EXPRESSIVE LANGUAGE DEVELOPMENT IN TWINS VERSUS SINGLETONS AT EARLY SCHOOL AGE

BY

JULIE K MAHIEU

THESIS

Submitted in partial fulfillment of the requirements for the degree of Master of Arts in Speech and Hearing Science in the Graduate College of the University of Illinois at Urbana-Champaign, 2011

Urbana, Illinois

Master’s Committee:

Assistant Professor Laura S. DeThorne, Chair
Associate Professor Nicoline Ambrose
Associate Professor Julie Hengst
Research Associate Jamie Mahurin-Smith
ABSTRACT

The current study investigated whether twins and singletons differed in their expressive language development at school age. Comparisons focused on language samples and formalized vocabulary assessments gathered from 28 twins and 28 singletons within the Western Reserve Reading Project; the two groups were matched on age, gender, and parent education level. The children’s language samples, taken from conversational interactions, were analyzed for mean length of utterance, number of different words, number of total words, and total number of conjunctions, all of which were converted to z-scores and averaged to form a Conversational Composite. Similarly, children’s scores on the Stanford Binet Vocabulary subtest (Thorndike, Hagen, & Sattler, 1986) and the Boston Naming Test (Goodglass & Kaplan, 2001) were converted into a Formal Language Composite. Mean comparisons via t-test revealed no group differences between twins and on either of the composites. In addition to the Conversational and Formal Composites, a pilot coding process for topic initiation and management was completed for 3 twin-sibling pairs, or a total of 6 children. The topics of modeling clay, school, and hobbies emerged as the most common across transcripts. Measures related to topic management revealed consistent overlap between the two groups. Examiners, in comparison to children, initiated twice as many topic shifts. In sum, results suggest that twinship in and of itself does not indicate a substantial risk for school-age expressive language difficulties, and provide early descriptive data to analyze the transactional nature of topic management within conversational interactions.
ACKNOWLEDGEMENTS

I would like to thank many people for their encouragement and assistance as I completed my thesis. First and foremost, I am thankful to Dr. Laura DeThorne for her support and guidance throughout the project. With her help and expertise I greatly increased my knowledge of language, research, and writing. It was her constant dedication to this project that enabled me to explore a topic of great interest to me. Her positive attitude kept the project continuously enjoyable, even as challenges arose. Dr. DeThorne helped to turn my curiosity into a project that I am deeply proud of, and for that I am extremely grateful.

I would also like to thank the members of my committee: Dr. Nicoline Ambrose, Dr. Julie Hengst, and Dr. Jamie Mahurin-Smith. Their insightful questions and comments at my Prospectus inspired revisions that took my project in exciting new directions. I’d specifically like to thank Dr. Ambrose for sharing her knowledge regarding twin studies, Dr. Hengst for encouraging me to explore discourse-level means of analyzing language, and Dr. Mahurin-Smith for her expertise in the area of prenatal factors.

I am also grateful to the Western Reserve Reading Project for providing the data used to fuel this project. The Western Reserve Reading Project is supported by NICHD (HD38075, HD46167). I appreciate the work of all associated investigators and participating families.

Finally, I’d like to thank my family and friends for their unwavering support throughout this endeavor. I would never have made it here without their love and encouragement.
Table of Contents

List of Tables ........................................................................................................... vi
List of Figures ............................................................................................................. vii
Chapter 1: Literature Review ..................................................................................... 1
  Language Comparison in Twins Versus Singletons ................................................. 1
  Linguistic Input ........................................................................................................ 7
  Obstetric Factors ...................................................................................................... 18
  Dietary Factors ......................................................................................................... 21
  Genetic Effects .......................................................................................................... 23
Chapter 2: Method ..................................................................................................... 26
  Participants .............................................................................................................. 27
  Conversational Language Samples ........................................................................ 30
    MLU ...................................................................................................................... 31
    NDW ................................................................................................................... 32
    NTW .................................................................................................................... 32
    TNC ...................................................................................................................... 32
  Formalized Measures .............................................................................................. 33
  Discourse Measures of Topic Management .......................................................... 33
  Topic-coding ............................................................................................................. 34
  Role in Topic Management ...................................................................................... 35
    Total number of topics ......................................................................................... 36
    Number of topic shifts initiated by the examiner/child ....................................... 36
    Total number of episodes .................................................................................... 36
Chapter 3: Results

Descriptives of Formalized, Conversational, and Topic-coding Measures

Correlation Analyses

Child measures

Examiner measures

Child and examiner measures

Mean Comparisons between Groups

Chapter 4: Discussion

Comparison to Prior Literature

Promise and Potential Pitfalls of Topic-coding

Limitations of the Current Study

Conclusion

Chapter 5: Tables and Figures

Tables

Figures

References

Appendix A: Guidelines for Topic-coding

Appendix B: Sample Page of Topic-coding from Transcript
List Of Tables

Table 1: Descriptive Data on Child Language Sample Measures........................................51
Table 2: Descriptive Data on Examiner Language Sample Measures........................................52
Table 3: List of Topics and Frequencies of Topics Across Episodes for Topic-Coding.................................................................53
Table 4: Results of Topic-Coding Measures for Twins and Singletons.................................54
Table 5: Bivariate Correlations for Standardized and Conversational Language Measures in Singletons......................................................55
Table 6: Bivariate Correlations for Standardized and Conversational Language Measures in Twins............................................................................56
Table 7: Bivariate Correlations for Conversational Language Measures in Singletons’ Examiners .......................................................................................57
Table 8: Bivariate Correlations for Conversational Language Measures in Twins’ Examiners.................................................................................................58
Table 9: Bivariate Correlations for Conversational Language Measures of Singletons and Examiners......................................................................................59
Table 10: Bivariate Correlations for Conversational Language Measures for Twins and Examiners.........................................................................................60
Table 11: Descriptive Data on Children’s Z-Scores, Divided into Twin and Singleton Groups........................................................................................................61
List Of Figures

*Figure 1:* Scatter Plots of Children’s Formal Composite Scores, With Twins
   Born Before 35 Weeks Highlighted………………………………………...62
*Figure 2:* Scatter Plots of Children’s Conversational Composite Scores, With Twins
   Born Before 35 Weeks Highlighted………………………………………...63
*Figure 3:* Scatter Plots of Children’s Formal Composite Scores, With Twins
   With Reported History of Twin Language Highlighted………………………64
*Figure 4:* Scatter Plots of Children’s Conversational Composite Scores, With Twins
   With Reported History of Twin Language Highlighted………………………65
Chapter I: Literature Review

The purpose of the present study is to examine potential differences in twins’ expressive language ability at school-age as compared to singletons. Previous studies of twins’ language development have focused primarily on younger ages, with differences interpreted in favor of singletons. This chapter will highlight previous research on language differences, including proposed causal mechanisms for potential group differences, with a focus on linguistic input, obstetric influences, dietary factors, and genetic effects. The literature reviewed here spans nearly a century of twin-singleton research. In those 100 years, different theoretical accounts of language acquisition have waxed and waned in favor and advances in medical technology, neuroimaging, and genomics have increased the interest and emphasis on biological influences on language development. As a result, the way that researchers designed, implemented, and interpreted studies of twin-singleton language differences has changed over time.

Much of the earliest research focused on documenting the presence of group differences between twins and singletons with predicted differences explained almost exclusively through differences in linguistic input. Later studies have expanded the list of environmental factors under consideration to include prematurity, and most recently potential early differences in nutrition. Regardless of causal orientation, there has been a consistent tendency to predict and interpret group differences in favor of singletons.

Language Comparison in Twins Versus Singletons

I first review those studies that focused exclusively on potential group differences between twins and singletons without explicit hypotheses about why such differences might exist. These studies, particularly the earlier ones, often lacked methodological
details that are considered of high importance today, such as clear reference to normative
data and controls for gestational age and perinatal history.

An early and highly cited study of twin-singleton language differences, Day (1932), sought to compare the language skills of young twins to those of age-matched
typically developing singletons. The study included 80 pairs of twins, with 20 pairs at
each of the following ages: 2, 3, 4, and 5 years. The twins were matched with singletons
based on age, sex, and parent occupation level. Information regarding perinatal history
and socioeconomic status was not reported. All children were administered the
Minnesota Preschool Scale to measure intelligence, which was administered at the time
of the home observation, as well as a sample of 50 utterances taken during an
independent child free play. All reported results were limited in presentation to bar
graphs, without clear reference to descriptive or statistical comparisons. In terms of IQ,
twins scored approximately 10 points behind the singletons, with twins averaging a score
of 94 and singletons a score of 103. The children’s language samples were compared in
terms of mean length of response, number of in comprehensible responses, functional
completeness, naming, and use of simple or incomplete sentences. The authors report that
twins were lower than singletons in mean length of response at age 2, with the
discrepancy increasing with age. By 54 months the twins averaged slightly more than 3
word utterances and singletons averaged approximately 4.5 word utterances. The author
found that at age 5 the twins averaged utterance lengths equivalent to those recorded in
singletons at age 3, which she described as a 2-year delay. In addition, the authors report
that the singletons’ speech, on average, contained more in comprehensible responses than
that of twins. When comparing the two groups for percentages of use of the various parts
of speech, the author reported similar percentages for both the twins and singletons. The author stated that the amount of naming that the singletons used in their speech decreased with age while it remained the same for twins, which she suggested could be a persistence of immature speech in twins. Additionally, the author reported that although twins produced more comments about the immediate situation than singletons, singletons commented more on events associated with the immediate situation than twins, with comments about associated situations reportedly not appearing at all in the twins’ speech until age 3. On average, twins were also reported to produce twice as many emotionally toned utterances as singletons, which the author also attributed to immature speech patterns. Twins were reported to increase in the number of questions asked with age, but even by age 5 did not reportedly ask as many questions as did singletons at the age of 2. Despite such report, the bar graphs presented for the two groups appeared to suggest that singletons’ average number of questions decreased over time, and that the number of questions asked by both groups at age 4 was similar. Finally, the author analyzed the number of different words produced by the children. At 5 years the twins had 158 different words, equivalent to the number of different words the singletons had at 3 years.

In addition to concerns regarding ambiguity and inconsistency in the data presented by Day (1932), the lack of normative data made it difficult to determine whether group differences represented actual delays on the part of twins or whether the singletons were relatively precocious. Furthermore, the author herself reported various validity concerns, such as inconsistencies in the home-play evaluation setups across families. Particularly, it was not always possible to see the twins one at a time, so the twins were occasionally both seen at once. In essence, the examiner was observing the
singletons and twins in different contexts, which could potentially account for some of the reported differences.

Similar to Day (1932), Mittler (1970) compared the psycholinguistic skills of 200 twins at age 4 with that of 100 singletons of a similar age. All twins were seen within one month of turning 4, and singletons were reported to be seen “as close as possible” to turning 4 (p. 743). Because age was the only selection criterion for participation, no exclusions or controls were made for children with histories of speech-language disorders, perinatal complications, or developmental delays. Twins were recruited from British public health records and singletons were recruited from local daycares. The study reported a strong bias toward professional and middle class families, which was true for both groups but more pronounced in the singleton group.

The language skills of the subjects were assessed using the Illinois Test of Psycholinguistic Abilities (ITPA), which includes categories of communication, organization, and processing. The twins scored what the authors interpreted as an average of 6 developmental months behind the singletons on the ITPA, with twins averaging a language age of 42 months and singletons a language age of 48 months. On the ITPA, the singletons were reported to score “exactly with that of the American standardization sample of the same age, despite the middle-class bias of the British sample, and the possible effect of socio-cultural and psycholinguistic differences between the two populations” (Mittler, 1970, p.12). The twins’ scores followed an overall pattern similar to those of singletons, but averaged 2/3 of a standard deviation below the singletons’ scores. This was a statistically significant difference on all but one of the six subtests.
In addition to the ITPA, the children were given two nonverbal assessments: the Seguin Form Board Test (SFB) and the Goodenough Draw-a-Man (DAM) test. In addition, the Peabody Picture Vocabulary Test was administered (PPVT). The twins scored 12 months ahead of the published norms on the SFB and 10 months ahead of the norms on the DAM. However, the author noted that the strong bias of the study toward the middle class or potentially outdated norms could have caused these results. The twins scored only slightly behind the singletons on the SFB. However, the author reported that the drawing section of the SFB was complicated by the fact that less than half of the twins completed this portion of the test compared to 2/3 of the singletons. On the PPVT, twins scored an average of 6 months behind the singletons for ‘mental age.’ The children’s scores on the PPVT were reported as a group comparison, but not reported in relation to the norms. The authors did not find any significant differences in psycholinguistic abilities when analyzing the results based on zygosity or sex.

