THE IMPACT OF PREMATURITY ON LANGUAGE SKILLS AT SCHOOL AGE

BY

JAMIE MAHURIN SMITH

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Speech and Hearing Science
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2010

Urbana, Illinois

Doctoral Committee:

Assistant Professor Laura S. DeThorne, Chair and Director of Research
Associate Professor Cynthia J. Johnson
Clinical Professor Charles Morton
Professor Ruth V. Watkins

ABSTRACT

This study compared school-age outcomes for 57 children born prematurely, at ≤32 weeks or <1500g, with outcomes for 57 children born at full term with no reported perinatal complications. The two groups were matched for age, sex, race, and parental education. Data came from the Western Reserve Reading Project and included discourse-level language samples collected at three points in time, each a year apart. In addition, standardized test results were obtained for IQ, digit span, and global language ability. The language samples were analyzed to yield a number of semantic and syntactic measures which were condensed via factor analysis to a semantic score and a syntactic score. Regression models showed statistically significant differences between the two groups for standardized test results, with more ambiguous results for the discourse-level language measures. The control group outperformed the premature group on both semantic and syntactic measures, but those differences never reached statistical significance and narrowed markedly at the third-year assessment point. These findings suggest that in the absence of frank neurological impairment, sophisticated semantic and syntactic skills may be relatively intact in the conversational and narrative language of children born prematurely. The decrements observed on standardized assessments of language and cognition may arise from deficits in domains such as attention or executive function, rather than reflecting significant impairment in their ability to learn language.

To the greater glory of God

and to my daughter

ACKNOWLEDGMENTS

I owe thanks to many people for their guidance and assistance. My advisor, Laura DeThorne, has provided an extraordinary balance of rigor and flexibility throughout and has taught me more than I ever expected to learn. I am grateful to my committee for its invaluable insights: to Cynthia Johnson for her knowledge of the literature and her steady encouragement, to Ruth Watkins for the inspiration to pursue a Ph.D., to Chuck Morton for the clinical knowledge that has assisted me in synthesizing medical research.

Thanks must also go to Ron Channell for his assistance with DSS, to Leonid Spektor for repeatedly and speedily repairing bugs in the CLAN code, and to Jessica Logan for her help with hierarchical linear modeling. I would also like to thank the crew of the University of Illinois Child Language and Literacy Lab, particularly Lisa Mellman, Elena Turner, Meredith Kresse, and Ellen Bogue. Additionally, this study would have been impossible without the participation of the WRRP families.

My family has borne up patiently during this project. I am grateful to my husband Martin for his willingness and his welcome sense of perspective. My sons JP, Andrew, Eliot, and Ben have been enthusiastic about this project and understanding about its demands; my daughter Eleanor has offered frequent unspoken reminders to prioritize. Without assistance from Marion Smith, Jenny Abraham, and Gianna Beretta Molla, this task would have been many times more difficult.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE: INTRODUCTION

At the end of pregnancy and the beginning of life outside the womb, brain development is taking place at a pace unequaled across the rest of the lifespan. When complications arise during a pregnancy, or when the birth process takes an unexpected turn, brain development can be affected: sometimes directly, as when birth asphyxia causes ischemic damage to brain structures, and sometimes indirectly, as when pathology of immature lungs impedes optimal oxygenation of the brain. For decades researchers have considered the long-term effects of such early complications, but the changing landscape of neonatology has meant that prognoses for babies in the neonatal intensive care unit (NICU) have shifted as well.

Much of the research into long-term outcomes for survivors of adverse perinatal events has focused on standardized measures of cognition, focusing on questions such as these: When a baby is born three months early, what is his or her IQ likely to be in kindergarten? How will that child perform on highly structured tasks in a formal testing situation? Few studies have looked explicitly at language outcomes, and even fewer have considered discourse-level abilities. What will these children be like as conversational partners? What kinds of storytellers will they become? The available research cannot provide adequate answers to these questions.

The study described here looked longitudinally at the conversational and narrative language abilities of a subset of participants in the Western Reserve Reading Project (WRRP), a population-based cohort of school-aged twin pairs. The study contrasted outcomes for a group with recognized perinatal risk factors (PRF) with outcomes for a second group, matched for age, sex, and parental education, whose perinatal course was uneventful. The sections that follow will explore the following areas to explain the rationale for this study: (a) the medical sequelae that most often occur in conjunction with preterm birth, particularly those associated with an increased risk of lasting impairment; (b) the long-term impact of adverse perinatal events, particularly on the development of language skills; (c) typical development of discourse-level language skills during the school years; (d) deficits, both global and language-specific, that might be anticipated in a group of children born prematurely, and (e) finally, support for the generalization of findings from twins to the general population.

Preterm Birth and Its Sequelae

Human gestation lasts, on average, 266 days from conception to delivery.[1] During the last trimester of pregnancy, tremendous changes take place within the fetus. Organs are maturing to prepare him or her to breathe independently and to take in and assimilate nutrients. The brain is developing at an astonishing rate, more than doubling in volume between 31 weeks' gestation and full term, with a fourfold increase in cortical gray matter volume (Hüppi et al., 1998). The body, too, is growing rapidly, from a mean weight of 1500g at 30 weeks' gestation (Kramer et al., 2001) to a mean weight of 3500g at full term (Ogden et al., 2002).

For 12% of all pregnancies in the US, and 17% of babies born to African-American mothers, delivery occurs before full term, or 37 weeks' gestation (Als & Butler, 2006). Broadly speaking, rates of morbidity (illness) and mortality (death) decrease as gestational age and birthweight increase. Babies born prior to 32 weeks' gestation are considered very premature; those born between 22 and 26 weeks' gestation are labeled extremely premature. The final month of

---

1 It may be useful for readers to note that pregnancies are traditionally dated from a mother's last menstrual period, adding two weeks to her baby's actual gestational age. Under this system, a pregnancy that actually began three weeks earlier is labeled a five-week pregnancy. Throughout this document, all gestational ages will refer to post-menstrual weeks and not to weeks post-conception.

pregnancy brings a surge in fetal weight gain, and babies born before that surge are virtually always smaller than their full-term counterparts. Here again, neonatologists have devised a classification system: babies born below 2500 grams, or 5 pounds 8 ounces, are called low birthweight; those born below 1500 grams, or 3 pounds 5 ounces, are considered very low birthweight (commonly abbreviated VLBW); babies with a birthweight below 1000 grams, or 2 pounds 3 ounces, are labeled extremely low birthweight (ELBW).

Babies born as early as 22 weeks' gestation can survive, but for these extremely preterm babies morbidity and mortality rates are astronomical: only 5% survive to be discharged from the hospital, and only 1% are free of moderate-to-severe disability as toddlers (Tyson, Parikh, Langer, Green, & Higgins, 2008). With advances in neonatal intensive care, overall mortality rates for premature babies have fallen; according to a 2008 review article, approximately 85% of babies born weighing less than 1500 grams survive to hospital discharge (Eichenwald & Stark, 2008). As more babies are surviving, however, the incidence of neurodevelopmental sequelae has remained constant, resulting in an increase in the population of impaired survivors of prematurity along with an increase in healthy survivors (Wilson-Costello & Hack, 2006).

It is important to note that even for a baby whose postnatal course appears unremarkable, preterm birth can have long-lasting effects on the brain (Peterson, 2003; Thompson et al., 2007). Als and Butler (2006) point out that babies with medically uncomplicated NICU stays still have increased risks of adverse developmental outcomes including, among others, learning impairments, decreased IQ, deficits in executive function, attention deficit disorder, and speech-language problems (see also Wilson-Costello & Hack, 2006). "It appears," Als and Butler state, "that development in the extra-uterine environment leads to different and potentially maladaptive developmental trajectories" (p. 1051). For example, differences in sensory input (such as the visual input that bombards babies in the extra-uterine environment but not those *in utero*) seem to affect the cortex differentially, potentially leading to differences in brain structure in infants born prematurely. It is reasonable, then, to be vigilant for neurodevelopmental differences in babies born prematurely, even when their NICU stays were uneventful.

When complications occur, however, they are often associated with an increased risk of poor outcomes. The sections that follow will review some common perinatal complications, which can arise either (a) prenatally, (b) during a baby's NICU stay, or (c) after discharge from the hospital.

