almost half of the older children with FASD, two thirds of the older children with microcephaly, and all of the older children with FAS in our sample exhibiting significantly impaired cohesive referencing when compared with peers without concerns for CNS abnormality. The fact that a significant correlation between error rates and age was found in the control group but was not found in the FASD group suggests that the CNS abnormalities found in the latter may be preventing many of these children from following the expected developmental course related to these errors. It is important to note there was nearly twice the incidence of significantly elevated NRE rates in the group of children with microcephaly (i.e., those with direct evidence of underlying CNS abnormality) when compared with those who had significant functional impairments but no microcephaly.

During their original clinical visit, 28 children in the FASD group left their clinical appointment without a language impairment being identified, with another 21 identified as having only mild to moderate language impairment. The remaining 23 children in the FASD group (32%) were identified as having significant language impairment that could contribute to a diagnosis of static encephalopathy under the 4-Digit Code. Of the 28 children in the FASD group who left their clinical appointment without language impairment being identified, 10 (36%) exhibited elevated NRE rates that would indicate language impairment severe enough to potentially contribute to a diagnosis of static encephalopathy under the 4-Digit Code. In addition, of the 21 children in the FASD group who left their clinical appointment with a diagnosis of mild to moderate language impairment, six (29%) exhibited elevated NRE rates that would be considered severe enough to potentially contribute to a diagnosis of static encephalopathy under the 4-Digit Code. Our results indicate, therefore, that the addition of TREIN analysis to the clinical process with this group of children would have identified an additional 16 children with this more severe level of impairment, potentially improving capture of "significant impairment" from 32% to 54%. This is consistent with the substantial effect size seen for both NRE measures (AUC = .78), indicating important diagnostic potential.

In the children aged 9 to 12 years, 14 of 45 children in the FASD group (31%) were identified as having significant language impairment in their clinical visit. Inclusion of the TREIN would have captured an additional 15 children, improving capture of "significant impairment" from 31% to 64%. In this age group, including a tally of NRE as part of standard clinical practice would have more than doubled sensitivity to underlying CNS abnormality when compared with a standard assessment of language capacity.

Conclusions

Clinical Utility

Speech-language pathologists have long recognized the need for information about the discourse-level integrative language abilities of children with language disorders. As members of interdisciplinary diagnostic teams, however, speech-language pathologists also play an important role in the diagnosis of a wide variety of other neurodevelopmental disorders, including FASD. The evidence presented here indicates that one discourse-level task has diagnostic utility for this purpose in that it may provide a behavioral marker of underlying CNS abnormality in many children with FASD whether or not they have a language disorder. Measuring the rNRE as defined by the TREIN during a simple story generation task has the potential to substantially increase sensitivity to underlying CNS abnormality in children with prenatal alcohol exposure. This was true whether NRE were quantified in terms of total words or referential opportunities. Both methods are easy to calculate, and although the rNRE (calculated on the basis of total words) slightly outperformed rNREopp (calculated based on referential opportunities), clinicians may find rNREopp easier to interpret for clinical purposes. Given the high correlation between these metrics (.99, $p < .0001$), the ease of interpretation may make rNREopp a more practical choice for future development and validation. As the narrative generation task used in this research allows for examination of other expressive language abilities in addition to cohesive referencing, it may be a highly efficient means for gathering evidence about the development of individuals seeking neurodevelopmental diagnoses more broadly.

Further Research

The current research is an early step in a program of research designed to understand the neurocognitive underpinnings of discourse-level integrative language capacities in school-aged children. Because this research was conducted using a retrospective sampling of existing data, we were not able to target specific hypotheses related to the connection between specific brain networks and cohesive referencing abilities in school-age children. Information related to the structural integrity of specific neural systems in the children with impaired cohesive referencing was unavailable. However, the strong relationship between increased risk for impaired cohesive referencing skills and increased risk of underlying CNS abnormality in children with FASD provides motivation to do prospective research of these skills that incorporates direct imaging of the relevant structures and their interconnectivity. Candidate structures would include the caudate nucleus of the basal ganglia and the frontal lobes, which appear to be particularly vulnerable to prenatal alcohol exposure (see e.g., Astley et al., 2009) and may have an important role to play in language functioning (see e.g., Alexander, 2006; Lee & Tomblin, 2012). These structures are part of the diverse network of brain structures that support the cognitive control needed to respond to dynamic contexts such as narrative discourse (see e.g., Alexander, 2006; Casey, Tottenham, & Fossella, 2002; Cools, Ivry, & D'Esposito, 2006; Hikosaka & Isoda, 2010; Kerstin, Kraft, Kehrer, & Brandt, 2014; Monchi, Petrides, Strafella, Worsley, & Doyon, 2006; Teichmann, Dupoux, Kouider, & Bachoud-Levi, 2006). If a more direct link between the structural impairment to this cognitive control network and impaired cohesive referencing can be made, this would provide an important step toward understanding the neurodevelopmental underpinning of integrative language capacities that support discourse functioning.