In addition to Day (1932) and Mittler (1970), a third more recent study, Levy, Hay, McLaughlin, Wood, & Waldman (1996), focused largely on group differences between twins and singletons. Although language development was not an explicit outcome in Levy et al., I include the study here due to its focus on the associated phenotypes of speech and reading disability. Levy et al. utilized parent report data from 1,938 families recruited from the Australian MHMRC Twin Registry to compare the behaviors of twins to their non-twin siblings within the same families. The authors relied on a parent-report measure, designed explicitly for the study, that included questions about the children’s behavioral history and speech and reading development. Mothers were asked to fill out the questionnaire for any of their non-twin children within the
target age range of 4-12 years. Twins averaged ~7.85 years in age and singleton siblings averaged ~8.38 years. Because twins and singletons were recruited from the same families, socioeconomic differences between groups were inherently controlled.

Multivariate analyses found that twins differed significantly from their siblings on ADHD symptoms, gestational age, birth weight, and speech and reading symptoms, with findings consistently favoring non-twin siblings. Of the twins, 10.6% met the criterion for ADHD while only 6.8% of their siblings qualified. For both the speech and reading measures the effect of twin/sibling status was found to be significant. The highest prevalence of both speech and reading problems and ADHD symptoms were found in male twins.

Although methodologically flawed, Day (1932) and Mittler (1970) both reported group differences between twins’ and singletons’ language abilities, which favored the singletons in most cases. Similarly, Levy et al. found that twins were more likely to display difficulties in related areas, such as attention, speech, and reading. Of particular interest in the earlier two studies especially was the tendency to interpret any emergent group differences as negative for twins, even in cases where the twins’ scores were higher than those of the singletons. This observation underscored the need to develop more explicit hypotheses related to why and how groups might be expected to differ from each other. I turn now to more recent studies that have focused not only on examining group differences but also on understanding why they might exist. Three primary factors have been proposed as key factors in the potential differences between twins and singletons, each of which I review in the text that follows: linguistic input, obstetric complications, and early diet.
Linguistic Input

Without question, the bulk of research examining twin-singleton language differences has focused on hypothesized differences in linguistic input, most often from the children’s mother. The focus on linguistic input may be explained in part by the popularity of social-interactionist and behaviorist theories. Though different in many regards, both theories stress the role of the primary caregiver in children’s early language development. In particular, emphasis has centered both on the quantity and quality of parent-child interactions, with general consensus that more language input geared somewhat above the child’s current level of language ability is most advantageous (Snow, 1972; Ellis & Wells, 1980).

The related argument then is that twins, in comparison to singletons, are likely to receive less individualized language input from caregivers. The specific rationale differs somewhat across studies, with some focusing on the proposed tendency for caregivers to address twins as a unit rather than individuals being disadvantageous, and others suggesting that the stress of caring for multiples leads to language that is focused more on environmental control than social interaction (Cox et al., 1987; Rutter et al 2003, Thorpe, Golding, MacGillivray & Greenwood, 1991). In addition there has been some suggestion that twins themselves, due to having a constant playmate, may less readily seek out other sources of more mature linguistic input (Luria & Ludovich, 1959).

Of interest, many studies of the linguistic input to twins attempted to recreate at least part of the dynamics of twinship in their singleton comparison group by including sibling dyads close in age, thereby making it less clear what group differences were expected and why. For example, Rutter, Thorpe, Greenwood, Northstone, & Golding
(2003) compared the language abilities of 96 twin pairs and 98 singleton pairs at both 20 and 36 months. Both the twin pairs and singletons were selected from the Avon Longitudinal Study of Parents and Children (ALSPAC). Children born earlier than 33 weeks gestation or found to have neurological deficits were excluded. Singletons were selected to match a twin pair in terms of gestational age and to have an older sibling close in age (i.e., within 30 months of age). Twins and singletons did not differ on mother’s vocabulary, parent occupation, or parent education level. Outcome measures included the MacArthur Communicative Development Inventory (MCDI), a parent report measure that estimates the child’s language abilities. The children were also given two standardized tests, the Preschool Language Scale (PLS-3) (Zimmerman, 1992) and the McCarthy Scales of Children’s Abilities. The authors reported that twins’ language was delayed 1.7 months behind the singletons at 20 months and 3.1 months behind the singletons at 36 months. Despite the fact that children born prior to 33-weeks gestation were not included in the study, perinatal histories, including birth weight and gestation, were considered and discarded through correlation analyses as possible causal factors for the twins’ reported delays (Rutter et. al, 2003).

As a means to explore linguistic input as a causal influence, twins were compared to singletons based on mother-child interaction at both 20 and 36 months. One twin within the pair was randomly selected for observation, as was the younger sibling in the singleton pair who was matched to the twin in terms of age and gender. At 20 months, the children were compared using the Caldwell HOME inventory. The HOME inventory is a combination of both mother report and home observation. On the overall HOME inventory scores, the mothers of twins scored significantly lower than the mothers of
singleton. Although the patterns of interaction between the mothers and their children were reported to be comparable for twins and singletons (minus the finding that the mothers of twins were more likely to address their children as a unit rather than individuals), the style and quality of interaction used by the two groups of mothers differed. Specifically, results of the HOME inventory suggested that during a free-play time with toys the mothers of twins were less likely to “provide strong motivation” to their children, although what the authors meant by strong motivation was never clarified (Rutter et al., 2003, p. 346). During the book observation time, in which the mothers were observed reading a book to the two children, the authors found that the mothers of twins were less likely to elaborate on different pictures, invite the child to comment, appear comfortable reading a book to the children, and report regular book reading to their children than the mothers of singletons. The authors make note that interaction differences observed between the groups of mothers could be due to the fact that the non-twin sibling pair had an older child that was potentially evoking more advanced interaction styles from the mother, resulting in differing maternal input (Rutter et al., 2003).

At 36 months, the authors found that the group differences on the HOME inventory were much smaller, with the singleton pair scoring only one half of a standard deviation higher than the twins. Additionally, there were no longer group differences in the mother-child book observation time. Like at 20 months, mothers of twins at 36 months were less likely to spend individualized time interacting with each twin during the mother-child interaction time. The authors state that the minimization of group differences between 20 and 36 months suggest that the earlier differences were due to
something specific to interacting with two young twins rather than differences in parenting style.

At 36 months the two home observations were repeated for 10 minutes each. The authors then used the family interaction information from the 20-month observations to determine whether they predicted language outcomes at 36 months. The authors found that the mothers’ tendency at 20 months to address the twins as a pair rather than individuals did not predict language outcomes at 36 months. The authors did find that the other mother-child interaction variables, such as encouraging speech from the child and engaging in the mutual reading of a storybook, were associated with language outcomes, and thus are possible causes for the group differences in language outcomes at 36 months. All variables were combined into a maternal input composite score to test this hypothesis. The composite correlated significantly with child measures: .37 with vocabulary at 20 months, .50 with language at 36 months. It was found that twin-singleton language differences at 36 months were no longer statistically significant when maternal interaction variables were controlled, with the difference falling from just over 3 months difference to .22 months difference. The authors interpreted this to mean that the different parent-child interaction styles observed at 20 months were strongly influencing language outcomes at 36 months (Rutter et. al, 2003); the authors did acknowledge, however, the possibility of child effects, whereby the children might have been eliciting differences in the interaction.

In a similar design, Stafford (1987) compared the role of maternal input on the language development of preschool age twins and singletons. The study included 44 total mothers, half of whom were mothers of twins and the other half mothers of singletons
with older siblings. The twins and singletons selected were between 24 and 36 months of age and described as middle class, well educated, and homogenous. The average age of the children was 28 months. The younger of the two singleton siblings was matched within two weeks of the birth date of a corresponding twin pair. There were 88 total children in the study, with 44 total target children. The target children included a randomly selected twin and the younger singleton sibling, who was matched with the twin based on age and sex. Twin pairs and singleton-sibling pairs were compared on parent education, mother’s work status, and parent income. A chi-squared analysis did not yield any statistically significant differences between the two groups. The twins were reported to be an average of 2.6 weeks premature, while the singletons were reported to be an average of 0.2 weeks premature, a statistically significant difference.

The twins’ and singletons’ expressive and receptive language abilities were analyzed using the Minnesota Child Development Inventory (MCDI). The MCDI is a parent report measure that estimates a child’s general abilities, including language. No statistically significant group differences in language ability were found. For the receptive measures, singletons scored an average of 3.1 months ahead of the norms for the test, while the twins scored .59 months ahead of the norms. For the expressive measures, singletons scored 3.1 months ahead of the norms while twins scored .64 months behind the norms. As such, the twins’ scores were a couple months behind those of the singletons on language measures. However, the twins’ scores were still within the typically developing range. It was reported that early birth was not associated with expressive or receptive language when assessed via correlation.
Despite the lack of substantial group differences, the authors sought to compare language input of parents of twins to that of parents of singletons. Due to her statement that twins have been “identified historically as having delayed speech,” the author hypothesized that mothers of twins might demonstrate different interaction styles with their children than mothers of singletons, which could be the cause of language differences between the two groups (p. 430, Stafford, 1987).

The two singleton siblings were observed interacting with their mother during the language sample in order to rule out the possibility that any observed group differences could be due to interacting with two children at once rather than interacting with a twin pair. However, it is not possible to recreate the actual situation of the mother of twins, since the mothers of twins are interacting with two children of the same age. The researchers transcribed 5 minutes of the playtime and 5 minutes of the snack time for analysis. The mothers’ interactions were then scored based on three categories: discourse categories, illocutionary categories, and conversational style. Discourse categories included behaviors such as imitations, expansions, and extensions of the children’s utterances. Illocutionary categories included behaviors such as occurrence of commands, repairs, questions, and acknowledgements. Conversational style involved behaviors including number of parent utterances, number of utterances directed at an individual child versus number of utterances directed to the two children simultaneously, and ratio of maternal utterances to child utterances.

Results of the study found that mothers of singletons were more likely to use imitations, expansions, extensions, items related to actions, and maternal self-repetitions when interacting with their children. Mothers of singletons also produced a statistically
significantly greater number of questions and positive acknowledgments than mothers of twins. Mothers of twins were more likely to produce utterances directed at both children simultaneously rather than individually. In addition, mothers of twins were more likely to use commands than mothers of singletons. The author interprets previous literature about interaction styles to suggest that use of commands represents a “desire to control” rather than lack “interest in talking with the child” (p. 455, Stafford, 1987). The author interprets the overall findings of her study to suggest that twins, who scored ahead of standardized norms despite being slightly delayed from singletons, are receiving a “less conversationally responsive linguistic environment” (p. 456). However, it is acknowledged that this conclusion is far reaching, as the correlation design of the study prohibits the ability to determine cause and effect. As noted in relation to Rutter et al. (2003), it is possible that the differences between parent styles noted in Stafford (1987) are the result of varying levels of language abilities in the children rather than the cause of them. For instance, an older child who is presumably more linguistically advanced than a younger child may evoke different language interactions from his or her parent, which could influence the interaction style that the parent uses with the older child and could extend to a younger sibling that is present for the interaction.