Prenatal Complications

Three types of prenatal complications will be introduced here: inappropriate fetal growth, twin-to-twin transfusion syndrome (TTTS), and maternal hypertensive disorders. Gestational age and birthweight are closely correlated, but a baby can also be large for gestational age (abbreviated LGA), a circumstance seen frequently in babies born to diabetic mothers, or small for gestational age (SGA). This latter condition, SGA, raises red flags about a baby's well-being since it can result from a variety of causes including placental insufficiency, maternal behaviors such as drug abuse and alcoholism, and prenatal infection. Multiple gestation is often associated with SGA status; it is important to monitor these babies' growth and development closely since intrauterine growth retardation (IUGR) or SGA status at birth is associated with an increased risk of suboptimal neurodevelopmental outcomes (Prada & Tsang, 1998).

IUGR is only one of the risks associated with multiple gestation, which is widely acknowledged to be more dangerous for both mothers and babies than singleton pregnancies

(Senat, Ancel, Bouvier-Colle, & Bréart, 1998). Of all twin pregnancies that reach 20 weeks, half will end before full term: 10% before 28 weeks, 10% between 28 and 31 weeks, and the remainder between 31 and 37 weeks (Lumley, 2003). Approximately half of preterm births in twin pregnancies are medically indicated (Moutquin, 2003), and a commonly cited reason for doctors to end a twin pregnancy prematurely is concern about twin-to-twin transfusion syndrome (TTTS). In a monochorionic pregnancy, when twins share a single placenta, interconnecting blood vessels known as anastomoses can form within the placenta. Outcomes vary depending on the direction and volume of blood flow. When one twin receives a disproportionately small volume of the available blood supply and nutrients, growth retardation can result. The twin receiving the larger blood volume can experience cardiac complications leading to heart failure. For both donor and recipient twins, neurodevelopmental complications can result: inadequate nutrient availability may affect the donor twin, and the recipient twin's overtaxed heart may be unable to deliver nutrients to the brain as needed or carry away accumulated cellular waste products. TTTS is the most common serious complication observed in monochorionic twin pregnancies (Cincotta, Gray, Phythian, Rogers, & Chan, 2000).

Mothers as well as babies face increased risks in multiple gestation; the most common serious maternal complication is pre-eclampsia, an umbrella term for a family of hypertensive disorders or pregnancy. For mothers, pre-eclampsia is diagnosed in the presence of high blood pressure and protein in the urine; it can include an alarming array of symptoms, including maternal seizures and damage to the liver or kidneys. Pre-eclampsia poses similarly grave risks to babies. Part of the problem for babies is the underlying pathology of pre-eclampsia, thought to arise from poor blood flow between uterus and placenta (Shah, 2006); decreased placental perfusion (blood flow) can mean diminished oxygenation, nutrient delivery, and waste removal. Part of the problem is that the treatment for poorly controlled pre-eclampsia is delivery of the baby. Protecting the mother's health may necessitate a preterm delivery, with its attendant risks.

One possible underlying factor which should be mentioned briefly is in-vitro fertilization (IVF). In IVF, a woman’s ova are retrieved from her ovaries and fertilized outside her body. After three to five days of growth in a culture medium, the resulting embryos are transferred back into the woman’s uterus. IVF has been associated with a number of potentially relevant complications, including an increased risk of multiple gestation. Even in singleton pregnancies, IVF is associated with higher rates of maternal complications such as pre-eclampsia (Shevell, Malone, Vidaver, Porter, Luthy, Comstock, et al., 2005). Long-term follow-up of children conceived via IVF has shown that it is a risk factor for adverse neurodevelopmental outcomes including cerebral palsy (Strömberg, Dahlquist, Ericson, Finnström, Köster, & Stjernqvist, 2002).

Whether or not they are affected by prenatal complications, babies born prematurely must often navigate a rocky course during their stay in the neonatal intensive care unit (NICU). Six complications of prematurity associated with adverse effects on long-term outcomes will be reviewed briefly here: bronchopulmonary dysplasia (BPD), patent ductus arteriosus (PDA), retinopathy of prematurity (ROP), necrotizing enterocolitis (NEC), neonatal sepsis, and intraventricular hemorrhage (IVH). The purpose of this section is to provide an overview of each condition and of the potential etiological mechanisms, both direct and indirect, by which these early pathologies might affect later development. For all the diagnoses described here, it is possible that causality is indirect: the pathology described could simply be a marker for a sicker, more vulnerable baby who is at increased risk of central nervous system complications. The literature on speech-language outcomes beyond infancy, cited briefly here, is reviewed in depth in the section headed "The Long-term Impact of Adverse Perinatal Events."

NICU Complications

Bronchopulmonary Dysplasia (BPD)

Bronchopulmonary dysplasia (BPD) is the most common complication observed in premature babies (Eichenwald & Stark, 2008), whose lungs are not mature enough to meet their bodies' demands for oxygen. The associated treatment, artificial ventilation and supplemental oxygen, can lead to damage in the form of scarring, overinflation, or inflammation of the alveoli in the lungs, resulting in BPD. A diagnosis of BPD increases a child's risk of adverse long-term outcomes and it is an independent risk factor for speech-language delay at school age (Lewis et al., 2002). The causal mechanism at work is unclear, but the differences observed among children with a history of BPD could be a function of chronic hypoxia arising from inadequate tissue oxygenation, potentially leading to diffuse cerebral damage with global effects on speech-language development (Short et al., 2003). BPD may be treated with steroids, which been linked to an increase in brain abnormalities such as reduction of gray matter volume and periventricular leukomalacia (Barrington, 2001). In addition, BPD is associated with a higher rate of hearing loss and of a movement disorder featuring rapid, jerky movements of oral structures, both of which could contribute to speech-language deficits (Singer, Siegel, Lewis, Hawkins, Yamashita, & Baley, 2001).

Patent Ductus Arteriosus (PDA)

In addition to respiratory complications, premature babies are at heightened risk of some pathologies of the circulatory system, particularly patent ductus arteriosus (PDA). The fetal circulatory system includes a connecting structure called the ductus arteriosus, which allows for communication between the aorta and the pulmonary artery during gestation. The ductus arteriosus protects the fetal heart against overexertion by allowing most of the blood from the pulmonary artery to bypass the fluid-filled lungs during gestation. Normally, this structure is largely closed by the end of the first day of *ex utero* life, a change triggered by the initiation of respiration in order to prevent communication between the two arteries. In some premature babies, however, the ductus arteriosus remains open, a condition known as patent ductus arteriosus (PDA). This may be a function of their immature ductal closure mechanisms, or it might result from inadequate oxygenation resulting from prematurity (Zahka & Erenberg, 2006). When PDA occurs, blood that would normally be pumped through the lungs can instead flow into the aorta, interrupting the normal process of oxygenation. PDA in the neonatal period is associated with poorer language outcomes in the preschool years (Singer, Siegel, Lewis, Hawkins, Yamashita, & Baley, 2001). As with BPD, the causal mechanism is incompletely understood, but it may be the case that inadequate blood flow through the pulmonary artery or associated blood pressure changes could result in suboptimal brain oxygenation during a critical window.

Retinopathy of Prematurity (ROP)

Early difficulties with oxygenation may also be associated, perhaps surprisingly, with vision impairment: retinopathy of prematurity (ROP) is a common complication with documented importance in the literature on long-term outcomes. ROP is characterized by abnormal growth of the retinal vasculature, accompanied by the development of fibrous tissue that can cause retinal distortion or detachment. While the condition often resolves completely, it can also lead to lasting visual impairment or blindness. ROP occurs only in premature babies, whose retinal vasculature is immature, and its incidence decreases with increasing gestational age: it affects more than 90% of babies born at 24 weeks' gestation, and less than 30% of those born at >31 weeks' gestation (Phelps, 2006). A history of severe ROP serves as a marker for impairment across multiple domains among kindergarten-aged children with a history of VLBW (Msall et al., 2000). This developmental pattern may result in part from the direct effects of ROP; it is more difficult to reach age-appropriate visual-motor integration benchmarks with a significant visual impairment. It is also important to note that one factor contributing to pathological growth of the retinal blood vessels is exposure to high levels of supplemental oxygen. The presence of ROP, then, could also indicate a history of hard-to-manage hypoxia, with all its attendant long-term risks.