Along this line, it seems unlikely that impairments of cohesive referencing identified in this research are specific to children with FASD because the CNS abnormalities found in FASD are unlikely to be unique to those with prenatal alcohol exposure. The findings from this research, therefore, would also provide motivation to examine the clinical utility of the TREIN analysis for use with other populations, particularly those that share overlapping and contrasting symptom profiles to those found in FASD. Candidate populations would include individuals without prenatal alcohol exposure diagnosed with attention-deficit/hyperactivity disorder, language disorders, auditory processing disorder, autism spectrum disorder, and intellectual disability.

A particular question that emerges from the results of this study is whether impairment of referential cohesion as captured by the TREIN reflects reduced "language capacity," making this a behavioral marker of a clinically important functional language impairment; or, alternatively, whether this impairment in cohesive referencing is instead a reflection of reduced capacity for social cognition (Bakopoulou & Dockrell, 2016; Coggins et al., 2007; Stevens, Dudek, Nash, Koren, & Rovet, 2015) and/or reduced cognitive control/executive functioning that manifests as reduced language performance during more complex, linguistically loaded tasks (Burden et al., 2011; Kingdon, Cardoso, & McGrath, 2015; Kodituwakku & Kodituwakku, 2014; Ye & Zhou, 2009). Prospective research examining the relationship between referential cohesion as captured by the TREIN and language, social cognition, and executive control capacities in those with and without prenatal alcohol exposure could be used to help clarify these important issues.

Limitations

The CG used here was gathered from public schools in the community with median family incomes and sociodemographic characteristics similar across school districts and representative of the area (Coggins, 1995). However, because details of race, income, and other socioeconomic variables were unavailable for the CG participants, no direct between-groups comparisons on these variables is possible. As is true of any clinically referred group, the clinical population from which the FASD group was pulled differs from the general population in the area in terms of

income and other demographic variables (see Astley, 2010b). This increases the chances of Type I errors in our results, with socioeconomic differences being an important alternative explanation for any contrast between the FASD group and CG participants. In addition, the lack of objective language and cognitive measures for the CG children increases the chances of Type I errors in the unlikely event that as a group their ability was significantly above average along the parameters of interest.

In contrast to the FASD group, the CG participants were not assessed by an interdisciplinary diagnostic team. However, in standard clinical practice, a socioeconomic, academic, and behavioral profile like that presented by the CG participants would not typically trigger a comprehensive team assessment. Thus, even though the CG participants did not undergo the same battery of cognitive, linguistic, and behavioral/social measures that the FASD participants completed, we believe that they nevertheless provide a reasonable basis for contrasting group performance in the context of this study—with the CG participants having a profile similar to most typically developing children attending public schools. It must be kept in mind that the lack of objective measures confirming that the CG children were indeed typically developing—and not subject to a prenatal alcohol exposure or unidentified language impairment—also increases the risk of Type II errors. The risk of both Type I and Type II errors resulting from our use of retrospective data enhance the need for the results of this initial research to be confirmed with prospective validation research.

Acknowledgments

The authors would like to acknowledge Susan J. Astley, Lesley B. Olswang, Stacy K. Betz, Heather Carmichael Olson, Sara Jerger, and everyone at the University of Washington Fetal Alcohol Syndrome Diagnostic and Prevention Network and the University of Washington Child Language Lab for their support in completion of this research.

Portions of the research reported here were conducted in partial fulfillment of the requirements for the degree of doctor of philosophy at the University of Washington by John C. Thorne and are discussed in his doctoral dissertation.

References

Alexander, M. P. (2006). Impairments of procedures for implementing complex language are due to disruption of frontal attention processes. *Journal of the International Neuropsychological Society, 12,* 236–247.

Ariel, M. (1994). Interpreting anaphoric expressions: A cognitive versus a pragmatic approach. *Journal of Linguistics, 30,* 3–42.