A longitudinal study by Lytton, Conway, and Suavé (1977) also examined whether twinship affected the way parents interacted with their children. The participants included 46 male twin pairs, of which 17 were monozygotic (MZ) and 29 were dizygotic (DZ), and 44 singleton males. The children were within 25-35 months of age (mean 32.4 months). The twins were recruited from local hospital records and the singletons were recruited from Child Health Clinics. Singletons were excluded from the study if they did
not come from a two-parent home. They were also excluded if they did not have a sibling within three years of age, either older or younger, in order to better emulate the twin situation of two young siblings being raised in the same home. The twins were matched with singletons based on age, father’s profession, and mother’s education. Chi-square analyses produced no significant differences in these areas.

Examiners administered the Peabody Picture Vocabulary Test to target children and informally observed them interacting with their parents and siblings over two 3-hour sessions. It was reported that a second sibling was almost always present during the singleton observations, and a third sibling was often present during the twin observations. Trained examiners coded child and parent behaviors, including the frequency with which the child sought a parent for comfort and frequency of use of speech per minute. Additionally, the mothers filled out a questionnaire concerning their perceptions of their children’s behaviors and their own parenting practices.

After conducting a multiple regression analysis to remove the potential influence of parent education level on outcomes, the following statistically significant group differences emerged: twins spoke less frequently than singletons, were spoken to less often by their parents than the singletons, and used “less mature” speech than the singletons (p. 103, Lytton et. al, 1977). This description of twins using “less mature” speech was not well defined, and appeared to be a qualitative judgment made by the examiners. The results for the PPVT favored twins when only children whose mothers attended college were considered, and favored singletons when only children of parents who did not attend college were considered. However, this difference in scores was not statistically significant. The authors concluded that the parents’ behaviors influenced
group language differences, in that the pressure of interacting with two children of the same age at once, as in the twin situation, reduced the overall input to both children.

In a follow-up study in 1987, Lytton, Watts, & Dunn looked at the same twins at 9 years of age. Of the original 46 male twin pairs, 35 agreed to participate in the follow-up study. The singletons in the follow-up study were newly selected from the twins’ classrooms, so that the same classroom teachers of the twins could also assess the singletons. In addition to classroom placement, the new singleton group was matched with the twins for age and sex. Because the twins came from the same school, they were considered to be of approximately similar social classes. However, explicit data comparing the two groups in this regard were not presented. Instead, the mothers’ education levels for the new singleton group were compared to those in the original singleton group and the two groups were not found to be significantly different. When analyzing the twins who chose to continue in the study in comparison to those who dropped out, the continuing group was found to have been more mature in speech at age 2 and to have higher mothers’ education levels. Results of group comparisons on birth history are reported below.

The classroom teachers of the twins and singletons were given a 28 question, 5-point Child Rating Scale to complete. The questions were selected to mimic those that were answered about the children at age 2. Teachers also rated the students on intellectual skills and accomplishments at school, with instructions to rate the students in comparison with the expected performance and behaviors of the entire class unit. The teachers were also asked to rate both the twins and the singletons during the same time period to avoid bias and remain consistent. None of the questions on the Child Rating Scales found
significant group differences. However, contrary to expectations, twins’ relationships
with peers were rated slightly higher than those of singletons.

In addition to the teacher report, children’s abilities were assessed via The
Peabody Individual Achievement Test, the Crichton Vocabulary Scale (CVS), and the
Raven’s Colored Progressive Matrices. The twins were reported to score significantly
lower on verbal intelligence than the singletons, with twins averaging a score of 63.5 and
singletons a score of 74.9. However, this was almost entirely due to the MZ twins’
average score of 45.7, which was significantly lower than the mean score of 74.0 attained
by the DZ twins. The DZ twins’ scores on verbal intelligence did not differ significantly
from the singleton group. The same trend was true for the groups in cognitive measures.
Group differences between twins and singletons on nonverbal abilities were not found.
An orthogonal covariate analysis that controlled for the covariates of mother’s education
level, birthweight, and prenatal stress, revealed that a “far greater” amount of twin-
singleton variance was explained by parent education than the twin factor. Similar results
were found for the PIAT math score.

The authors hypothesized that the shift from the findings at 2 years, where group
differences were associated with input variables, could be caused by a reduction in the
“near exclusive role” that parents play in the toddlers’ cognitive development, in
comparison with the involvement of many additional parties as the children mature (p.
367, Lytton et. al, 1987). However, this conclusion appears to interpret the results in
order to be consistent with the initial hypothesis, while failing to consider all of their
data, such as why MZ twins’ scores were substantially lower than DZ twins’ scores and
why twins’ were reportedly experiencing stronger peer relationships.
While most hypotheses surrounding the linguistic input to twins appears to center on the quantitative and qualitative differences in input (i.e., fewer opportunities for quiet one-on-one interaction), some investigators have centered instead on the potentially deleterious effects of children close in age developing their own language. A proposed cause of language impairment in twins versus singletons is the use of an independently created language between the twins, which is often unintelligible to outsiders. Often coined as “twin language,” some research investigators have hypothesized that use of a private language might result in a language delay among twins. It has been suggested that because these twins have a co-twin of the same age who understands them and often serves as a companion, twins do not find as much need as singleton children to initiate communication with peers and practice the language of the wider community. Of interest, this hypothesis appears at odds with the emergent consensus that exposure to more than one language does not lead to language delays, at least not in the singleton population (Holowka, Brosseau-Lapre, & Petitto, 2002; Winsler, Diaz, Espinosa, & Rodriguez, 1999).

In particular regard to twin language, the previously mentioned study by Mittler (1970), interviewed 200 mothers of four-year-old twins regarding their children’s prior use of a twin language. Nearly half of the twins were reported to use a “twin language”; however, discriminant function analyses did not find this to be associated with the children’s scores on the ITPA.

A study by Bishop and Bishop (1998) explicitly examined the role of reported “twin language” use in children’s development of future speech and language problems. The study consisted of two groups of participating twins. Sample G contained 94 twin
pairs between the ages of 7 to 13 who were selected to represent a sample of the general twin population. Sample L consisted of 82 twin pairs between the ages of 7 and 13, in which at least one of the two twins had a speech language impairment. The authors compared the language abilities of the twins at school-age based on parent report of whether or not the twins were reported to have used a “twin language” as toddlers. The authors found that parents of twins with language impairment were significantly more likely to report a previous use of “twin language” between their children than parents of twins with typically developing language abilities (50% vs. 11%). Although such results align with the idea that children’s use of a twin language may impede their language development, the authors suggested an alternative interpretation: perhaps what is considered by caregivers as a “twin language” may actually be similarly deviant speech patterns used between two twins at a similar developmental level (see also Dodd & McEvoy, 1994). Of interest, this interpretation is supported by Lastres and DeThorne (2009), a study that found more similar speech errors in MZ pairs concordant for speech difficulties than in concordant DZ pairs.

In sum, studies of linguistic input as a cause of language delay in twins have led to mixed findings that are often limited by unspecified rationales and correlational data. Similar to other studies of twin-singleton differences, interpretations often appeared colored by the tendency to view any twin-singleton differences as negative. In addition to these concerns, other potential causal factors, such as obstetric complications, often remained unexplored.

**Obstetric Factors**
Medical advances have not only increased the survival rates of infants born with medical complications, but have also altered the means in which many women conceive their children. The increasing number of women seeking alternative routes to pregnancy, including fertility drugs and in vitro fertilization, has greatly raised the prevalence of multiple births. According to the Center for Disease Control’s 2007 report on annual births, the incidence of twin births in the United States between 1970-2004 increased by 70%. This number remained stable between 2004-2007, where approximately 30 out of every 1000 live births consisted of twin pairs.

In addition to, or perhaps as a result of, the increasing number of multiple births, the number of premature births is also on the rise. The same Center for Disease Control 2007 report on annual births reported that the number of premature births between 1990-2007 rose around 1% per year, with a total increase of 20%. Additionally, the number of infants born before 39 weeks gestation is steadily rising, with a 15% increase since 1990 (52% vs. 67%)(Martin, 2010). According to the Center for Disease Control’s 2008 report on annual births, 58.63% of twins born in the United States complete fewer than 37 weeks gestation (of which 11.6% were born before 32 weeks), compared with 10.63% of singletons that complete fewer than 37 weeks gestation (of which 1.6% were born before 32 weeks).

With increases in premature births and the survival rates came enhanced concerns with the prognosis for children born with medical complications and the various pre- and peri-natal factors that affect outcomes. Such factors have particular relevance for twins, who on average are born several weeks earlier than singletons and consequently have lower birth-weights. Lower birth-weights are associated with greater risk for later health
problems (Escobar, 1991) and developmental challenges, including language (Hille, Den Ouden, Bauer, van den Oudenrijn, Brand, & Verloove-Vanhorick, 1994; Luoma, Herrgard, Martikainen, & Ahonen, 1998). To date only a couple studies have explicitly considered the impact of obstetric factors on twins’ language development.

In Rutter et al.’s (2003) previously mentioned twin study, possible causal factors for “mild language delay” in twins were examined. The twin pairs and singletons, who were studied at both 20 and 36 months, were compared on a variety of obstetric measures including birth weight, gestation age, and optimality indices. The optimality indices gave each child a composite score based on obstetric and perinatal complications. The authors found that obstetric features did not account for the slightly lower language development in the twin pairs; however the full range of prematurity was restricted by their exclusion of twin pairs born prior to 33 weeks gestation. Similarly, the aforementioned study by Stafford (1987) did not find associations between prematurity and language abilities.

Another previously reviewed study, Mittler (1970), offered somewhat mixed results related to pre- and perinatal factors. Interviews with questions surrounding the children’s obstetric histories were obtained from the mothers of the twins, while a short questionnaire was obtained from the parents of singletons. Of the participating mothers of twins, over 50% reported pregnancy abnormalities, with 43% reported birth anomalies. Once again these abnormalities were not found to result in group differences when compared with twins whose mothers did not experience birth abnormalities. The twins averaged a birth weight of 5 pounds 10 ounces compared to 7 pounds 6 ounces for singletons. Lower birth weights were associated with lower test scores, but the
association did not reach statistical significance. Length of gestation was, however, reported to be significantly associated with ITPA scores.

In sum, few studies considered the role of obstetric factors in explaining twin-singleton language differences, and none that I found explicitly set out to examine it. However, given evidence of long-term consequences of obstetric complications (Escobar, 1991), it seems critical to examine or control for such factors when comparing language outcomes in twins versus singletons.

**Dietary Factors**

Perhaps most recent on the horizon of research in twin versus singleton development is consideration of potential differences in early diet, particularly the prevalence and longevity of breastfeeding. Breast milk is thought to contain important antibodies that help to protect infants from infections, as well as long-chain fatty acids that are important for infant brain development (Gartner et al., 2005).

A meta-analysis by Anderson et al (1999) combined the research of 20 previous studies concerning breastfeeding’s influence on cognitive development. The results showed that children who were breastfed scored an average of 5.32 points higher in cognitive function than children who were not breastfed. After taking into account possible confounding variables such as socioeconomic status, birth weight, and gestational age, the study found the children who were breastfed scored a statistically significant 3.16 points higher in cognitive function than children who were not breastfed.

In light of the potential cognitive benefits associated with breastfeeding, it seems relevant to note the particular challenges mothers of twins face with early feeding, which are twofold (pun intended). First, mothers of twins are faced with the logistical challenge
of breast-feeding two infants simultaneously, which is likely to decrease the prevalence and longevity of breast-feeding in twins relative to singletons. In addition, it is feasible that mothers of twins who do breastfeed are providing their infants with qualitatively different milk than mothers who breastfeed singletons, as their bodies must produce enough milk to sustain two infants, which challenges the milk supply but may also influence the constituents of the milk produced. The role of early diet in understanding twin versus singleton differences may be particularly germane for children born before 2001, when infant formula in the United States was not yet supplemented with the long-chain polyunsaturated fatty acids (LCPUFA) thought to be important for brain development (cf. Willatts et al., 1998; see Beyerlein et al., 2010 for conflicting evidence).