Necrotizing Enterocolitis (NEC)

Another relevant condition that might seem at first glance to be unrelated to neurodevelopment is necrotizing enterocolitis (NEC), a serious pathology of the bowel affecting up to 10% of VLBW babies. NEC is a condition in which bowel tissue necroses, leading in severe cases to bowel perforation and widespread infection. It is associated with a high rate of poor neurodevelopmental outcomes for 20-month-old children with a history of VLBW (Sonntag, Grimmer, Scholz, Metze, Wit, & Obladen, 2000). While it may seem unlikely that a short-term, albeit serious, disease of the bowel could contribute to long-lasting impairment of the brain, the authors propose multiple potential mechanisms for the observed relationship. Poor perfusion (blood flow), leading to hypoxia, could affect both the bowel and the brain and be a single underlying causal factor for both conditions. Alternatively, a state of shock, characterized by diminished blood flow throughout the body, could lead to hypoxic encephalopathy, or the body's own immune response could cause damage to the developing brain. It could also be the case that unmet nutritional needs could lead to subtle neurological dysfunction, either during the critical early window before 40 weeks' gestational age, or afterward, if injury to the bowel mucosa were to impede absorption of important trace nutrients.

Neonatal Sepsis

During the time that they are hospitalized, more than half of VLBW infants will be treated for sepsis (systemic infection), a serious condition that has been identified as a marker for later neurodevelopmental impairment (Adams-Chapman & Stoll, 2006). It is thought that the massive inflammatory responses mounted by the body in response to the infection may explain the association between sepsis and brain anomalies such as periventricular leukomalacia, which is described in the following section. Adams-Chapman and Stoll report that the production of inflammatory cytokines is associated with free radical production and subsequent damage to or death of neural tissue.

Intraventricular Hemorrhage (IVH)

Finally, many premature babies experience direct insults to the brain. Intraventricular hemorrhage (IVH), bleeding within the ventricles of the brain, affects up to half of babies born before 32 weeks (Als & Butler, 2006). Premature babies have more delicate blood vessels and an underdeveloped ability to regulate blood flow through the brain, leaving them more vulnerable to damage as a result of variable blood pressure (de Vries, 2006). IVH is graded from I through IV, with higher numbers reflecting more extensive bleeding and greater potential for lasting ischemic damage. While the language regions of the brain are unlikely to be directly affected by bleeding in and around the ventricles, it is important to note that even milder forms of IVH (grades I-II) are associated with poorer neurodevelopmental outcomes at age 20 months: it appears that an early insult can have long-lasting and diffuse effects (Patra, Wilson-Costello, Taylor, Mercuri-Minich, & Hack, 2006). A related type of brain lesion, periventricular leukomalacia (PVL), describes white matter injury or ischemia in the regions surrounding the ventricles. PVL is often associated with spastic diplegia (Ward & Beechy, 2003), a form of cerebral palsy with pronounced lower extremity involvement. It is familiar to many speech-language pathologists due to the prevalence of speech-language impairment among children with this diagnosis.

Moderating Variables

The preceding subsections have focused on medical complications that can compound effects of prematurity, but some attention should be given to two additional moderating variables that can influence the degree to which these complications will affect a child. A brief explanatory note may be in order here: mediating variables are those which exert a causal influence on an outcome, while moderating variables are those which can alter the strength or direction of a relationship between independent and dependent variables (Baron & Kenny, 1986). Some variables can serve as either or both. Baron and Kenny explain that "moderator variables specify when certain effects will hold; mediators speak to how or why such effects occur" (p. 1176). For an illustration contrasting the two types of variables, see Figures 1.1 and 1.2. (Further discussion of the variable illustrated, maternal responsiveness, appears in the section headed “Psychosocial Variables and Post-Infancy Outcomes.”)

The first moderating variable under consideration is sex, since it figures in any efforts to explain variable outcomes among preterm babies. Though the reasons are not well understood, it has long been observed that female babies appear less vulnerable to adverse perinatal events (Stevenson et al., 2000) and show greater resilience when they occur. Ingemarsson (2003) reported that male babies are more likely to be born prematurely, more likely to suffer fetal distress and acidemia (an index of labor stress) during birth, more likely to die following a premature birth, and more likely to be diagnosed with pulmonary hypoplasia (inadequate lung development). He further stated that during labor, female preterm babies have significantly higher levels of catecholamines, hormones secreted in response to stress that can protect against intrapartum hypoxia, than their male counterparts. When preterm birth is judged to be inevitable, prophylactic steroids are administered to enhance postnatal pulmonary function; Ingemarsson also noted that female babies respond better to this treatment than males do. Moreover, after delivery, male babies are more likely to suffer immediate neurological complications, and they have higher rates of long-term sequelae such as attention deficit disorder (Reijneveld et al., 2006). Although attempts to explain these observed sex differences remain highly speculative, it is prudent to consider sex as a moderating variable in studies of long-term outcomes.

In addition to sex, infant feeding patterns may moderate the impact of preterm birth. Human milk contains a number of nutrients with the potential to influence neurodevelopment, notably the long-chain polyunsaturated fatty acids known as docosahexaenoic acid (DHA) and arachidonic acid (AA); see Riordan (2005) for additional details. A 1992 study by Lucas and colleagues assessed IQ in 8-year-old children who had been born preterm. After correction for maternal education and social class, they reported an 8.3 point IQ advantage for those who had been fed human milk versus those who were fed infant formula. This difference of more than half a standard deviation is larger than that seen in studies of children born at term (cf. Kramer et al., 2008). One possible explanation for the sizable discrepancy observed by Lucas and colleagues is that premature babies do not accrete the fat reserves *in utero* that full-term babies do, and are more affected during infancy by variability in dietary fatty acid profiles. This hypothesis is supported by the findings of Farquharson, Cockburn, and Patrick (1992), who studied the composition of autopsy samples of infants' neural tissue and found that formula-fed premature babies had unusually low levels of DHA. Breastfeeding may also play a less direct role in outcomes by boosting immune function; NEC, for instance, is far less common in breastfed babies (Updegrove, 2004). Additionally, breastfeeding may interact with other environmental influences such as SES, conferring some protection against adverse outcomes associated with low SES. Daniels and Adair (2005) reported such an interaction in their study of breastfeeding and cognition and noted that the protective effect was particularly pronounced for low birthweight children. Taken together, these reports suggest that breastfeeding duration is another relevant moderating variable in studies of long-term outcomes for babies born prematurely.

The two moderating variables just described can each exert some influence on the extent to which a given child's NICU course will shape his or her long-term neurodevelopment. Most VLBW survivors, however, will spend far more time outside the hospital than in it, and moderating variables that can affect children after their NICU discharge must be considered as well.

Psychosocial Variables and Post-Infancy Outcomes

Long after their discharge from the NICU, outcomes for preterm babies may be shaped by psychosocial factors. In addition to the immediate impact of medical complications on a baby, a number of researchers have considered the impact of a NICU stay on parents. Their findings suggest that the effects of adverse perinatal events on parents and families, as well as on the affected babies, can be significant. Bryan (2003) surveyed the recent literature on the psychosocial sequelae of preterm multiple births, and reported a number of findings with implications for long-term outcomes. First, when preterm multiples go home after their hospitalization, their mothers are more vulnerable to depression and are less responsive to their children, as a group, than mothers of singletons. In addition, she noted that their siblings are more likely to develop behavior problems. The impact of maternal interaction on early language development is controversial, but it is plausible that diminished parent-child interactions arising from maternal depression or siblings requiring behavioral support could have a deleterious effect on language acquisition. (For additional information on maternal depression and its impact on mother-baby interaction, readers are referred to Field, 2002).

The impact of maternal responsiveness and depression could not be assessed within the WRRP cohort, but another important moderating variable, parental education, was. Studies of the interaction between biomedical and social risk factors have indicated that parental education can play a role in developmental outcomes (Msall, Bier, LaGasse, Tremont, & Lester, 1998; Holloman, Dobbins, & Scott, 1998). A more extensively educated parent may serve as a protective influence in the presence of perinatal risk factors, while a lower degree of parental education appears to be an exacerbating factor. The causal influences underlying this observed relationship are doubtless complex, but may be tied to factors such as community support for raising a disabled child, resource availability for therapies and medical follow-up, the quality of available childcare, and variations in language-based interactions during the early years.

To sum up, then, twins face substantially higher odds of prenatal problems, of preterm birth with all its attendant risks, and of postnatal complications as well, any one of which can be associated with impairments at school age. The following section will review the literature on long-term outcomes in an effort to specify and quantify those heightened risks.