Ariel, M. (2004). Accessibility marking: Discourse functions, discourse profiles, and processing cues. *Discourse Processes, 37,* 91–116.

Ariel, M. (2009). Discourse, grammar, discourse. *Discourse Studies, 11,* 5–36.

Astley, S. J. (2004). *Diagnostic guide for fetal alcohol spectrum disorders: The Four-Digit Diagnostic Code* (3rd ed.). Seattle, WA: FAS Diagnostic and Prevention Network, University of Washington. Electronic version available from http://fasdpn.org

Astley, S. J. (2010a). [Narratives produced by children with prenatal alcohol exposure]. Unpublished raw data.

Astley, S. J. (2010b). Profile of the first 1,400 patients receiving diagnostic evaluations for fetal alcohol spectrum disorder at the Washington State Fetal Alcohol Syndrome Diagnostic & Prevention Network. *Canadian Journal of Clinical Pharmacology, 17,* e132–e164.

Astley, S. J. (2013). Validation of the fetal alcohol spectrum disorder (FASD) 4-Digit Diagnostic Code. *Journal of Population Therapeutics and Clinical Pharmacology, 20,* e416–e467.

Astley, S. J., Aylward, E., Olson, H. C., Kerns, K., Brooks, A., Coggins, T. E., . . . Richards, T. (2009). Magnetic resonance imaging outcomes from a comprehensive magnetic resonance study of children with fetal alcohol spectrum disorders. *Alcoholism: Clinical and Experimental Research, 33,* 1671–1689.

Bertrand, J., Floyd, L. L., & Weber, M. K. (2005). Guidelines for identifying and referring persons with fetal alcohol syndrome. *Morbidity and Mortality Weekly Report: Recommendations and Reports, 54*(RR-11), 1–14.

Bakopoulou, I., & Dockrell, J. E. (2016). The role of social cognition and prosocial behaviour in relation to the socio-emotional functioning of primary aged children with specific language impairment. *Research in Developmental Disabilities, 49,* 354–370. doi:10.1016/j.ridd.2015.12.013

Bakoyiannis, I., Gkioka, E., Pergialiotis, V., Mastroleon, I., Prodromidou, A., Vlachos, G. D., & Perrea, D. (2014). Fetal alcohol spectrum disorders and cognitive functions of young children. *Reviews in the Neurosciences, 25,* 631–639. doi:10.1515/revneuro-2014-0029

Burden, M., Westerlund, A., Muckle, G., Dodge, N., Dewailly, E., Nelson, C. A., . . . Jacobson, J. L. (2011). The effects of maternal binge drinking during pregnancy on neural correlates of response inhibition and memory in childhood. *Alcoholism: Clinical and Experimental Research, 35,* 69–82. doi:10.1111/j.1530-0277.2010.01323.x

Carmichael-Olsen, H., & Astley, S. J. (2005). [Intervening with children/adolescents with FAS/ARND]. Unpublished raw data.

Casey, B. J., Tottenham, N., & Fossella, J. (2002). Clinical, imaging, lesion, and genetic approaches toward a model of cognitive control. *Developmental Psychobiology, 40,* 237–254.

Chudley, A. E., Conry, J., Cook, J. L., Loock, C., Rosales, T., & LeBlanc, N. (2005). Fetal alcohol spectrum disorder: Canadian guidelines for diagnosis. *Canadian Medical Association Journal, 172*(5, Suppl.), S1–S21.

Coggins, T. E. (1995). [Narratives produced by typically developing school-aged children]. Unpublished raw data.

Coggins, T. E., Timler, G. R., & Olswang, L. B. (2007). A state of double jeopardy: Impact of prenatal alcohol exposure and adverse environments on the social communicative abilities of school-age children with fetal alcohol spectrum disorder. *Language, Speech, and Hearing Services in Schools, 38,* 117–127.

Cook, J. L., Green, C. R., Lilley, C. M., Anderson, S. M., Baldwin, M. E., Chudley, A. E., . . . Canada Fetal Alcohol Spectrum Disorder Research Network. (2016). Fetal alcohol spectrum disorder: A guideline for diagnosis across the lifespan. *Canadian Medical Association Journal*, *188,* 191–197. doi:10.1503/cmaj.141593

Cools, R., Ivry, R. B., & D'Esposito, M. (2006). The human striatum is necessary for responding to changes in stimulus relevance. *Journal of Cognitive Neuroscience, 18,* 1973–1983.