Unfortunately, only one study to date considered the potential role of early breast-feeding differences on developmental outcomes in twins. Rutter et al. (2003) noted that on average the singleton children in their study were breastfed for a longer duration than twin pairs (51% vs. 31% breastfed for 6 months or more). The authors found that singletons who were breast-fed were significantly more likely to have higher language scores at 20 and 36 months, which was not true for the twins at either age. Consequentially, breastfeeding duration was considered a possible contributing factor involved to the reported language delay found in twins versus singletons. After comparing the twins and singletons based on the duration of time they were breastfed by their mothers, the authors concluded that it was not likely that longer duration of breastfeeding accounted for language differences between twins and singletons.
However, the study failed to consider possible differences in the nutritional make-up of breast milk between the two groups.

**Genetic Effects**

Although environmental factors ranging from linguistic input to early diet have been considered as explanation for twin-singleton language differences, it is important to acknowledge the role that genetics play in language abilities. A meta-analysis by Stromswold (2001) analyzed over 100 twin, genetic linkage, and adoption studies pertaining to language development. The results of the review revealed that there is a strong genetic component for both typical and impaired language. Of particular relevance to the study at hand, DeThorne et al. (2008) examined the genetic and environmental influences on both formal and conversation-based measures of child language within 380 school-age twins (78 MZ pairs, 112 DZ pairs) from the Western Reserve Reading Project. Results indicated significant heritability for both conversational and formal language factors, with estimates of .70 and .45 respectively. Although twin studies are inherently well-positioned to study causal effects, concerns regarding the extent to which findings from twin studies can be generalized to singletons are often raised (Evans & Martin, 2000), particularly due to reports of language delay in twins relative to singletons.

Because genetics play an important role in language abilities, it is possible, as in the case of “twin language,” that twins, who appear delayed because of some factor implicit to being a twin, are in fact both delayed due to their highly shared genetics. The dual-presence of a language delay may be more noticeable, and become more anecdotal.
memorable, when it appears in twins, thereby fueling more conjecture regarding the language challenges of twins.

Although the previous literature comparing twins’ and singletons’ language abilities suggests that twins may be delayed in language development relative to singletons, the prior studies focused primarily on children in the preschool years and suffered from several methodological limitations. Much of the previous research lacked proper controls for potential confounding factors such as age, socioeconomic status (SES), and perinatal history. In addition, studies of discourse-related measures often suffered from underspecified hypotheses, unclear procedures, and conditions that were not easily comparable between the twin and singleton groups.

The purpose of the present study was to examine whether twins and singletons differ in expressive language abilities during school age, while controlling for potential confounding variables such as age, assessment context, gender, and SES. In addition, the full range of variation in birth history, early diet, and use of twin language were included to examine the potential role of these factors in observed group differences. Finally, I piloted an assessment of topic-maintenance use in small subgroup of twins and singletons to help direct future analyses of potential group differences in discourse. The specific research questions were as follows:

1. Do twins differ in their expressive language skills compared to singletons at early school age?

2. If yes, are group differences greater for conversational measures vs. formalized test tasks?

3. And if differences are present, do they seem to be due to perinatal risk
factors or reported twin language use?

4. Finally, what measures related to topic management appear promising as meaningful descriptors of discourse in school-age children?
Chapter II: Method

The data used in the current study came from the Western Reserve Reading Project (WRRP), a longitudinal twin study that focuses on the development of reading, math, and related skills. Twins lived primarily in the cities of Columbus, Cleveland, and Cincinnati and were recruited via Ohio State birth records, school nominations, and media advertisements. Data for the project at large was collected via annual home visits and questionnaire data, which was initially completed after twins entered kindergarten but before they completed first grade. Assessments focused primarily on the development of reading; however, conversational language samples were also collected during the second and third annual home visits. The data for the present study focused on twin data from the third home visit, during which twins were in either second or third grade, depending on the time at which they originally enrolled in the project.

To serve as an explicit comparison for the twins’ conversational language data at the third home visit, singleton siblings within the age range associated with HV3 (i.e., 7-10 yrs) were recruited for the present study while their older twin siblings were participating in later home visits. When both the parent and singleton sibling provided consent, a third examiner was sent to the home visit for assessment of the singleton sibling and the child was given a $5 gift card as a token of appreciation. Whereas all twins, including those in the present study, completed a two-hour test battery related primarily to reading, the singletons were only asked to complete the language measures, which took approximately an hour. The current study focused on language data collected from the singletons in comparison to twin data from HV3. Note that the singletons for the
present study were recruited from families with twins in the larger study, but they were not siblings of the specific twins included in the present study.

**Participants**

At the time of the present study, 38 singleton siblings had participated in the language assessment. Of those 38, one child was excluded because his language sample, although conducted, was not recorded due to an examiner error. Of the 37 remaining singletons, 9 did not fall within the target age range of 7-10 years and were not included in the study. After taking these factors into account, the final singleton group consisted of 28 children (43% male, 57% female) with a mean age of 8.32 years (range: 6.33-9.83 years, $SD= .95$). Because the singletons were siblings of twins from WRRP, and their language samples were taken later on in the project, it necessitates that the singleton population had at least 2 older siblings and were consequently not first-born children. For the 26 singletons for which birth information was available, all 26 were the latest born in their families. In contrast, the twin group had varied birth orders, with 14 of the 28 twins born from their mothers’ first pregnancy, and the rest falling between the 2$^{nd}$-5$^{th}$ pregnancies.

Twins from the larger WRRP database were individually selected to match singletons in terms of gender, race, age, and parent education. Twin zygosity was not considered during the matching process, in order to allow the twins’ zygosity to more freely resemble the greater WRRP database. The process began by searching the twin database for all available twins that were an exact match to a given singleton in terms of (a) race, (b) parental education, and (c) gender. Of the twins that emerged from this search, the closest match in age (given in total months) was then selected. There were
three cases in which more than one exact age-match was available. In such cases, all possible matches were written down on slips of paper and one was drawn to be the twin participant.

There were 16 cases in which an exact match in terms of months of age was not possible. In such cases, the twin with the next closest age match was selected. In the 4 cases in which the closest age match could have been either older or younger, the selection alternated between older versus younger matches to keep group means most similar. In all cases, age was matched within four months.

In the two cases where an age match within four months could not be made while also matching for age, race, parent education, and gender, the twin search was widened one level in terms of parental education. This resulted in one pair with uneven parent education level in which the singleton level was 6, corresponding to graduation from a four year college, and the twin level was 7, corresponding to attendance of some graduate or professional school without graduating. In one case, age and parental education could not be matched within the given parameters. For this final case, the search was widened to include all variations in child race, and a Caucasian twin was paired with an African American singleton. In this case, gender and parent education matched exactly, with age differing by one month.

The matched twin group had a mean age of 8.30 (Range: 6.42 to 9.75, $SD = .94$), within .02 years of the mean age of the singleton group. The twin group consisted of 28 children (43% male, 57% female). The race of the children was based on information provided by participating parents related to each child in the family. The final sample of twins consisted of 27 Caucasian children and 1 Asian child, differing from the singleton
sample only in presence of one more Caucasian and one fewer African American. Parent education level was based on self-report. All parents in both groups had high school diplomas or the equivalent, with the majority having received a 4-year college degree. Categorized by highest level of education, the primary caregivers of the twin sample included 1 with a high school diploma, 1 with some college, 4 with 2 years of college, 10 with 4 years of college, 3 with some graduate/professional training, and 9 with completed graduate or professional degrees. The singleton group differed only by one more parent with some college and one fewer parent with some graduate training.

Of the 27 twins for which zygosity information was available, 12 were reported to be monozygotic (44%), 14 were reported to be dizygotic (52%), and 1 was undetermined (4%). In the vast majority of cases, zygosity was confirmed by DNA testing obtained via buccal swab. Information on developmental history obtained via questionnaire was available for most twins but not singletons. Specifically, results from the Speech-Language Survey (DeThorne et al, 2006), were available for 25 twins, in which parents indicated through a series of yes/no responses if their children had persistent or recovered challenges in various domains of speech-language development. Of those 25 twins, 23 (92%) were reported to have had no reported history of expressive language difficulty and 2 (8%) were reported to have a resolved expressive language difficulty. Similarly, 19 (76%) of the twins had never seen a speech-language pathologist, 5 (20%) reportedly had seen a speech-language pathologist at one time but not at the time of entrance in the study, and 1 (4%) was seeing a speech-language pathologist at entry into the study. Of the 25 twins for which hearing history information was available, 23 (92%) were reported to have no history of hearing loss, 1 (4%) was reported to have had a hearing problem
that resolved, and 1 (4%) parent reported the child’s hearing history status to be unknown. Of the 22 twins for which birth information was available, again based on parent-report via questionnaire, 20 were reported to have had no birth complications (91%) and 2 were reported to have had birth complications (9%). The 2 birth complications were described as pre-eclampsia with high blood pressure for one mother and large loss of blood after delivery of the second twin for the other mother. Of the 23 twins for whom length of delivery information was available, 6 (26%) were reportedly born before 35 weeks gestation while 17 (74%) reported births after 35 weeks gestation. Of the 6 twin pairs born before 35 weeks gestation, 2 pairs were born at 32 weeks, 2 pairs were born at 33 weeks, and 2 pairs were born at 34 weeks.

In a survey that was given one year after home visit 3, during wave 5 of WRRP, the parents of the twins were asked to answer questions regarding the twins’ previous use of twin language. Of the 15 parents who responded to a question about whether they had ever noticed the twins using a twin language, 4 (27%) reported that they had noticed use of twin language, 10 (67%) reported no use of twin language, and 1 (6%) was unknown. Of the 5 children who were noted to use twin language, 2 were reported to begin using it between 0-12 months and 3 were reported to begin using it between 13-24 months. In terms of when the twin language use abated, 1 pair reportedly stopped between 13-24 months, 1 between 25-36 months, 2 between 37-48 months, and 1 after 48 months.

**Conversational Language Samples**

Language samples for both groups were collected within the children’s homes by an examiner trained in procedures from Leadholm and Miller (1992). Specifically, examiners were instructed to interact conversationally with each child for a 15-minute
time period while playing with modeling clay. Examples of interaction tips included (a) ask open ended questions, (b) allow time for the child to speak, (c) make comments, and (d) avoid correcting the child’s speech (see DeThorne & Hart, 2009 for published guidelines). Some of the suggested topics of conversation included school, pets, extracurricular activities, and movies. The language samples were recorded onto cassette tapes or memory cards and shared with the Child Language and Literacy Laboratory at the University of Illinois for transcription based on standard procedures for Systematic Analysis of Language Transcripts (SALT; Leadholm & Miller, 1992). Sentences were parsed into C-Units based on pauses, natural intonation, semantic relationship, and conjunction use. Transcription was completed by research assistants who were trained in SALT transcription to a minimum level of 85% agreement for individual morphemes and utterance boundaries compared to a seasoned transcriber. Individual output measures of linguistic complexity were selected to mirror prior language sample analyses from prior WRRP publications (DeThorne et al., 2008; DeThorne & Hart, 2009).

The conversational measures taken from SALT are listed below. These measures are conversational in that they come from semi-structured conversations and not formalized language tests. However, they are not discourse-based in the sense that they capture the dynamic between conversational partners.

**MLU.** Mean Length of Utterance (MLU) provides an index of average number morphemes per utterance by dividing the total number of morphemes from all child complete and intelligible utterances in the sample by the total number of complete and intelligible child within the sample. MLU is often considered to be a useful measure of syntactic development, although it is a relatively gross estimate and correlates highly
with semantic measures as well (DeThorne, Johnson, & Loeb, 2005). In addition to its correlation with other measures, MLU has demonstrated developmental change and group differences within children’s school-age years (e.g., Scott 1995, Heilmann 2010).