The Long-term Impact of Adverse Perinatal Events

Overview

For at least 50 years scientists have been investigating long-term outcomes for adverse perinatal events, attempting to address questions that still trouble parents today (Lilienfeld, Pasamanick, & Rogers, 1955). How often might a difficult beginning foreshadow future impairment? What factors determine vulnerability and resilience? Will affected children "catch up" over time, or will the increasingly complex demands of the later school years widen the gap between them and their peers? Dozens of papers have been published addressing these questions, a majority of which have focused on global cognitive outcomes. Two considerations have shaped this literature review: first, the late 80s and early 90s brought changes in NICU treatment strategies that led to dramatic reductions in neonatal morbidity and mortality, notably the introduction of exogenous surfactant to improve lung function in premature babies, together with the use of antepartum steroids to aid in lung maturation (Schwartz, Luby, Scanlon, & Kellogg, 1994). To provide realistic estimates of effect sizes for children in the WRRP cohort and their contemporaries, all of whom were born after these changes had been implemented, this review will prioritize studies of children born in the early 90s and thereafter. Second, after a brief review of some recent data on global outcomes, it will focus more explicitly on studies of language function.

Global Outcomes

It is useful to address global outcomes briefly, to provide a frame of reference for the studies focused on language outcomes. Many studies have followed children with a history of VLBW; Johnson (2007) provides a review of recent research. The consensus is that VLBW survivors face an increased risk of myriad morbidities, including impaired neurodevelopment, deficits in vision and hearing, and behavior problems. Johnson notes that significant differences in IQ persist between very preterm children and controls even when researchers control for factors such as SES. ELBW children, even those free of overt neurosensory impairments, score an average of 10 standardized points lower on tests of cognition than their siblings do. As outlined in the preceding section, certain conditions, including SGA, IVH, BPD, ROP, and NEC, further increase survivors' risks of long-term impairment.

It is crucial to note, however, that early outcomes may have limited predictive power for any given child. Hack et al. (2005) followed 330 children with a history of ELBW, born between 1992 and 1995. Of the 238 children who survived to age 8, 200 underwent cognitive testing at both 20 months (corrected age), using the Bayley Scales of Infant Development, Second Edition (BSID-II; Bayley, 1993) and at 8 years, using the Kaufman Assessment Battery for Children (KABC; Kaufman & Applegate, 1988). In logistic regression models, test scores at 20 months did not serve as reliable predictors of cognitive function at 8 years for children with a history of ELBW, though better predictive power was noted for the children with neurosensory impairment. Significant neurosensory impairment, a label which included conditions such as cerebral palsy, shunt-dependent hydrocephalus, and unilateral or bilateral blindness or deafness, was identified in 46 of the children.

In a similar study, Koller et al. (1997) looked prospectively at patterns of cognitive change over time in 203 children with a history of VLBW, administering cognitive tests four times during their first six years: at age 1, 2, either 3 or 4, and either 5 or 6 years. At 12 and 24 months of age, the assessment tool was the Bayley Scales of Infant Development (Bayley, 1969); at the 3- or 4-year assessment it was the Stanford-Binet Intelligence Scale (Thorndike, Hagen, & Sattler, 1986). For 6-year-olds the investigators used the Wechsler Intelligence Scale for Children -- Revised (WISC-R; Wechsler, 1974). The researchers found that outcomes at age 6 varied quite a bit: 13% of the sample stayed in the average range throughout; 24% of the sample slipped from average to low-average cognitive performance; 43% moved from average to below average; 8% improved from very low to low average; and 12% remained stable in the very low range. In this cohort, the results of a neurological examination at 1 year were classified as normal (no atypical findings on a neurological exam), suspicious (minor anomalies in gait, muscle tone, or movement), or abnormal (reports of seizures, hydrocephalus, CP, blindness or deafness). Based on the results of discriminant analysis, the authors concluded that the results of this exam served as an important predictor for six-year outcomes, accounting for 84% of the variability between groups. They noted, too, that the trend was particularly clear for infants in the smaller end of the VLBW range.

In a more encouraging study, Ment and colleagues (2003) contrasted scores on the Peabody Picture Vocabulary Test - Revised (PPVT-R; Dunn & Dunn, 1981) for 296 VLBW survivors followed prospectively as part of an IVH-prevention trial, and found that the median standard score improved from 88 at 36 months to 99 at 96 months. The authors cited these results as evidence of resilience in a majority of children with a history of VLBW, though they noted that scores declined by approximately one standard deviation for the group of 8 children with a history of early-onset IVH and CNS injury.

Taken together, these longitudinal studies highlight the unpredictable nature of long-term outcomes. As a group, children born preterm are at increased risk of impairment at school age, but within that larger group many children will show an encouraging degree of resilience. Frank neurological complications are a robust negative prognostic factor.

Related Impairments in Cognitive Function

In addition to these findings from studies of cognitive development, a further noteworthy trend observed in school-aged children with a history of adverse perinatal events is an increased risk of behavioral and emotional problems (Reijneveld et al., 2006; Chapieski & Evankovich, 1997). Children with a history of prematurity and/or VLBW are more likely to be diagnosed with attention deficit/hyperactivity disorder and deficits in executive function. They are more likely to be evaluated by their peers as sensitive and isolated (Nadeau, Boivin, Tessier, Lefebvre, & Robaey, 2001). It is plausible that these well-established differences could contribute to impaired performance in other domains: IQ scores for a child with ADHD may reflect his attentional impairment more than his true cognitive abilities; diminished executive function could make it more difficult for a child to devise and relate a coherent narrative; a child perceived as a loner by his peers might have fewer opportunities to develop strong social language skills. For a review of recent research into cognitive and behavioral outcomes at school age, with considerations of the methodological issues involved, readers are referred to Johnson (2007).

Speech-Language Outcomes

A subset of studies has explicitly considered the effect of VLBW and adverse perinatal events on children's future speech-language abilities. Some of these studies have considered a possible correlation between adverse perinatal events and late emergence of language in toddlers; others have followed children through the preschool years and beyond. Only brief consideration will be given here to studies of very young children, because this study addressed school-aged outcomes.

Early Speech-Language Outcomes

Even in the general population the relationship between late talking and language skills at school age remains controversial (Paul, 1996; Nippold & Swartz, 1996). For VLBW survivors, given the research that suggests non-linear growth between the preschool years and middle childhood (Ment et al., 2003), studies of preschool abilities are unlikely to have much predictive value for school-age outcomes. Interested readers are referred to Zubrick, Taylor, Rice, and Slegers (2007) for a recent example and a review of related studies. These authors found that preterm birth (<37 weeks' gestation) or suboptimal growth (<85% of expected birthweight) were associated with almost twice the risk of late language emergence at age 2.

One study of early outcomes addresses some questions which are particularly relevant to the present study because of its conclusions about the interplay between genetic and environmental variables in children born prematurely. Koeppen-Schomerus and colleagues (2000) reviewed data from the Twins Early Development Study, a prospective study of all twins born in England and Wales during 1994. Prior reports on the cohort had excluded any children born very prematurely, but these authors looked explicitly at outcomes for children born preterm. They divided the sample into a high-risk group (born between 25 and 31 weeks' gestation, including

5% of the total sample of 2223 twin pairs), a moderate-risk group (born at 32 or 33 weeks' gestation, including 8.6% of the total sample), and a low-risk group (born at or above 34 weeks' gestation, including 86.4% of the total sample). Two parent-reported measures of children's abilities were used to measure the children's cognitive and language skills at age 2: the MacArthur Communicative Development Inventory (Fenson et al., 1991) and the Parent Report of Children's Cognitive Abilities (Saudino et al., 1998). Standard maximum likelihood modeling was used to derive estimates of the relative importance of genetic and environmental influences. For the moderate- and low-risk groups, heritability accounted for a significant fraction of the overall variability in language skills (33% and 22%, respectively, with shared environmental influences accounting for 65% and 73%). In the high-risk group, however, shared environmental factors explained 84% of the variance in language outcomes while heritability accounted for only 9%, a value that did not reach statistical significance. For children born prior to 32 weeks' gestation, the authors concluded, environmental influences are far more important than genetic heritage in determining language outcomes at age 2.