Donald, K. A., Eastman, E., Howells, F. M., Adnams, C., Riley, E. P., Woods, R. P., . . . Stein, D. J. (2015). Neuroimaging

effects of prenatal alcohol exposure on the developing human brain: A magnetic resonance imaging review. *Acta Neuropsychiatrica,* 251–269. doi:10.1017/neu.2015.12

Gillam, R. B., & Pearson, N. A. (2004). *Test of Narrative Language*. Austin, TX: Pro-Ed.

Grittner, J. M., Coggins, T. E., Thorne, J. C., & Olswang, L. (2009). *Classification accuracy of nominal reference errors for fetal alcohol syndrome*. Poster presented at the American Speech-Language-Hearing Association National Convention, New Orleans, LA.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London, UK: Longman.

Heilman, J., Miller, J., Nockerts, A., & Dunaway, C. (2010). Properties of the narrative scoring scheme using narrative retells in young school-aged children. *American Journal of Speech-Language Pathology, 19,* 154–166.

Hikosaka, O., & Isoda, M. (2010). Switching from automatic to controlled behavior: Cortico-basal ganglia mechanisms. *Trends in Cognitive Sciences, 14,* 154–161.

Hoyme, H. E., May, P. A., Kalberg, W. O., Kodituwakku, P., Gossage, J. P., Trujillo, P. T., . . . Robinson, L. K. (2005). A practical clinical approach to diagnosis of fetal alcohol spectrum disorders: Clarification of the 1996 institute of medicine criteria. *Pediatrics, 115,* 39–47. doi:10.1542/peds.2004-0259

Johnston, J. R. (1982) Narratives: A new look at communication problems in older language-disordered children. *Language, Speech, and Hearing Services in Schools, 13,* 144–155.

Kerstin, I., Kraft, A., Kehrer, S., & Brandt, S. (2014). Mechanisms and neuronal networks involved in reactive and proactive cognitive control of interference in working memory. *Neuroscience & Biobehavioral Reviews, 46,* 58–70. doi:10.1016/j.neubiorev.2014.06.014

Kingdon, D., Cardoso, C., & McGrath, J. J. (2016). Research Review: Executive function deficits in fetal alcohol spectrum disorders and attention-deficit/hyperactivity disorder—A meta-analysis. *Journal of Child Psychology and Psychiatry*, *57,* 116–131. doi:10.1111/jcpp.12451

Kodituwakku, P., & Kodituwakku, E. (2014). Cognitive and behavioral profiles of children with fetal alcohol spectrum disorders. *Current Developmental Disorders Reports, 1,* 149–160. doi:10.1007/s40474-014-0022-6

Lee, J. C., & Tomblin, J. B. (2012). Reinforcement learning in young adults with developmental language impairment. *Brain and Language, 123,* 154–163.

Liles, B. Z. (1993). Narrative discourse in children with language disorders and children with normal language: A critical review of the literature. *Journal of Speech and Hearing Research, 36,* 868–882.

May, P. A., Baete, A., Russo, J., Elliott, A. J., Blankenship, J., Kalberg, W. O., . . . Hoyme, H. E. (2014). Prevalence and characteristics of fetal alcohol spectrum disorders. *Pediatrics, 134,* 855–866. doi:10.1542/peds.2013-3319

Mayer, M. (1969). *Frog, where are you?* New York, NY: Dial.

Miller, J. F. (2004). *Systematic Analysis of Language Transcripts* (Version 8.02, Research Edition). Madison: University of Wisconsin-Madison.

Monchi, O., Petrides, M., Strafella, A. P., Worsley, K. J., & Doyon, J. (2006). Functional role of the basal ganglia in the planning and execution of actions. *Annals of Neurology, 59,* 257–264.

Popova, S., Lange, S., Shield, K., Mihic, A., Chudley, A. E., Mukherjee, . . . Rehm, J. (2016). Comorbidity of fetal alcohol spectrum disorder: A systematic review and meta-analysis. *The Lancet, 387,* 978–987. doi:10.1016/S0140-6736(15)01345-8

Roozen, S., Gjalt-Jorn, Y. P., Kok, G., Townend, D., Nijhuis, J., & Curfs, L. (2016). Worldwide prevalence of fetal alcohol spectrum disorders: A systematic literature review including meta-analysis. *Alcoholism: Clinical and Experimental Research, 40,* 18–32. doi:10.1111/acer.12939

Sampson, P., Streissguth, A., Bookstein, F., & Barr, H. M. (2000). On categorizations in analyses of alcohol teratogenesis. *Environmental Health Perspectives, 108*(Suppl. 3), 421–428.