**NDW.** Number of Different Words (NDW) reflects diversity in expressive vocabulary by totaling the number of different root words used in a language sample. Although NDW can be calculated from various sample permutations, the present study utilized the first 100 complete and intelligible utterances from each child’s sample. During school-age years, NDW has correlated with overall narrative quality (Fey, Catts, Proctor-Williams, Tomblin, & Zhang, 2004) and has shown promise in distinguishing typically developing children from children with specific language impairment (Hewitt, Scheffner Hammer, Yont, & Tomblin, 2005).

**NTW.** Number of Total Words (NTW) often serves as a measure of volubility (Leadholm & Miller, 1992), and is represented by the total number of tokens used within a sample. Similar to NDW, NTW was calculated for this study on the first 100 complete and intelligible utterances. NTW has been found to reflect developmental changes through school age (Leadholm & Miller, 1992) and to distinguish typically developing children from those with specific language impairment.

**TNC.** Total Number of Conjunctions (TNC) is considered an additional measure of syntactic complexity (Nippold, 1998) and represents the number of conjunctions produced by a child within a particular sample. In the present study, TNC was derived from the child’s first 100 complete and intelligible utterances. TNC has been found to show developmental change through school age (Leadholm & Miller, 1992).
**Formalized Measures**

In addition to the conversational language samples, examiners administered two semantic measures that represented a standard part of the WRRP assessment battery: the Boston Naming Test (BNT; Goodglass & Kaplan, 2001) and the Stanford Binet Vocabulary subtest (Thorndike, Hagen, & Sattler, 1986). The Boston Naming Test consists of a series of pictured objects associated with high- and low-frequency words that the child is asked to name. The vocabulary portion of the Stanford Binet Intelligence Scale requires children to give the definitions of various words. To be clear, I did not select these measures for the present study, but included them in the analyses as they were the only formal language measures available from HV3 of the larger WRRP assessment.

**Discourse Measures of Topic Management**

In order to pilot a coding system for topic management in the semi-structured conversational samples of school-age children, three twin-singleton matches (i.e., six total children) were selected for additional discourse analysis based in part on work by Mentis and Prutting (1998). The process for selecting the six children was as follows:

The IDs of all singletons within 3 months of the total twin-singleton sample mean of 8.31 years were written down on slips of paper, resulting in a total of 4 singletons within the desired age range of 8.08-8.58 years. Of these 4 singletons, IDs were drawn ‘blindly’ one at a time until a total of three singletons were selected. Each time an ID was chosen, that ID was returned to the pile in order to keep the chance of any ID being drawn the same. The topic-coding was completed for the three selected singletons and their previously paired twin matches. The result was a group of 4 males and 2 females.
with an average age of 8.33 years. A different examiner facilitated each of the 6 transcripts. This was not done intentionally, but was a chance result of the selection process. As the coding process was being developed for those six children, practice transcripts from two of the singletons in the study were coded, in order to obtain an idea of issues that might arise during the topic-coding. Once selected, the six transcripts were coded for both type and frequency of topics included, as well as the role each child and examiner utterance played in the process of topic management.

Each transcript went through three passes. In the first pass, each utterance was coded for topic and number of topic repetitions within the transcript. Because new topics were added as the coding progressed, a second pass was made to adjust for utterances that could now better be coded under one of the newly added categories. A final pass through the transcripts was made in order to check for examiner errors, such as failing to code the number of repetitions of a topic within the transcript accurately, and to total each of the measures, listed below.

**Topic-coding**

Before beginning the topic-coding, a list of anticipated topics was taken from the guidelines for language sample collection given to examiners during training (see Hart & DeThorne, 2008). Suggested topics for the interaction included, but were not limited to, (a) school, (b) family, (c) pets, (d) movies and television, and (e) holidays. These broad topics were then expanded to more specific topics based on my explicit review of two practice samples and experience transcribing numerous samples for the project as a whole. For example, the subject of school was divided into three specific topics: classes, teachers, and field trips. Through this process, a list of 25 topics was generated prior to
the start of the topic-coding, each of which was assigned a number for coding purposes. I personally reviewed all 6 transcripts from the 3 singleton-twin matches, coding each child and examiner utterance with one of the pre-established topics. As coding for each transcript progressed, there were instances in which topics that were discussed could not fall under any of the topic categories previously generated. In these cases, a new topic was created and added to the list of topics. For example, a topic about live performances and plays was created after one child spent a large portion of a transcript discussing plays that she attended.

The number of times specific topics were repeated within a single transcript was tracked by expanding the topic code by one decimal code. For example, if the topic ‘pets’ (assigned topic #11) was being discussed for the first time within the transcript, all utterances that occurred consecutively within that specific conversational exchange would be coded as 11.1, or topic 11 (pets). If the examiner and child returned to talk about pets later in the same transcript, those additional utterances would be coded as 11.2, for the second occurrence of topic 11 (pets).

**Role in Topic Management**

In addition to being coded for topic, each child and examiner utterance in a transcript was coded for one of two roles it played in the process of topic management: either as a topic shift (TS) or a topic continuation (TC). When the distinction between the two roles was not discernable, the utterance was considered uncodeable for its role in topic management (TNC). The majority of utterances considered uncodeable in terms of role were those marked as abandoned or interrupted during transcription. Topic shifts were defined as any utterance that differed notably from the topic of the previous
utterance. For instance, if the child and the examiner were discussing the child’s favorite subject in school and the examiner then asked if the child had any plans for summer vacation, the examiner’s utterance would be coded as a topic shift, from the topic of school to the topic of vacation. Topic continuations were used to denote any utterance in the transcript that did not function as a discernable topic shift. Consequently, each utterance that occurred after a topic shift was coded as a topic continuation (either by examiner or child: TCE, TCC) until a new topic was introduced by either the child or examiner. Utterances that functioned as comments or signs of engagement such as “yeah” or “mhm” were coded as topic continuations.

In sum then, each utterance in the six selected transcripts was coded for its topic (type and frequency) and for its role in the process of topic management. Readers are referred to Appendices A and B for final coding guidelines as well as an example page of a coded transcript. Subsequent to coding, the following six dependent measures were derived from each transcript and compared across children and between twins and singletons to derive a preliminary estimate of individual and group variability.

**Total number of topics.** The Total Number of Topics (TNT) provided a frequency count of how many different topics emerged within any particular transcript. In theory, this value could range from 1 to 40, the total number of different topics coded.

**Number of topic shifts initiated by the examiner/child.** The total number of topic shifts that were initiated by the examiner and the total number of topic shifts initiated by the child were separately summed for each transcript.

**Total number of episodes.** Each transcript was coded for the total number of episodes that occurred during the transcript. Episodes were conversational segments
delineated by topic shifts. Specifically, each time a topic shift occurred, that topic shift and all subsequent continuations of the topic were considered one episode. An episode ended and a new episode began when the child or examiner initiated a topic shift. If the topic was shifted to a previously discussed topic, it was still considered a new episode (of a repeat topic). As such, language transcripts could contain a larger number of episodes than topics due to repeat topics throughout the interaction.

**Length of shortest/longest episode.** The total number of utterances that occurred within each episode were counted to derive measures of both the shortest and longest episode for each transcript. Utterances that could not be coded in terms of their role in topic management were not counted in the total.

**Total number of coded utterances.** The total number of utterances that remained in the 15-minute language samples after all utterances that could not be coded were removed was calculated for each transcript.

**Average length of episode.** The total number of utterances that occurred within each episode were averaged across episodes within a transcript to derive the average length of episode per transcript.

**Analyses**

To address the primary and secondary research question, I derived mean comparison statistics for twins versus singletons related both to conversational and formal language measures. Prior literature focused on language differences in younger children suggested I would find differences in favor of singletons, particularly if a substantial number of twins experienced peri- and post-natal complications. If the predicted group differences were found, the third question regarding the role of prenatal
risk factors would be addressed by completing the same analysis after removal of the twins with reported perinatal risk factors to see whether or not significant group differences remained.
Chapter III: Results

Descriptives of Formalized, Conversational, and Topic-coding Measures

Descriptive data for both the formalized and conversational language measures are presented in Table 1 for twins and singletons. Because NDW, NTW, and TNC required 100 complete and intelligible utterances in order to be calculated, this information is missing for the 5 twins and 3 singletons that did not reach 100 C-Units. As a result, the number of participants for the descriptive data varies among measures. There is a noticeable trend for singletons to score higher than the twins across all measures except for the Stanford Binet vocabulary test, where the twins scored slightly higher. Descriptive data for the examiners who interacted with the twins and singletons are presented in Table 2. There was a trend for the twins’ examiners to have higher scores than the singletons’ examiners across measures, with the exception of TCICU.

Two singleton outliers, whose scores were at least two standard deviations above or below the mean on NLU, NDW, and NTW were identified via visual inspection. No outliers were identified in the twin group. Of the two singleton outliers, one scored 2 to 3 standard deviations above the mean on the listed measures, while the other outlier fell 2 to 3 standard deviations below the mean. Viewing scatter plots of age by conversational measure, it was determined that the child who performed well above the mean was among the oldest children from the singleton group. Despite being one of the older children, this child still performed well above the other singletons of the same age range. The singleton outlier who scored 2 to 3 standard deviations below the mean was of average age, thus his lower scores could not be explained by age.
A trained examiner re-listened to the language samples of the two singleton outliers, and their transcripts were checked for any recording or transcription errors that could have resulted in their deviant scores. Although no evidence of procedural error was identified, the outliers were removed to examine their impact on the descriptives. When the separate analysis was run without the singleton outliers, the singletons’ average age became one month younger than the twins’, while the singletons still had a trend of scoring slightly higher than the twins on all measures except for the BNT. Given that outliers were not due to procedural error and did not demonstrate notable effects on group means, the values were maintained in remaining analyses.

The topic-coding process, described in the Methods, resulted in a total of 40 topics, with the frequency of each topic provided for each child. Of the 40 total topics listed, only 35 were utilized in the six transcripts coded (see Table 3); the other five topics were expected based on examiner training and practice samples but never emerged in the selected samples. The most common topics that emerged were modeling clay, with 34 total episodes devoted to the topic, school, with 17 episodes, hobbies/extracurriculars, with 8 episodes, and the immediate environment other than the modeling clay (e.g. something in the room, something happening outside) with a total of 8 episodes. Modeling clay was the only topic that emerged across all six interactions, although 21 of the 34 episodes on the topic came from one sample.

Descriptive results of the preliminary topic-coding analyses are presented in Table 4. Given that topic-coding was coded for a subsample of six children, three singletons and their twin matches, individual data is provided in Table 4, as well as the group means. Group means are to be considered with caution given that there were only three...
children in each group. That said, the singletons as a group had higher values on 6 of the 8 measures, all measures except Total Number of Topic Shifts by the child and Total Number of Episodes); however ranges overlapped across groups for all individual measures. For both groups of children the examiners averaged more than half of the topic shifting.

Correlation Analyses

Child measures. To examine associations across variables, Tables 5 and 6 contain bivariate correlations across individual language measures for singletons and twins, respectively. Across correlation analyses, alpha was set at .01 to reduce the chance of spurious associations. Correlations for both groups, twins and singletons, followed the same trends. As expected, the conversational language measures were highly correlated for both groups, with the exception of TCICU. The significant correlation coefficients ranged from .63 to .94 for twins and .58 to .97 for singletons. Additionally, the two formalized language test scores were highly correlated, with a correlation coefficient of .70 for both groups. The formalized language scores were also found to highly correlate with age. For singletons, age correlated with the BNT at .76 and the SB-Vocab at .71, and for twins, age was correlated with the BNT at .55 and the SB-Vocab at .65. Of interest, formal and conversational language measures were not highly correlated with one another within either group, with all correlation coefficients being less than .40.

Examiner measures. Tables 7 and 8 contain bivariate correlations of conversational language measures from the examiners, for the singletons and twins respectively. As was the case for the children, the examiners’ conversational language measures were highly correlated (with the exclusion of TCICU). Significant correlation
coefficients ranged from .43 to .90 for the singletons’ examiners and .34-.90 for the twins’ examiners.