Later Speech-Language Outcomes

While the Koeppen-Schomerus paper raises intriguing issues, studies of older children are of more immediate import to the present investigation. In their studies of later-developing speech-language skills, researchers have most often approached the question of speech-language skills in VLBW survivors using one of two methodologies: either prospective studies of VLBW cohorts that document the range of speech-language outcomes within the cohort, or retrospective studies assessing for overt language disability in children with a history of VLBW. This distinction is significant because there tend to be differences between the two types of studies. Generally speaking, prospective studies in this population are slower, more costly, and more logistically challenging because of the need to follow a cohort from birth onward; retrospective studies, on the other hand, are more vulnerable to bias and confounding. Cohort composition may vary across methodologies as well, with the prospective studies typically including children with a wider range of cognitive abilities and family SES. Outcome measures also vary, with most of the retrospective studies assessing explicitly for specific language impairment (SLI). Here the prospective studies are presented first, followed by the retrospective studies.

Prospective Studies of Later-Developing Speech-Language Abilities in VLBW Survivors

Researchers at Case Western Reserve University have been investigating long-term outcomes in multiple domains for children with a history of very low birthweight and bronchopulmonary dysplasia (BPD) as compared with outcomes for control children (Singer, Siegel, Lewis, Hawkins, Yamashita, & Baley, 2001; Lewis et al., 2002; Short et al., 2003). In a 2001 study, the authors contrasted outcomes at age 3 for three groups of children: one group of 122 children with a history of VLBW and BPD, a second group of 84 children with VLBW and no BPD, and a third control group of 123 children, born at full term. The Bayley Scales of Infant Development II and the Battelle Developmental Inventory's Communication Domain Subscale were administered to all participating children. Children whose history included both VLBW and BPD scored significantly lower on all Battelle domains (an average of 7 standardized points on both receptive and expressive subtests) than children with VLBW alone, whose receptive scores on the Battelle were significantly lower (6 points, on average) than those obtained by control children. Children with a history of BPD had a significantly higher risk of scoring in the impaired range than children in either of the other two groups, with group means falling in the low end of the normal range. After controlling for IQ, the researchers found that only the differences in receptive language function remained significant. Multiple regression analyses revealed that the following variables were significant predictors of poorer language outcome, collectively accounting for 21% of the variability in outcomes: history of patent ductus arteriosus (PDA), minority race, lower SES, and greater neurologic risk (history of intraventricular hemorrhage or seizures). A history of PDA, reported for 56% of the BPD group and 18% of the VLBW group, was the single strongest predictor of impaired language skills, though not of impaired cognitive or motor skills, a finding which led the authors to suggest that PDA might have a language-specific effect although the mechanism was not specified.

Singer and colleagues concluded that a history of VLBW and BPD was associated with an increased risk of specific receptive language delay at age 3, even beyond deficits associated with impairment in global cognition. They suggest three possible explanations for their finding: first, that the Battelle may be particularly sensitive to receptive deficits, since children in all three groups scored lower on the receptive scale than on the expressive scale; second, that vulnerability to BPD and to receptive language deficits might have a similar genetic foundation; and third, that the results might reflect underlying deficits in auditory or other sensory processing abilities, in which case expressive language skills might be less affected. A further alternative, not addressed by the authors, is that attention deficits in the cohort could have played a part in their lower receptive language scores; this possibility should be considered whenever receptive deficits are reported in the absence of expressive delays.

The speech-language skills of this same cohort were assessed again at age 8 (Lewis et al., 2002). In the five years between the two studies, some children were lost to attrition; at the 8-

year follow-up there were 89 children with a history of BPD and VLBW, 71 with a history of VLBW without BPD, and 93 control children. Speech-language abilities were evaluated using the Goldman-Fristoe Test of Articulation (GFTA; Goldman & Fristoe, 1986) and the Clinical Evaluation of Language Fundamentals-3 (CELF-3; Semel, Wiig, & Secord, 1995). In addition, the Test of Oral Structures and Functions (TOSF; Vitali, 1986) was used to assess oral motor abilities, the children's fine and gross motor skills were evaluated via the Bruininks-Oseretsky Test of Motor Proficiency (BOMT; Bruininks, 1978), and the Weschler Intelligence Scale for Children-III (WISC-III; Wechsler, 1991) was administered to measure their cognitive skills.

The BPD group scored significantly lower on the GFTA than the other two groups, which did not differ statistically from each other. On all subtests of the CELF-3, the BPD and VLBW groups scored significantly lower than the control group. The VLBW group scored slightly higher than the BPD group on the Concepts and Directions subtest, a measure of receptive language, but the CELF-3 subtest scores for these two groups were otherwise statistically indistinguishable. Similarly, the BPD and VLBW groups scored significantly worse than the control group on verbal IQ and on the test of oral motor skills, but the two experimental groups did not differ from each other on these measures. On performance IQ and full-scale IQ, all three groups were significantly different, with the BPD group scoring lowest and the control group scoring highest. This same outcome was observed for fine and gross motor skills. The authors also noted that the BPD group included the highest percentage of children then receiving speech-language therapy (48%), followed by the VLBW group (21%) and then the control group (9%). They suggested that speech-language therapy would diminish group differences on language measures, citing this as evidence of the robustness of their findings. As in their 2001 study of these children at age 3, a history of BPD was linked to later receptive language impairment and to deficits in articulation skills. All children with a history of VLBW, whether or not their history included BPD, had lower scores on measures of language, IQ, and motor skills. It is noteworthy that across measures, both sets of experimental group means tended to fall in the lower end of the normal range and not outside it, suggesting a trend toward subtle deficits rather than overt impairment.

Rutter and colleagues (2003) considered obstetric risk factors in their population-based study of mild language delay. The authors identified 98 twin pairs and 96 age-matched singletons, each of whom had a singleton sibling not more than 30 months older. Only one child from each twin pair was assessed, to circumvent the problem of non-independent observations. The children were assessed using the MacArthur Communicative Development Inventory (Fenson et al., 1991) at 20 months, as well as the Preschool Language Scales - 3 (PLS-3; Zimmerman, Steiner, & Pond, 1992) and the McCarthy Scales of Children's Abilities (McCarthy, 1972) at 36 months. The authors interviewed families to obtain information on sociodemographics and a wide array of perinatal complications, and measured correlations between perinatal risk factors and language skills at 20 and 36 months. Neither birthweight, gestational age, nor a history of obstetric complications explained differences in language outcomes. It is important to note, however, that their original analyses excluded the 19 children in the cohort who were born before 33 weeks' gestation. When results for those children were included, 6 of the 19 were found to have significantly impaired language. For the remaining 13, language outcomes were described as "marginally lower" than those of the larger sample (no further details were available in this report), lending additional support to the trend toward subtle deficits observed in other prospective studies of this population.

Two Scandinavian studies (Luoma, Herrgård, Martikainen, & Ahonen, 1998; Jennische & Sedin, 1999) also looked prospectively at groups of children born preterm between 1980 and 1986. Because these data are older they will not be reviewed in detail here, but both sets of authors reported findings similar to those reported in more recent prospective studies: a tendency toward subtle difficulties with language in children born prematurely, but not a pattern of overt language impairment.

Retrospective Studies of Later-Developing Speech-Language Abilities in VLBW Survivors

While prospective studies have shown a clear correlation between very low birthweight and an increased risk of various types of neurodevelopmental delay, including lower scores on language measures, language researchers looking retrospectively have not reported a link between adverse perinatal events and specific language impairment (SLI), a controversial diagnosis in which performance IQ and sensory abilities are normal but language ability is significantly impaired. These studies will be reviewed in chronological order, from oldest to most recent.

Aram and colleagues (1991) retrospectively investigated the impact of very low birthweight on speech-language outcomes in their comparison of 249 VLBW children with 363 normal birthweight children, all born between 1977-1979 and assessed at 8 years of age. Children in both groups completed a battery of standardized tests including the Peabody Picture Vocabulary Test - Revised (PPVT-R; Dunn & Dunn, 1981), two subtests of the Token Test (DiSimoni, 1974), the Rapid Automatized Naming Test (RAN; Denckla & Rudel, 1976), a subtest of the Clinical Evaluation of Language Fundamentals (CELF; Semel-Mintz & Wiig, 1982), the Photo

Articulation Test (PAT; Pendergast, Dickey, Selmar, & Soder, 1984), a test of diadochokinetic rate (Fletcher, 1978), and the Wechsler Intelligence Scale for Children - Revised (WISC-R; Wechsler, 1974).