Sampson, P., Streissguth, A., Bookstein, F., Little, R., Clarren, S., Dehaene, P., . . . Graham, J. (1997). Incidence of fetal alcohol syndrome and prevalence of alcohol-related neurodevelopmental disorder. *Teratology, 56,* 317–326.

Schneider, P., & Hayward, D. (2010). Who does what to whom: Introduction of referents in children's storytelling from pictures. *Language, Speech, and Hearing Services in Schools, 41,* 459–473.

Stevens, S. A., Dudek, J., Nash, K., Koren, G., & Rovet, J. (2015). Social perspective taking and empathy in children with fetal alcohol spectrum disorders. *Journal of the International Neuropsychological Society, 21,* 74–84. doi:10.1017/S1355617714001088

Streissguth, A. P., Bookstein, F. L., Barr, H. M., Sampson, P. D., O'Malley, K., & Young, J. K. (2004). Risk factors for adverse life outcomes in fetal alcohol syndrome and fetal alcohol effects. *Journal of Developmental and Behavioral Pediatrics, 25,* 228–238.

Strong, C. J. (1998). *The Strong Narrative Assessment Procedure*. Eau Claire, WI: Thinking Publications.

Teichmann, M., Dupoux, E., Kouider, S., & Bachoud-Levi, A. C. (2006). The role of the striatum in processing language rules: Evidence from word perception in Huntington's disease. *Journal of Cognitive Neuroscience, 18,* 1555–1569.

Thorne, J. C. (2006). *Tallying Reference Errors in Narrative* [manual]. Seattle: University of Washington. Available at http://johncthorne.wordpress.com/tallying-reference-errors-in-narrative-trein/

Thorne, J. C., & Coggins, T. E. (2008a). A diagnostically promising technique for tallying nominal reference errors in the narratives of school-aged children with foetal alcohol spectrum disorders (FASD). *International Journal of Language & Communication Disorders, 43,* 570–594.

Thorne, J. C., & Coggins, T. E. (2008b, June). *Signals of CNS damage in the discourse behavior of school-aged children with prenatal alcohol exposure during a narrative generation task*. Poster presented at the Symposium on Research in Child Language Disorders, Madison, WI. Electronic version available at http://johncthorne.files.wordpress.com/2010/05/webversion.ppt

Thorne, J. C., Coggins, T. E., Carmichael Olson, H., & Astley, S. J. (2007). Exploring the utility of narrative analysis in diagnostic decision making: Picture-bound reference, elaboration, and fetal alcohol spectrum disorders. *Journal of Speech, Language, and Hearing Research, 50,* 459–474.

Van Hoek, K. (1997). *Anaphora and conceptual structure*. Chicago, IL: The University of Chicago Press.

Wiig, E. H., & Secord, W. A. (1989). *Test of Language Competence–Expanded Edition*. San Antonio, TX: The Psychological Corporation.

Ye, Z., & Zhou, X. (2009). Executive control in language processing. *Neuroscience & Biobehavioral Reviews, 33,* 1168–1177. doi:10.1016/j.neubiorev.2009.03.003

Appendix A

Details of *Tallying Reference Errors in Narrative* (TREIN; Thorne, 2006)

The TREIN tracks how a participant introduces and maintains reference to concepts throughout the narrative. It quantifies errors of cohesive referencing in noun and pronoun phrases. For example, neither "the frog" nor "it" would unambiguously refer to a *specific* frog following the sentence "They saw **two frogs** sitting on a log" because the forms do not unambiguously indicate to the naïve listener which frog is intended in context. They fail to make an appropriate "referential tie" (see Halliday & Hasan, 1976). Similarly, upon first mention, the use of a definite form, such as "The frog" or "it" (rather than the indefinite "a frog" or a possessive form such as "the boy's/his pet frog"), would be considered a cohesive error because those forms imply that the naïve listener has prior knowledge of that frog as part of the common ground of knowledge shared by the listener and storyteller. This makes them inappropriate forms to introduce a new concept into the discourse. Table A1 provides a description of each of the specific codes used in a TREIN analysis.

Table A1. Summary of TREIN codes.