**Child and examiner measures.** To examine potential associations between child and examiner measures, Tables 9 and 10 contain the bivariate correlations of the singletons’ and twins’ conversational language measures with their examiners’ conversational language measures, respectively. Of the 35 correlations derived in relation to the singleton samples, 9 reached statistical significance, with effect sizes ranging from -.53 to -.72. Of the 35 correlations run for the twin samples, 4 reached statistical significance, with effect sizes ranging from -.50 to -.60. Singleton and examiner TCICU correlated at .406; however, this number did not reach significance at the .01 level. The consistently negative associations indicated a tendency toward inverse associations between examiner and child measures, such that when one was higher the other was lower.

**Mean Comparisons between Groups**

Due to the significant correlation across dependent variables and the factor analyses reported in prior study of the same variables (e.g., DeThorne et al., 2008), two standardized composites were formed to compare group means, one for conversational measures and one for formal measures. Specifically, all measures were converted to z-scores based on the means and standard deviations from the combined twin and singleton samples. As such, a z-score of -.5 for MLU would indicate that a child’s MLU was half a standard deviation lower than the mean for the group as a whole. Descriptive data for the children’s z-scores are presented in Table 11. The z-scores were then averaged as follows: (a) MLU, NDW, NTW, and TNC for the Conversational Composite, and (b)
BNT and SB-Vocab for the Formal Composite.

The Conversational and Formal Composites were utilized to examine this study’s primary research question of potential language differences between school-age twins and singletons. Specifically, each composite was entered separately as the test variable in an independent-samples t-test via SPSS with twin versus singletons as the grouping variable. The group means for the Conversational Composite were .11 for singletons, with a standard deviation of .99, and -.06 for twins, with a standard deviation of .84. Means for the Formal Composite were .02 for both singletons and twins, with a standard deviation of .95 for singletons and .89 for twins. Results of the t-tests indicated there were no significant differences between the scores of twins and singletons on either the Conversational Composite \((t=.64, df=46, p=.52)\) or Formal Composite \((t=-.18, df=54, p=.86)\).

Because at least one prior study found that dizygotic twins scored significantly higher than monozygotic twins on language measures (Lytton et. al, 1977), the composite scores of the monozygotic twins were compared to those of the dizygotic twins on both the formalized and conversational composites. Consistent with the previous finding, the DZ twins averaged higher scores than MZ twins for both composites, with z-scores of .18 and -.05 respectively for the Formalized Composite and .32 and -.41 on the Conversational Composite. Although as a group the twins’ scores did not differ significantly from those of the singletons, Figures 1 and 2 provide scatter plots of the Formal and Conversational Composites as a function of age for all twins, with the six twins born prior to 35 weeks gestation highlighted. Note that the Conversational Composite was available for only 3 of the 6 twins born prior to 35 weeks due to samples...
with fewer than 100 complete and intelligible utterances. Similarly, Figures 3 and 4 provide scatterplots of the Composites by child age for the four twins who were reported to use twin language.
Chapter IV: Discussion

The primary finding from this study is that twins as a general rule do not differ from singletons in their expressive language abilities at school age. The results were consistent across both formal and conversational language measures. In addition, the topic-coding on twin-singleton matches revealed substantial overlap between groups across all measures. Results from the topic-coding were promising in term of capturing the type and frequency of topics across interactions; however concerns regarding coding reliability and the need for validity evidence would need to be addressed before the measures were implemented on a fuller scale. The remainder of the ensuing Discussion section will address how my findings (a) relate to prior literature, (b) highlight promise and potential pitfalls related to topic-coding, and (c) are tempered by limitations of our sample and specific measures.

Comparison to Prior Literature

The findings of the current study contrast to previous studies, which have reported expressive language delays in twins in relation to singletons (i.e., Day, 1932; Mittler, 1970; Rutter, Thorpe, Greenwood, Northstone, & Golding, 2003; Stafford, 1987; and Lytton, Conway, & Suavé 1977). One possibility is that differences across studies are related to child age. Language differences in twins have been reported in children under the age of five years, whereas the present study included children age 7-10 years. It is feasible that early language differences between twins and singletons are evident at younger ages but are either no longer discernible or entirely dissipate by school age. Consistent with this interpretation are longitudinal findings from Lytton, Watts, & Dunn (1987), in which differences found between twins and singletons during the preschool
years were largely nonexistent by the school age years (with the exception of monozygotic twins). Hypothesing a specific mechanism or rationale for twins’ ability to ‘catch up’ seems particularly important, especially in light of the fact that early language difficulties have also been hypothesized to contribute to a negative downward spiral in terms of language learning and social interaction (Redmond & Rice, 1998). In terms of twins ‘catching up,’ Lytton et al. (1987) proposed that the early differences in twins compared to singletons relates to differences in parental input that dissipate once children’s circle of communicative partners widens. Another possibility is that the higher rate of prematurity and birth complications in twins (Martin et al., 2010) sets them temporarily behind in early language learning. For example, if twins who were born prematurely or suffered birth complications were consequently in the hospital for the first weeks or months of life, their language exposure could be slightly delayed from full term singletons in the first years of life. If this was the case, however, it’s unclear why such differences would dissipate rather than continue or widen. It is possible that as long as children, twins included, aren’t suffering from long term delays and receive adequate language exposure within a certain “window of opportunity” for language acquisition, they will be able to reach full potential by school age regardless of the nature of their earliest environmental experience. Arguably parallel evidence for this position could be found in literature on bilingual language acquisition, which suggests that children are able to master additional languages if adequate experience is available within early years (Johnson & Newport, 1989; DeKeyser, 2000).

A second possibility for contrasting findings across studies is differences in methodology. Specifically, many prior studies reporting twin-singleton language
differences did not implement consistent assessment procedures across twin and singleton
groups or ensure adequate control for group differences in socioeconomic status and birth
histories (Day, 1932; Mittler, 1970; Stafford, 1987; Lytton et al., 1977). In the studies
that did control for birthweight and birth complications (e.g., Rutter et al., 2003) only
very mild group differences were discovered. As such, perhaps the previous research was
detecting delays due to birth complications or prematurity, rather than differences
inherent to being a twin, a finding that would be generally consistent with the present
study.

Promise and Potential Pitfalls of Topic-coding

Topic-coding from the present study was intended to pilot procedures for
analyzing the transactional nature of topic management within the conversational
discourse of school-age children with adult examiners. The results provided preliminary
coding guidelines (see Appendix A) as well as an initial estimate of individual variability
across measures. In addition, the influential role of the examiner became apparent from
their relatively high proportion of topic shifts in comparison to the singletons.

Concerns about the reliability of the topic-coding measures would need to be
addressed before further implementation. Specifically, topic transitions were often subtle
and difficult to delineate confidently. For instance, during an exchange related to pets,
one child included details about specific pets within his family (e.g. lizards, hermit crabs,
goldfish). A coder would need to decide if transitions to specific pets would constitute
individual topic shifts or fall under the larger topic of ‘pets’. I opted for the larger
category of pets in this case for simplicity; however, different decisions could have easily
lead to contrasting values.
In addition to reliability concerns, validity concerns also surfaced. In particular, I was concerned that the frequency counts themselves could be misleading. For instance, it might be easy on face-value to interpret the transcript containing 21 examiner topic shifts and 24 child topic shifts as a more balanced interaction than the transcript containing 20 examiner topic shifts and 0 child topic shifts. In actuality, the former transcript primarily contained examiner topic shifts that attempted to introduce new topics to the interaction as the child continuously shifted the topic back to the modeling clay with which he was playing. The child in this interaction often did not acknowledge the examiner’s clear attempt at initiating new topics, and displayed difficulty discussing any topics outside of modeling clay. On the contrary, the transcript containing 20 topic shifts by the examiner and 0 topic shifts by the child contained more balanced back and forth between the child and examiner. Although the child was not introducing topics to the interaction, the child was able to maintain topics introduced by the examiner (resulting in the longest episode of all 6 transcripts-72 utterances).

Despite concerns with both reliability and validity, the topic coding did help illuminate the role that the examiner played in the interaction. The examiners averaged over half of the topic shifting for both groups, which demonstrates the role they likely played in guiding the interaction. What is more, the examiners were largely the initiators of the most frequently repeated topics across transcripts (with the exception of one interaction in which the child repeatedly shifted the topic to modeling clay). Additionally, three of the four most commonly repeated topics were in the guidelines of suggested topics for examiners to use during the interaction, which is indication that the examiners were controlling parts of the interaction based on the given guidelines. Converging
evidence for the shared contribution of the individuals in the interaction came from the significant associations that emerged between child and examiner conversational language measures (see tables 9 and 10). Of interest was the finding that conversational measures (with the exception of TCICU) were negatively correlated between child and examiner, suggesting perhaps that increased examiner complexity lead to decreased child complexity or vice versa.

**Limitations of the Current Study**

Despite the contribution of the present study overall, I’ll highlight two limitations in relation to external validity. First and foremost, although WRRP was intended as a population-based sampling of twins in Ohio, the study fails to represent the full range of individual variability, particularly in regard to children with medical complications and developmental difficulties. For example, despite the estimated 11.6% of twins born prior to 32 weeks in the populations as a whole (Center for Disease Control, 2008), none of the twins in the current study were reportedly this premature and only two reported birth complications. However, in terms of the group average length of gestation, the reported average length of gestation of 35.5 weeks for twins in the present study reasonably matched the average of 35.2 weeks for twins in the CDC’s 2009 report.

A second limitation relates to the measures themselves. With the exception of the topic coding, measures in the current study were taken from the larger WRRP study, which was not originally designed as a study of child language. Most explicitly, the Boston Naming Test is not designed as an expressive vocabulary measure for young children, and the SB-Vocab is one subtest from a larger IQ battery; neither are considered best practice assessment tools for child language development (Bogue & DeThorne,
In addition, the conversational language measures, though supported by validity evidence and best practice in the field, are limited in their ability to reflect the interactional nature of conversation. In the future, examiners interested in measuring the linguistic complexity of children’s language samples may benefit more from the use of SALT measures such as MLU and NDW, whereas examiners interested in measuring the communicative competence of the participants may benefit more from the use of discourse measures, such as the topic-coding employed here. In addition, the nature of the samples in some ways resembled more of an informal interview than a reciprocal conversation, in part because it included largely unfamiliar partners and was embedded within a larger assessment protocol.