Neurological assessments revealed major neurological impairments (e.g., CP, hydrocephalus, severe hearing loss) in 24 of the 249 VLBW children, and the results of group comparisons varied if those 24 children were included in the analysis. When those scores were considered, Bonferroni-corrected group differences were significant for every measure except the PPVT-R and the test of diadochokinetic rates, with the control children outscoring the VLBW children by roughly one-third of a standard deviation across the board. When these 24 children were excluded, corrected group differences were smaller but remained significant for the WISC-R, the Token Test subtests, the RAN, and the CELF. In addition to looking at results for individual measures, the researchers contrasted language outcomes and cognitive results to evaluate the prevalence of SLI. Within the group of VLBW children, the authors looked at a wide range of perinatal risk factors, including type of delivery, APGAR scores, respiratory distress, apnea episodes, and NEC, among others, to evaluate their potential role as causal factors for SLI. They defined SLI initially as a discrepancy of >1 SD between performance IQ (PIQ) and any of the five measures of expressive or receptive language together with normal cognition ($PIQ \geq 85$), normal hearing in at least one ear, and no significant neurological impairments; following their initial analysis, they re-analyzed the data looking for a discrepancy of >2 SD between PIQ and language function. Using chi-square analyses and *t* tests to compare outcomes for the VLBW group and the control group, they concluded that perinatal risk factors were not associated with SLI in the VLBW cohort. Furthermore, the authors reported a higher incidence of SLI among the control children than among the VLBW children: 45.7% of control children vs. 33.7% of VLBW children using their initial definition, and 18.7% of control children vs. 14.5% of VLBW children using their revised definition. These values far exceed the incidence of SLI in the population as a whole, reported at 7% (Leonard, 1998; Tomblin, Smith, & Zhang, 1997), perhaps because a child who performed poorly on any of five subscales could have been diagnosed with SLI under the criteria used here. The investigators concluded that VLBW is a risk factor for pervasive developmental problems that encompass language deficits, but not for SLI. They noted, too, that their reliance on word- and sentence-level assessments was a limitation of their study, and suggested that the inclusion of discourse-level tasks might have revealed additional deficits.

Discourse-level abilities were considered in a study by Feldman, Janosky, Scher, and Wareham (1994). They investigated the possibility of a link between perinatal risk and specific language impairment in their study of 18 3-year-old boys born prior to 36 weeks' gestation: 6 with CP and periventricular brain lesions, 6 with similar brain lesions but no motor impairment, and 6 preterm control children, born prematurely but with no evidence of brain injury. Selection criteria for the study included normal hearing, the absence of any seizure disorder, and a score above 80 on the McCarthy Scales General Cognitive Index (McCarthy, 1972). This study is unique in its use of language sample measures: the children and their parents talked together for approximately 20 minutes in the children's clinic of a Pittsburgh hospital. Investigators measured the children's mean length of utterance (MLU), evaluated their number of total words per minute and number of different words per minute, and derived a score for the Index of Productive Syntax (IPSyn), a measure of syntactic complexity. They found no significant group differences on any of the measures, although they did report concerns about outliers in each group. While this study is intriguing in its use of language sample measures, it is limited both by its small sample size and by its inadequate descriptions of children's medical histories. A child born at 25 weeks, for instance, faces many more challenges than a child born at 35 weeks and yet this study does not distinguish between them, relying instead on a 36-week cutoff for its definition of prematurity. It does, however, provide some additional evidence that early neurological insults are not associated with specific language impairment in particular.

Bishop (1997) looked at risk factors for specific speech-language impairment in a retrospective twin study, which compared 84 twin pairs in which one or both twins had SLI or an articulation disorder (total affected $n = 122$) with 36 twin pairs in which neither twin had any history of speech-language impairment. She divided affected participants into four subgroups based on their impairment profile: articulation disorder ($n = 22$), expressive language disorder ($n = 65$), expressive language disorder with articulation impairment ($n = 24$), and pure receptive language disorder ($n = 11$). (Receptive deficits were also present in at least some of the individuals diagnosed with expressive language disorders, but little information was available on deficit profiles.) Intriguingly, Bishop found evidence of genetic influences for all but the last subgroup; no familial factors were observed for pure receptive disorder.

Bishop's exploration of the distribution of perinatal risk factors in her experimental and control groups yielded no significant group differences. She reported only one weak trend: an association between toxemia and speech-language impairment that failed to reach statistical significance. It should be noted that this cohort as a whole experienced few perinatal complications, with a mean gestational age of 36 weeks at delivery and a mean 5-minute

APGAR score slightly above 9 out of a possible 10; it is more difficult to address the impact of perinatal risks in a relatively low-risk group. Bishop addressed the discrepancy between prospective studies that show an increased risk of language deficits in children who experienced adverse perinatal events, and those retrospective studies, focused narrowly on language outcomes, that have found no predictable relationship. She suggested that the etiological factors at work in specific speech-language impairment might differ from those responsible for general developmental delays, and this disparity could account for the seemingly contradictory findings. In other words, global neurocognitive impairment arising from early insults could certainly be associated with language deficits. Frank language impairment in the presence of typical cognitive skills, however, may fall in a different diagnostic category.

Tomblin, Smith, and Zhang (1997) considered perinatal risks in their population-based case-control study of 177 kindergarten-aged children with SLI and 925 control children. Children's language skills were assessed using the Test of Language Development-2: Primary (Newcomer & Hammill, 1988); performance IQ was evaluated via the Wechsler Preschool and Primary Scale of Intelligence -- Revised (Wechsler, 1989). Families were interviewed at length to collect information on a number of potential prenatal and perinatal risk factors for SLI, including delivery complications and low birthweight. Logistic regression was used to find odds ratios and confidence intervals for all risk factors under consideration. There was a trend toward a higher incidence of low birthweight in the SLI group (4.7% of cases versus 2.8% of controls), but the odds ratio of 1.7 was not statistically significant (95% CI = 0.8-3.8).

In contrast to the authors of the preceding studies, Briscoe, Gathercole, and Marlow (1998) described preterm birth as an important risk factor for SLI in a "sizable minority" of children.

They compared outcomes for 26 children born at 26-32 weeks' gestation (mean GA = 28 weeks) with results for 26 children born at term. The children were tested at ages 3;0-4;0 using the British Picture Vocabulary Scales (Dunn, Dunn, Whetton, & Pintilie, 1982), the Oral Vocabulary subtest of the McCarthy Scales of Children's Abilities (McCarthy, 1972), the Bus Story Test of Continuous Speech (Renfrew, 1985), a digit-span task and a non-word repetition task to assess phonological short-term memory, and Raven's Progressive Coloured Matrices (Raven, 1977) to measure nonverbal skills. For the preterm group, results were also available from the Griffiths Mental Development Scales (Griffiths, 1974), administered at 12 and 24 months. The Bus Story Test is particularly relevant because it assesses narrative language. Children listen to a story and retell it using picture stimuli; their transcribed stories are scored for content and sentence length.

For all measures, the preterm group performed more poorly than the full-term group, with their scores typically falling half of a standard deviation below those of the full-term children. These deficits were only statistically significant for two measures, however: the British Picture Vocabulary Test and the content subscale of the Bus Story Test. Scores on the Bus Story Test were not uniformly lower for the preterm group; rather, the distribution was characterized by a cluster of scores at the extreme low end of the scale. Using a criterion developed by Bishop and Edmundson (1987) to pinpoint children at risk for SLI, the authors divided children into "at-risk" and "no-risk" subgroups based on results from the Bus Story Test content subscale. The "at-risk" subgroup included 8 of the 26 preterm children and none of their full-term counterparts. The researchers compared perinatal histories for the "at-risk" and "no-risk" preterm children, reviewing the incidence of a number of perinatal events including fetal distress, respiratory illness, intracranial lesions, and rates of infection, along with gestational age, birthweight, and length of NICU stay. They found no predictable relationships between adverse perinatal events and at-risk status, though a trend was noted toward lower birthweight and increased need for supplementary oxygen among the lowest-scoring children. They emphasized that the small size of their groups diminished the study's statistical power and encouraged further research with larger cohorts. The authors concluded that prematurity on its own does not cause specific language impairment, but that events associated with prematurity may raise the risk of SLI in vulnerable children, perhaps as a result of deficits in phonological short-term memory. To reconcile the apparent discrepancy between their conclusions and those of the preceding studies, none of which found an association between perinatal risk factors and SLI, it is useful to note that the other studies cited here all diagnosed actual cases of language impairment while Briscoe and colleagues only noted the existence of a risk factor for the diagnosis at preschool age.