Code	Definition
Codes for introduction	
+ indefintro	*Indefinite introduction* of concepts (e.g., "**A boy** was looking in **a jar.**").
+ defintro	*Definite introduction* of common concepts/concepts with supporting context (e.g., "**The moon** was out." – MOON is a common concept; or "He knocked down a beehive and **the bees** swarmed." – BEEHIVES have BEES).
+ possintro	*Possessive introduction* of concepts (e.g., "The boy had **his dog** with him")
– ambigintro	*Ambiguous introduction* of concepts using a definite form not supported by context (e.g., "**The frog** escaped" on first mention). Also inappropriate 2nd use of an indefinite form (e.g., "A boy had a frog. **A boy** had a dog.").
– pnintro	*Pronominal introduction* of concept (e.g., "**It** was in there," on first mention).
Codes for referential maintenance	
+ ntie	Clear *referential tie* using nominal form (e.g., "**A boy** captured a frog. Then **the boy** put it in a jar.").
– ambigntie	*Ambiguous referential tie* using nominal form (e.g., "He saw **two frogs**. **The frog** was…"); also mislabeling available concepts (e.g., "dog" for "frog").
+ pntie	Clear *referential tie* using pronominal form (e.g., "**A boy** found a frog. **He** put it in a jar.").
– ambigpntie	*Ambiguous referential tie* using pronominal form (e.g., "**The boy and the dog** were looking for the frog. **He** looked under the bed.").

Note. + indicates an appropriately cohesive strategy; – indicates an inappropriate strategy.

By tallying pronouns and other noun phrases separately, the TREIN acknowledges the subtle differences in behavior between these two syntactic classes when it comes to cohesion. Definite noun phrases are used in English (primarily) to refer to specific and known concepts that are part of the common ground, whereas pronouns are used to refer to specific, known concepts that are the current focus of the discourse (see e.g., Ariel, 2009). Tallying each error type separately allows for examination of the impact that these differences may have on error rates during narrative production. The TREIN training manual and details of the system as implemented in this research are available online: http://johncthorne.wordpress.com/tallying-reference-errors-in-narrative-trein/

Appendix B

4-Digit Codes for *FASD group (n = 72) organized by degree of CNS impairment.

Static encephalopathy			Neurobehavioral disorder	
2444 -FAS	4234	1124	1224	1123
2443 -FAS	4234	1223	1124	1224
3442 -FAS	4234	3223	1124	1224
1443 -FAS	3233	3123	1124	1224
4343 -FAS	1234	1424	1124	1223
4344 -FAS	1234	1424	1124	1223
1343 -FAS	1234	1423	1124	1223
3244	2234	1423	1124	1223
3243	1233	1423	1123	1223
1244	1234	1324	2123	1223
1343	1234	1324	1424	1223
1243	1233	1323	1224	***3422***[a]
4244	1233	2223	1224	
3244	2134	2124	1224	
2244	1134	1224	1123	

Note. FAS = full or partial fetal alcohol syndrome according to Astley (2004), following Astley et al., 2009.

Sources: Unpublished raw data from Astley (2010a), Carmichael-Olson and Astley (2005), and Coggins (1995).

[a]Neurobehavioral disorder, alcohol unknown with full face of FAS.

Appendix C

Story Elicitation Protocol

Frog, Where Are You? (Mayer, 1969)
Administration Guidelines

1. Assess each child individually. Provide a simple barrier for the child to block your view of the book.
2. Examiner presents child with two envelopes containing copies of *Frog, Where Are You?*
   - "*I have a big box in my office with a bunch of these special books. They are neat because they don't have any words. I want you to pick one AND DON'T LET ME SEE IT* (child picks; put the other book away, out of reach).
3. Examiner says:
   - "*I want you to look at the pictures in the book. You'll want to look at each picture carefully, because when you are done, I want you to tell me the best story that you can.*
   - *Take as much time as you need to look at all the pictures in the book. When you are done looking at all the pictures, turn back to Page 1 and use the pictures to help you tell me the best story that you can.*"
4. After the child looks through the book, confirm that they are ready and have them turn back to the first page of the book.
5. Examiner says, "*Now remember, I want to hear your very best story*."
6. Examiner should be positioned so that it is clear that she or he can't directly see the Frog book.
7. If necessary, the examiner can help the child begin by saying:
   - *"I will help you get started. Once upon a time there was…"*
8. The child should go through each page and tell the story. The child should be turning the pages as she or he tells the story. No specific prompts should be given by the examiner.
9. General encouragement statements are permitted (e.g., *"What happened next?"*). It is okay if the child does not have something to say about each page.
10. After the child finishes telling the story, stop the tape and praise the child for his or her storytelling skills.