**Conclusion**

In sum, this study provides evidence that twinship itself is not inherently linked to differences in school-age language skills. This finding was true for both formal and conversational assessments. These results are encouraging, both for those who are raising and educating twins, as well as those who are studying them for the purpose of understanding language development at large. In short, these findings suggest that if twins are at greater risk for language difficulties early in development, it is likely due to causal factors like birth complications that are not exclusive to twins. As such, further studies would benefit from including twins who differ in regard to explicitly hypothesized causal factors of language differences rather than assuming twinship in general as a disabling condition.
Chapter V: Tables and Figures

Tables

Table 1. Descriptive Data on Child Language Sample Measures, Divided into Twin and Singleton Groups

<table>
<thead>
<tr>
<th>Group Language Measure</th>
<th>Twin</th>
<th>Mean (SD)</th>
<th>Range</th>
<th>Singleton</th>
<th>Mean (SD)</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>n</td>
<td></td>
<td>n</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCICU</td>
<td>28</td>
<td>132.86 (39.730)</td>
<td>44-199</td>
<td>28</td>
<td>153.64 (43.891)</td>
<td>63-250</td>
</tr>
<tr>
<td>MLU</td>
<td>28</td>
<td>6.2693 (1.43546)</td>
<td>4.16-</td>
<td>28</td>
<td>6.6082 (1.64803)</td>
<td>2.81-11.47</td>
</tr>
<tr>
<td>NDW</td>
<td>23</td>
<td>203.57 (25.763)</td>
<td>150-263</td>
<td>25</td>
<td>207.88 (36.252)</td>
<td>112-286</td>
</tr>
<tr>
<td>NTW</td>
<td>23</td>
<td>582.35 (113.385)</td>
<td>428-784</td>
<td>25</td>
<td>597.92 (153.968)</td>
<td>250-1047</td>
</tr>
<tr>
<td>TNC</td>
<td>23</td>
<td>42.96 (23.019)</td>
<td>13-95</td>
<td>25</td>
<td>50.08 (20.025)</td>
<td>15-99</td>
</tr>
</tbody>
</table>

Note: TCICU = Total Complete and Intelligible C-units, MLU = mean length of c-unit, NDW = number of different words, NTW = total number of words, TNC = total number of conjunctions, BNT= Boston Naming Test, SB-Voc= Stanford Binet Vocabulary subtest
Table 2. *Descriptive Data on Examiner Language Sample Measures, Divided into Twin and Singleton Groups*

<table>
<thead>
<tr>
<th>Group</th>
<th>Examiner- Twin Group</th>
<th>Examiner- Singleton Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language Measure</td>
<td>n</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>TCICU</td>
<td>28</td>
<td>125.11 (33.421)</td>
</tr>
<tr>
<td>MLU</td>
<td>28</td>
<td>5.9114 (.83875)</td>
</tr>
<tr>
<td>NDW</td>
<td>22</td>
<td>172.82 (18.412)</td>
</tr>
<tr>
<td>NTW</td>
<td>22</td>
<td>550.41 (58.810)</td>
</tr>
<tr>
<td>TNC</td>
<td>22</td>
<td>27.32 (7.840)</td>
</tr>
</tbody>
</table>

*Note: TCICU = Total Complete and Intelligible C-units, MLU = mean length of c-unit, NDW = number of different words, NTW = total number of words, TNC = total number of conjunctions*
Table 3. *List of Topics and Frequencies of Topics Across Episodes for Topic-Coding*

<table>
<thead>
<tr>
<th>Topic #</th>
<th>Topic</th>
<th>Twin 1 Topic Frequency</th>
<th>Twin 2 Topic Frequency</th>
<th>Twin 3 Topic Frequency</th>
<th>Sibling 1 Topic Frequency</th>
<th>Sibling 2 Topic Frequency</th>
<th>Sibling 3 Topic Frequency</th>
<th>Frequency of Topic Across Transcripts</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Model Magic/Play Doh'</td>
<td>3</td>
<td>3</td>
<td>21</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>34</td>
</tr>
<tr>
<td>2</td>
<td>TV shows/TV Characters</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Movies/Movie Characters</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Books</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>School-Classes/Subjects/Grade level related questions/Recess School-Teachers/Talk of specific teachers School-Fieldtrips/Special Events Parents Siblings Extended Family/Family friends Pets Friends/Playing with friends Holidays Vacations/Breaks/Trips Games-Outdoor/Indoor non-board games Games- Board/Card Games- Video Hobbies/Extracurriculars/Interests ts Food/Restaurants/Cooking Team Sports- Participating Sports- Favorite Teams/Watching sports New house Weather Animals/Fish/Reptiles Immediate Environment/Situation Making clothing/Jewelry /Crafts/Crocheting Rides/Theme park attractions Candy Live shows/Entertainment/Plays Books/Reading for leisure/Characters and plotlines Boy scouts BB Guns/Bow and Arrows School- Playground Favorite Colors Dance/Cheerleading Fishing Toys Zoo Skiing/Snowboarding Snow/Playing outside in snow/Building snowmen</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>16</td>
<td>22</td>
<td>45</td>
<td>20</td>
<td>35</td>
<td>17</td>
<td>155</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4. *Results of Topic-Coding Measures for Twins and Singletons by Group, Followed by Individual Scores*

<table>
<thead>
<tr>
<th>Measure</th>
<th>Twins' Means (n=3)</th>
<th>Singletons' Means (n=3)</th>
<th>Twin 1</th>
<th>Twin 2</th>
<th>Twin 3</th>
<th>Singleton 1</th>
<th>Singleton 2</th>
<th>Singleton 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Number of Topics</td>
<td>9.3</td>
<td>15.3</td>
<td>8</td>
<td>11</td>
<td>9</td>
<td>14</td>
<td>23</td>
<td>9</td>
</tr>
<tr>
<td>Total # of Coded Utterances</td>
<td>300</td>
<td>315</td>
<td>254</td>
<td>305</td>
<td>340</td>
<td>403</td>
<td>342</td>
<td>200</td>
</tr>
<tr>
<td>Total # Topic Shifts_Chiild</td>
<td>12.3</td>
<td>3.3</td>
<td>6</td>
<td>24</td>
<td>7</td>
<td>0</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>Total # Topic Shifts_Examiner</td>
<td>15.3</td>
<td>20.7</td>
<td>10</td>
<td>15</td>
<td>15</td>
<td>20</td>
<td>33</td>
<td>9</td>
</tr>
<tr>
<td>Total # Episodes</td>
<td>28</td>
<td>24</td>
<td>16</td>
<td>22</td>
<td>45</td>
<td>20</td>
<td>35</td>
<td>17</td>
</tr>
<tr>
<td>Length of Shortest Episode</td>
<td>2</td>
<td>2.3</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Length of Longest Episode</td>
<td>42.3</td>
<td>48.3</td>
<td>43</td>
<td>54</td>
<td>30</td>
<td>72</td>
<td>45</td>
<td>28</td>
</tr>
<tr>
<td>Average Length of Episode</td>
<td>12.4</td>
<td>20.8</td>
<td>15.9</td>
<td>13.9</td>
<td>7.6</td>
<td>20.2</td>
<td>9.8</td>
<td>11.8</td>
</tr>
</tbody>
</table>
Table 5. Bivariate Correlations for Standardized and Conversational Language Measures in Singletons

<table>
<thead>
<tr>
<th></th>
<th>BNT</th>
<th>SB-Voc</th>
<th>MLU</th>
<th>NDW</th>
<th>NTW</th>
<th>TNC</th>
<th>TCICU</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>BNT</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SB-Voc</td>
<td>.70**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MLU</td>
<td>.20</td>
<td>.20</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NDW</td>
<td>.40*</td>
<td>.39</td>
<td>.83**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NTW</td>
<td>.27</td>
<td>.28</td>
<td>.97**</td>
<td>.86**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TNC</td>
<td>-.04</td>
<td>.07</td>
<td>.77**</td>
<td>.58**</td>
<td>.77**</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCICU</td>
<td>-.31</td>
<td>-.14</td>
<td>.16</td>
<td>-.10</td>
<td>.06</td>
<td>.37</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>.76**</td>
<td>.71**</td>
<td>.27</td>
<td>.30</td>
<td>.27</td>
<td>.09</td>
<td>-.26</td>
<td>1</td>
</tr>
</tbody>
</table>

**Correlation is significant at the 0.01 level (2-tailed)
*Correlation is significant at the 0.05 level (2-tailed)

Note: BNT=Boston Naming Test, SB-Voc=Stanford Binet Vocabulary Subtest, MLU=Mean length of c-unit, NDW=Number of different words, NTW=Total number of words, TNC=Total number of conjunctions, TCICU=Total complete and intelligible c-unit.
Table 6. Bivariate Correlations for Standardized and Conversational Language Measures in Twins

<table>
<thead>
<tr>
<th></th>
<th>BNT</th>
<th>SB-Voc</th>
<th>MLU</th>
<th>NDW</th>
<th>NTW</th>
<th>TNC</th>
<th>TCICU</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>BNT</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SB-Voc</td>
<td>.70**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MLU</td>
<td>.11</td>
<td>.29</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NDW</td>
<td>.19</td>
<td>.36</td>
<td>.76**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NTW</td>
<td>.18</td>
<td>.29</td>
<td>.94**</td>
<td>.75**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TNC</td>
<td>.19</td>
<td>.28</td>
<td>.87**</td>
<td>.63**</td>
<td>.86**</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCICU</td>
<td>-.15</td>
<td>-.03</td>
<td>.52**</td>
<td>.24</td>
<td>.51*</td>
<td>.30</td>
<td>.13</td>
<td>1</td>
</tr>
<tr>
<td>Age</td>
<td>.55**</td>
<td>.65**</td>
<td>.21</td>
<td>.30</td>
<td>.25</td>
<td>.38</td>
<td>.13</td>
<td>1</td>
</tr>
</tbody>
</table>

**Correlation is significant at the 0.01 level (2-tailed)
*Correlation is significant at the 0.05 level (2-tailed)

Note: BNT=Boston Naming Test, SB-Voc= Stanford Binet Vocabulary Subtest, MLU = Mean length of c-unit, NDW = Number of different words, NTW = Total number of words, TNC = Total number of conjunctions,, TCICU = Total complete and intelligible c-units
Table 7. *Bivariate Correlations for Conversational Language Measures in Singletons'* Examiners

<table>
<thead>
<tr>
<th></th>
<th>MLU</th>
<th>NDW</th>
<th>NTW</th>
<th>TNC</th>
<th>TCICU</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLU</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NDW</td>
<td>.76**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NTW</td>
<td>.90**</td>
<td>.77**</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TNC</td>
<td>.43*</td>
<td>.44*</td>
<td>.77**</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>TCICU</td>
<td>.03</td>
<td>.12</td>
<td>.22</td>
<td>.22</td>
<td>1</td>
</tr>
</tbody>
</table>

** Correlation is significant at the 0.01 level (2-tailed)
* Correlation is significant at the 0.05 level (2-tailed)

*Note: MLU = Mean length of c-unit, NDW = Number of different words, NTW = Total number of words, TNC = Total number of conjunctions., TCICU = Total complete and intelligible c-units*
Table 8. *Bivariate Correlations for Conversational Language Measures in Twins’ Examiners*

<table>
<thead>
<tr>
<th></th>
<th>MLU</th>
<th>NDW</th>
<th>NTW</th>
<th>TNC</th>
<th>TCICU</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MLU</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NDW</td>
<td>.83**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NTW</td>
<td>.89**</td>
<td>.90**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TNC</td>
<td>.34</td>
<td>.46*</td>
<td>.46*</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCICU</td>
<td>-.14</td>
<td>-.30</td>
<td>-.32</td>
<td>-.14</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

**Correlation is significant at the 0.01 level (2-tailed)**

*Correlation is significant at the 0.05 level (2-tailed)*

*Note: MLU = Mean length of c-unit, NDW = Number of different words, NTW = Total number of words, TNC = Total number of conjunctions, TCICU = Total complete and intelligible c-units*
Table 9. *Bivariate Correlations for Conversational Language Measures of Singletons and Examiners*

<table>
<thead>
<tr>
<th></th>
<th>MLU-E</th>
<th>NDW-E</th>
<th>NTW-E</th>
<th>TNC-E</th>
<th>TCICU-E</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLU-C</td>
<td>-.68**</td>
<td>-.37</td>
<td>-.53**</td>
<td>-.30</td>
<td>-.60**</td>
</tr>
<tr>
<td>NDW-C</td>
<td>-.58**</td>
<td>-.32</td>
<td>-.55**</td>
<td>-.37</td>
<td>-.69**</td>
</tr>
<tr>
<td>NTW-C</td>
<td>-.65**</td>
<td>-.29</td>
<td>-.51*</td>
<td>-.21</td>
<td>-.72**</td>
</tr>
<tr>
<td>TNC-C</td>
<td>-.53**</td>
<td>-.40</td>
<td>-.52*</td>
<td>-.22</td>
<td>-.37</td>
</tr>
<tr>
<td>TCICU-C</td>
<td>-.38*</td>
<td>-.31</td>
<td>-.19</td>
<td>.04</td>
<td>.41*</td>
</tr>
<tr>
<td>BNT</td>
<td>-.10</td>
<td>.21</td>
<td>.02</td>
<td>-.26</td>
<td>-.16</td>
</tr>
<tr>
<td>SB-Vocab</td>
<td>-.03</td>
<td>.14</td>
<td>-.02</td>
<td>-.10</td>
<td>.02</td>
</tr>
</tbody>
</table>