In summary, then, it appears that a history of adverse perinatal events is not clearly associated with specific language impairment in retrospective studies. At the same time, however, subtle but measurable language deficits, along with global delays, have been reported in many prospective studies of VLBW survivors when they are compared with their full-term counterparts. Two reasons suggest themselves for this apparent discrepancy: first, an SLI diagnosis may not be the most useful outcome measure in the population under consideration. Since SLI is defined as language impairment in the presence of normal cognition and sensory skills, and since adverse perinatal events are associated with both cognitive and sensory deficits, true SLI might be, as Aram reported (1991), less prevalent in children born prematurely than in the population as as whole. Second, the prospective studies may cut a wider swath in terms of both sociodemographics and children's abilities. Because they follow all the children born in a particular hospital within a particular window, they may include more low-SES families and more significantly impaired children than the retrospective studies. This will, of course, vary with recruitment and follow-up strategies.

A final trend is evident in this review of the existing literature: most of the available data have been obtained from standardized tests, with little emphasis on naturalistic language tasks. Briscoe, Gathercole, and Marlow (1998), with its inclusion of the Bus Story Test, is one exception; Feldman, Janosky, Scher, and Wareham (1994), which featured language sample data, is the other. Both of these studies looked at the abilities of 3-year-olds, leaving questions about discourse-level language skills at school age unanswered. Given this gap in the available literature, the task of quantifying these group differences in samples of conversational and narrative language has the potential to provide a valuable addition to our understanding of the development of spoken language. In order to tackle the question of subtle deficits in language skills at school age, it is necessary to review the process of typical development.

The Importance of Naturalistic Language Tasks

For NICU parents concerned about their children's future performance on standard tests, ample evidence is available. For NICU parents concerned about their children's abilities to participate normally in day-to-day conversations, or to tell appealing and culturally appropriate stories, far less research exists on long-term outcomes. Factor analysis comparing outcomes on standardized language tests with outcomes on conversational language measures indicates that the two types of tasks load on different factors (Mather & Black, 1984; DeThorne et al., 2008), tapping different sets of skills. Real-life examples of this difference abound: a student whose

SAT verbal score was perfect but who struggles in ordinary conversation, or a student who devises engaging and complex stories but who freezes up in a formal testing situation.

Standardized testing levies a number of different requirements, and test results may reflect a child's struggles with any one of those, rather than deployment of his or her true language abilities in everyday contexts. Gipps (1999) highlights a number of factors which can affect scores on standardized measures. For instance, a child may do better or worse on a test depending on how motivating he or she finds the tasks and stimulus materials. Children's levels of test anxiety can vary widely, and performance may vary as a result. A formal testing situation can test a child's frustration tolerance, and willingness to comply with adult demands, and his or her eagerness to please and perform (see Speltz, DeKlyen, Calderon, Greenberg, & Fisher, 1999). Finally, Gipps points out that cultural considerations can be important, since in some cultures it is considered rude for children to tell an adult something he or she already knows. Even when a child is willing and able to participate in testing, standardized tests vary in terms of sensitivity and specificity. A review of widely used language assessments (Spaulding, Plante, & Farinella, 2006) raised questions about their utility in identifying areas of disability consistently.

Alternatively, language sample analysis can be a rich source of information about a child's language abilities in a less formal context. How much of his vocabulary consists of unusual words? Is she stringing clauses together with repeated use of "and" or "then," or are more complex conjunctions like "however" appearing in her speech? Are the sentences dense with phrases and ideas, or do they tend toward sparseness? Language sample measures are multifaceted, since they may be affected by personality traits such as introversion or by the connection (or lack thereof) between an examiner and a child, but they offer a revealing window into a child's capacities and they possess an inherent social validity for language use that is absent from the standardized testing milieu. Regardless of context, meaningful evaluation of spoken language hinges on an understanding of the process of language development, which will be reviewed in the following section.

Typical Development of Semantic and Morphosyntactic Skills in School-Aged Children

Semantic Development

During the early years of language development, the rapid pace of expressive vocabulary acquisition for typically developing children is a source of surprise and delight for many parents. On average, a child's vocabulary grows from a few spoken words at age 1, to approximately 40 words by 16 months, to an average of 570 words by age 2½ (Fenson et al., 1994). Few parents recognize that vocabulary growth becomes even more rapid, at least in absolute terms, as school-aged children acquire new words via reading in addition to spoken language. It is estimated that typically-developing school-aged children learn between 10 and 13 new words each day, or 3,000-5,000 per year (Nagy, Anderson, & Herman, 1987; Nippold, 1998). Later-developing vocabulary is characterized by its increasing proportion of abstract terms which lack any concrete referent. While a young child can look at a picture of an alpaca and infer that "alpaca" means a long-necked long-haired hoofed animal, an older child has a more complex task ahead as he or she tries to decipher abstract terms like "perception" or "unjust." As children's vocabularies grow, they learn increasing numbers of words that are synonyms and close cousins; an important component of early semantic development is untangling the nuances that distinguish closely related words. In addition to different words with similar meanings, English abounds in polysemous words, which have multiple distinct meanings. One example is "set," a word with 100 different definitions at one online dictionary (http://dictionary.reference.com, retrieved 9/10/08). As children learn that different words can mean similar things, they continue to learn that the same word can mean very different things. The preschooler who could distinguish between instructions to *set* the table and pick up the Lego *set* grows into a school-aged child who understands that *set* can refer to part of a tennis game, to the configuration of a ship's sails, or to what happens at grandma's weekly beauty shop appointment.

An additional characteristic of school-aged vocabulary development is specialization. Since children acquire words largely through reading and conversation (Nippold, 2007; Beals & Tabors, 1995; Beals, 1997; Weizman & Snow, 2001), different interests can lead to different vocabularies: "cutlass" and "moiety" for a child who reads old-fashioned adventure stories; "theropod" and "minmi" for a dinosaur enthusiast. Family culture also plays a role in shaping vocabulary, so that a child might be familiar with the word "wasabi" but puzzled by "scrapple," or vice versa. During the preschool years, parents are routinely asked to complete the MacArthur Communicative Development Inventory, an instrument that asks them to report which of roughly 700 words their child can produce. But it is difficult if not impossible to imagine a comparable instrument that could capture the breadth and diversity of school-aged vocabulary development. For this reason, as will be discussed further in chapter two, one useful strategy for analyzing semantic development in children this age is comparing their use of low-frequency vocabulary (Weizman & Snow, 2001; Marinellie & Chan, 2006; Beals, 1997, Beals & Tabors, 1995).

In addition to a acquiring a wide variety of root words, English-speaking children expand their vocabularies via derivational morphemes, which are especially common in academic reading materials (Nippold & Sun, 2008). The English language includes more than a hundred affixes that can be used to form new words (Nippold, 2007). Some of these are acquired early, as evidenced by a two-year-old who shouts "*Un*-eat it!" on discovering that the last slice of cake has been consumed. Others, such as the *-ent* suffix that transforms "reside" into "resident," emerge much later. During the school years, typically-developing children are broadening their ability to use affixes such as *-ness, -ship, -ful, -able,* along with many others (Anglin, 1993).

Syntactic Development

Along with their enormous vocabulary growth, school-aged children are acquiring new syntactic skills as well. Some of this change is evident in their increasing utterance length, evident both in conversational and in narrative language. Leadholm and Miller (1992) reported a steady increase in mean length of utterance for children from 3 to 13 years of age, with consistently longer utterances used for narratives than for conversation.

Syntactic development is evident not only in the growing length of children's utterances, but also in their density. Scott and Stokes (1995) itemized some changes of interest in this population, highlighting the types of phrases that typically-developing children acquire at school age. Adjectives become adjective phrases, with "huge" expanding into "unbelievably huge." Adverbs, too, can become adverb phrases (e.g., "very quickly," not just "quickly"). Eisenberg and colleagues (2008) found a steady increase in the frequency and complexity of children's noun phrases over the grade school years. The simple label, "dog," is replaced by "the brown dog," which is supplanted by "the brown dog with black spots who dug up my mother's tomatoes." Children's speech also begins to incorporate more sophisticated verb forms, such as passive voice and perfect aspect (Nippold, 1998).

In addition to within-sentence changes, also called intrasentential growth, intersentential development is also occurring in school-aged children (Mentis, 1994). Adverbial conjuncts are increasingly used as connectors between two similar sentences ("so then I...") or as bridges between dissimilar statements ("anyway," "on the other hand"). Intersentential transitions are particularly important in writing tasks (consider the frequency of words like "however" and "similarly" in scholarly writing), and growth in this domain continues across a child's years in school and into early adulthood (Nippold, Schwarz, & Undlin, 1992).