** Correlation is significant at the 0.01 level (2-tailed)
* Correlation is significant at the 0.05 level (2-tailed)

*Note: MLU = Mean length of c-unit, NDW = Number of different words, NTW = Total number of words, TNC = Total number of conjunctions, TCICU = Total complete and intelligible c-units*
Table 10. Bivariate Correlations for Conversational Language Measures for Twins and Examiners

<table>
<thead>
<tr>
<th></th>
<th>MLU-E</th>
<th>NDW-E</th>
<th>NTW-E</th>
<th>TNC-E</th>
<th>TCICU-E</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLU-C</td>
<td>-.59**</td>
<td>-.52*</td>
<td>-.48*</td>
<td>-.30</td>
<td>-.20</td>
</tr>
<tr>
<td>NDW-C</td>
<td>-.57**</td>
<td>-.50*</td>
<td>-.50*</td>
<td>-.21</td>
<td>-.09</td>
</tr>
<tr>
<td>NTW-C</td>
<td>-.57**</td>
<td>-.37</td>
<td>-.33</td>
<td>-.22</td>
<td>-.30</td>
</tr>
<tr>
<td>TNC-C</td>
<td>-.53**</td>
<td>-.52*</td>
<td>-.42</td>
<td>-.17</td>
<td>.23</td>
</tr>
<tr>
<td>TCICU-C</td>
<td>-.23</td>
<td>-.21</td>
<td>-.08</td>
<td>-.24</td>
<td>.20</td>
</tr>
<tr>
<td>BNT</td>
<td>-.12</td>
<td>.14</td>
<td>.10</td>
<td>.22</td>
<td>-.29</td>
</tr>
<tr>
<td>SB-Voc</td>
<td>-.15</td>
<td>.12</td>
<td>.14</td>
<td>.31</td>
<td>-.37</td>
</tr>
</tbody>
</table>

** Correlation is significant at the 0.01 level (2-tailed)
*Correlation is significant at the 0.05 level (2-tailed)

Note: MLU = Mean length of c-unit, NDW = Number of different words, NTW = Total number of words, TNC = Total number of conjunctions, TCICU = Total complete and intelligible c-units
Table 11. Descriptive Data on Children’s Z-Scores, Divided into Twin and Singleton Groups

<table>
<thead>
<tr>
<th>Group Measure</th>
<th>Singleton</th>
<th>Twin</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>MLU</td>
<td>28</td>
<td>.11 (1.07)</td>
</tr>
<tr>
<td>NDW</td>
<td>25</td>
<td>.07 (1.15)</td>
</tr>
<tr>
<td>NTW</td>
<td>25</td>
<td>.06 (1.14)</td>
</tr>
<tr>
<td>TNC</td>
<td>25</td>
<td>.16 (.92)</td>
</tr>
<tr>
<td>BNT</td>
<td>28</td>
<td>-.16 (1.05)</td>
</tr>
<tr>
<td>SB-Voc</td>
<td>28</td>
<td>.11 (1.00)</td>
</tr>
</tbody>
</table>

Note: MLU = mean length of c-unit, NDW = number of different words, NTW = total number of words, TNC = total number of conjunctions, BNT = Boston Naming Test, SB-Voc = Stanford Binet Vocabulary subtest
Figures

Figure 1. Scatter Plots of Children’s Formal Composite Scores, With Twins Born Before 35 Weeks Highlighted
**Figure 2.** Scatter Plots of Children’s Conversational Composite Scores, With Twins Born Before 35 Weeks Highlighted
Figure 3. Scatter Plots of Children’s Formal Composite Scores, With Twins With Reported History of Twin Language Highlighted
Figure 4. Scatter Plots of Children’s Conversational Composite Scores, With Twins With Reported History of Twin Language Highlighted
References


Appendix A

Guidelines for Topic-coding

Codes:

**Individual utterance codes** (all individual utterances should have one or more of the following codes):
- **TSC**- Topic shift, initiated by child
- **TSE**- Topic shift, initiated by examiner
- **TCC**- Topic continuation, child
- **TCE**- Topic continuation, examiner
- **E1**- Episode 1
- **E2, E3, E4**... (Episodes 2, 3, 4...)
- **T1**- Topic 1
- **T2, T3, T4**... (Topics 2, 3, 4...)
- **X.1**- Topic #, first occurrence
- **X.2, X.3, X.4** (second, third, and fourth occurrence of topic number)
- **TNC**- topic not codeable, includes abandoned and interrupted utterances as well as utterances for which topic could not be determined

**Episode codes** (each episode should contain totals for the following):
- **TNT**- Total number of topics
- **ALE**- Average length of episode
- **TNCU**- total number of coded utterances in the transcript excluding those marked TNC and utterances at the beginning or end of transcript which did not receive coding (see later)
- **LSE**- length of the shortest episode within the transcript
- **LLE**- length of the longest episode within the transcript

**Topics:** See list of topics

**Coding rules:**

1. Begin coding at the utterance that comes before the first response.

E.g.

E so I’m go/ing to pull mine out.
E and I don’t know what i’m go/ing to make.
E have you thought about if you’re going to make something? –**TSE**
C (um) i’m just gonna like really make just these color/s mixed up. –**TCC**

2. Code the utterance before the first response as a topic shift (TS- E or C).
3. Each utterance following the initiation of a topic will be coded as topic continuation (TC, examiner or child), until the topic shifts.

e.g.
(talking about Model Magic clay)
e uhhuh and then do you know what make/3s green? **TCE**
c (um) black and blue? **TCC**
e you’re close. **TCE**
e well I think it’s yellow and blue actually. **TCE**
e I saw you playing basketball outside when we got here. **TSE**

4. All of the utterances that occur prior to a topic shift (TS) are considered to be one episode. Make note of each episode in the side margin (e.g. E1, E2, E3).

5. If a topic repeats often but is missing from the list, or a large portion of a transcript is devoted to a topic that is not on the list, the topic can be added to the list as a new topic. Previous transcripts will then need to go through a new coding pass in order to check for use of any added topics.

7. It is possible for an episode to contain only one utterance if the child or examiner attempts to initiate a new topic and the attempt at initiation is not returned by the other person in the interaction

e.g.
c it look/3s like tiedye. **TCC 1.17**
e uhhuh it does **TCE 1.17**
c I can’t make it. **TCC 1.17**
e have you ever made anything that look/3s like tiedye? **TSE 26.1**
c dang! **TSC 1.18**
e yep. **TCE 1.18**
c we’ll mix a little bit (of yellow) more yellow and more red. **TCC 1.18**

In the above example, the examiner attempted to shift the conversation to the topic of making tie-dye. However, the child did not acknowledge this attempt at shifting topics and instead returned the interaction to the topic of modeling clay.

8. Sometimes topic shifts are very subtle. Think of each block of conversation as being its own “story.” By asking what the topic of each utterance is, it becomes easier to differentiate between a continuing topic and an offshoot from the previous topic that transitioned into a new topic.

9. Each utterance in the transcript (minus those coded as TNC- topic not codeable- see later) should be marked with a number that signifies the topic number, and the number of times the current topic has been repeated within the utterance.
-the number 11.1 signifies the topic of pets, the first time pets are being discussed in the current transcript
-the number 11.4 signifies the topic of pets, the 4\textsuperscript{th} time pets are being discussed in an episode of the current transcript

9. After the transcript is coded, the total number of episodes, topics, topic shifts by both the child and examiner, length of the longest and shortest episode, average length of episode, and number of coded utterances should be calculated. In order to calculate average length of episode, the total number of utterances in each episode must be totaled and then averaged. All calculations can be recorded at the top of each transcript and later entered into the database.

10. The final utterances of the transcript are not coded if they involve the examiner saying he/she is going to turn off the recording.

13. Abandoned or interrupted utterances are not assigned a topic, but instead are coded as Topic Not Codeable (TNC). In the case that it is impossible to determine the topic of an utterance, it may also be coded as TNC. Topics that are coded TNC are not included in any of the calculations.

14. At the end of each transcript, every single utterance should have a code for type of utterance (e.g., topic shift, topic continuation, topic not codeable), number of topic, and occurrence number of topic, as described earlier (e.g. TCC 11.4, TSE 2.3, TSC 1.1). Topic not codeable (TNC) utterances do not receive a topic or occurrence number.
Appendix B

Sample Page of Topic-coding from Transcript

E Second? TCE 5.2
E And (what are) what kind/s of thing/s are you doing in school? TCE 5.2
C Math. TCC 5.2
E <Oh>. TCE 5.2
C <Rocket>math and writing, spelling. TCC 5.2
E Wow, so what kind of thing/s do you write about? TCE 5.2
C (Um) we don’t really do writing that much we do (like) cursive. TCC 5.2
E Oh okay is that really hard? TCE 5.2
C Kind of. TCC 5.2
E (What/s your favorite) what/s your favorite subject to do in school? TCE 5.2
C Math. TCC 5.2
E Math? TCE 5.2
E And what do you do in math right now? TCE 5.2
C Rocketmath. TCC 5.2
E What does that mean? TCE 5.2
C (It/s like) it goes A through Z. TCC 5.2
C And every time you get to> TNC
C (Like) if you pass A and you go to B it get/3s harder. TCC 5.2
E Oh wow so what kind/s of problem/s are there in these like different level/s? TCE 5.2
E What do you do? TCE 5.2
C You just do plus. TCC 5.2
E Oh. TCE 5.2
C And minus. TCC 5.2
E Wow. TCE 5.2
C But on Rocketmath we only do plus. TCC 5.2
E Okay. TCE 5.2
E So how far have you gotten in it? TCE 5.2
C (Um) N. TCC 5.2
E Wow that/s really far. TCE 5.2
E And you said it goes A through what? TCE 5.2
C Z. TCC 5.2
E A through Z. TCE 5.2
E Very cool. TCE 5.2

_________________________________________Episode 7
E So (wh*) what did you do over (di*) winter break? TSE 14.2
C We went skiing. TCC 14.2
E Mhm. TCE 14.2
C And (um) we had Christmas at our house. TCC 14.2
E Oh yeah? TCC 14.2

_________________________________________Episode 8
E And what did you get for Christmas? TSE 13.1

73
C  I got a soccer ball, soccer bag, (um) a lipstick on Christmas Eve. TCC 13.1
C  I got a Nerfgun. TCC 13.1

___________________________________________________________Episode 9

E  A Nerfgun? TSE  37.1
C  Yeah. TCC 37.1
E  Where do you play with that? TCE 37.1
E  Can you play with that inside the house or? TCE 37.1
E  Yeah? TCE 37.1
E  Can you tell me about it (what it how) how it work/3s or what it look/3s
   like? TCE 37.1
C  (It) it has a laser gun that it show/3s you where you can shoot. TCC 37.1
E  Cool! TCE 37.1
C  And there/3s a back part a (f*) middle part and a front part. TCC 37.1
E  Mhm. TCC 37.1
C  And you can take part/s off. TCC 37.1
C  And (it) the gun still work/3s. TCC 37.1
E  Oh my gosh how does it still work? TCE 37.1
E  Wow that/3s pretty cool. TCE 37.1
E  So (where do you) do you play with your sister/s? TCE 37.1
C  Or who do you play with it? TCE 37.1
C  My sister/s. TCC 37.1
E  Yeah? TCE 37.1
C  And mama and papa. TCC 37.1
E  Oh wow.TCE 37.1
C  We play with those (um) bullseye/s. TCC 37.1
E  Oh cool. TCE 37.1
E  And it does/n't break or anything? TCE 37.1

___________________________________________________________Episode 10

E  Oh my gosh that/3s really cool. TSE  1.2
E  What are you make/ing? TCE 1.2
C  Mm. TCC 1.2
E  Yeah I don't know what I/3m making either. TCE 1.2