These changes are observed both in children's conversational speech and in their maturing narratives. A number of developments have been reported in the storytelling abilities of school-aged children (Botvin & Sutton-Smith, 1977; Roth & Spekman, 1986). Their stories become longer and more detailed, with more sophisticated plots and a greater degree of cohesion among plot elements. Characters become more three-dimensional, and children pay more attention to describing characters' thoughts and feelings (see Nippold, 2007). In addition to these changes in story elements, children's stories also undergo stylistic changes. Children who employ a literate language style in their narratives are typically judged as producing more mature stories than children who do not. Greenhalgh and Strong (2001) described four features that have been identified in connection with a literate language style: use of conjunctions other than *and*, use of adverbs, use of mental and linguistic verbs (e.g., think, know, tell), and use of elaborated noun phrases (e.g., "the blond girl whose birthday was Tuesday," rather than "the girl"). Gradually, over the course of the school years, children move away from the terse and often confusing narratives of early childhood, toward the more sophisticated, engaging, cohesive types of stories that characterize mature storytelling ability.

Language Development in Twins

Given the focus on twins in this study, a specific note regarding language development in this population is warranted: it is important to address the concern that causal influences on language acquisition for twins differ from causal influences for singletons. While twins may develop language more slowly than their singleton counterparts, particularly at younger ages, this discrepancy can be explained in large part by environmental variables such as perinatal insults and later patterns of family interaction (Thorpe, Rutter, & Greenwood, 2003). Such factors may be more common or more pronounced in twins, but are by no means irrelevant to singletons (e.g., Luke & Keith, 1992). Since no evidence suggests that distinct causal factors are at work in twins' language deficits, results from twin studies are assumed to generalize to broader populations (cf. Evans & Martin, 2000). In addition, the increased prevalence of adverse perinatal events in a twin population makes it especially appropriate.

Of particular relevance to this study, the Western Reserve Reading project provided the first large-scale study of conversational language in twins (DeThorne et al., 2008), with a focus on both vocabulary and syntactic development. Findings indicated that approximately half of the variance in such measures could be explained by genetic factors, with no significant shared environmental influences (factors experienced by both twins, such as home environment). The findings raise provocative questions in light of the literature reviewed here. A number of potential explanations were considered in the attempt to reconcile the absence of shared environmental effects in DeThorne et al. (2008) with the literature on perinatal risk factors. One possibility was that children with serious perinatal complications constituted a small enough fraction of the WRRP population-based sample that these environmental effects could not be detected with the statistical approach used by the authors (see, for instance, Koeppen-Schomerus, Eley, Walke, Gringras, & Plomin, 2000). It might also be the case that discordance in perinatal outcomes (e.g., only one twin suffers an intraventricular hemorrhage) could explain why no shared environmental effects were reported, since disparities in the perinatal course of twin pairs are not infrequent and such factors would constitute nonshared rather than shared environmental effects. A third possibility was that the increased risks associated with monochorionic or monoamniotic pregnancies (Stromswold, 2006) might cause the monozygotic twins in the sample to resemble each other more closely than the dizygotic twins, in which case the variance could be mistakenly attributed to genetic effects. Alternatively, the language measures used in previous studies of this cohort might be less sensitive to subtle variations in older children's semantic and syntactic development.

Two other possibilities must also be entertained, neither of which could be dealt with definitively by the present study: genetic effects on conversational language skills might be so robust that they trump environmental factors. This would contrast dramatically with the findings of longitudinal studies of premature babies, but those studies tend to rely on standardized tests, which tap somewhat different skills. Finally, it could be the case that unspecified positive environmental factors (potential influences might include parent-child interactions, optimal nutrition, or high-quality educational experiences) are mitigating the negative impact of perinatal complications in a substantial portion of this sample, with the result that no significant shared environmental effects are detected. Whatever the answer may be, the questions are intriguing.

Rationale for the Present Study and Summary of Hypotheses

The available literature suggests that for groups of school-aged children, performance on language tasks is likely to be subtly impaired for those who experienced adverse perinatal events versus those whose perinatal course was unremarkable. The present study looked longitudinally at language skills in conversational and narrative tasks for children ranging from first to fourth grade.

The study explored the following research questions:

1. How do the semantic and morphosyntactic skills of children in the PRF group compare with those of their peers in the control group, as assessed at each of three annual home visits?

It was hypothesized that the PRF group as a whole would exhibit more impoverished vocabulary and morphosyntax for both conversational and narrative language tasks as compared with children in the control group. It was not anticipated that the PRF group as a whole would fall outside the normal range; rather, that they would display more subtle impairment, with low-normal skills in these domains.

2. Within the PRF group, what is the relationship between birthweight or degree of prematurity and language abilities at school age?

In prior studies, a gradient has been reported, with greater prematurity generally corresponding to greater impairment. A similar finding was expected in the present study.

3. Do the group profiles shift over time, across the three home visits?

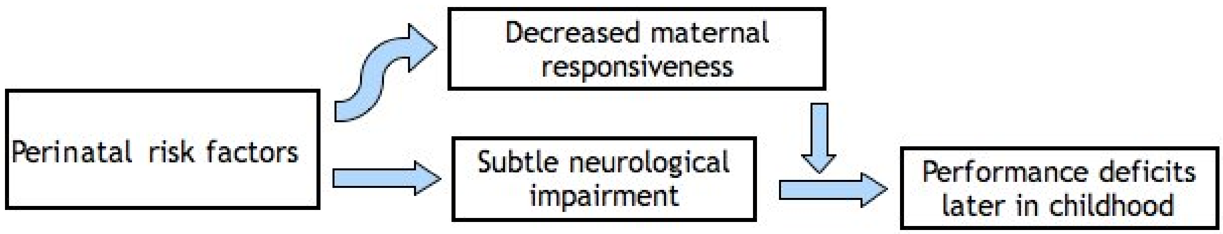
On the one hand, the heightened demands of a narrative language task could exacerbate any group differences noted in conversational language skills. Alternatively, since some studies of

NICU survivors have reported improved performance with advancing age, these group differences might diminish over the period of time between HV2 and HV5.

4. What variables appear to moderate the impact of adverse perinatal events? Specifically, do these data bear out the previously noted trends toward increased resilience in girls, and improved performance among children who were breastfed and those with more educated parents?

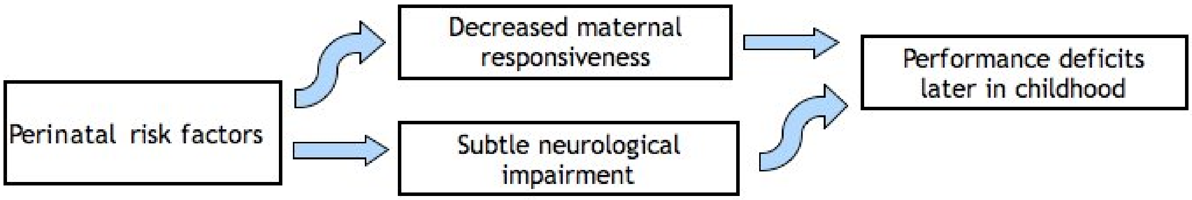
It was expected that these data would support existing trends, showing larger effect sizes among the boys in the sample and small but measurable advantages for breastfed children as well as those with more educated parents.

Figure 1.1 Maternal responsiveness as a moderator.



In this model, decreased maternal responsiveness does not directly cause performance deficits in children but influences the degree to which children are affected by subtle neurological impairment.

Figure 1.2. Maternal responsiveness as a mediator.



In this model, decreased maternal responsiveness has a direct causal role in children's performance.

CHAPTER TWO: METHOD

Participants

Data for this study were drawn from the Western Reserve Reading Project (WRRP; Petrill, Deater-Deckard, Thompson, DeThorne, & Schatschneider, 2006b), a longitudinal study of the genetic and environmental components of children's abilities in reading, mathematics, and related skills. The WRRP includes 350 same-sex twin pairs, chiefly from Ohio, who began participating in the study when the children were in either kindergarten or first grade. Families were recruited through Ohio state birth records, media advertising, clubs for mothers of twins, and school nominations. As part of the general WRRP protocols, both parents and teachers completed questionnaires designed to elicit information on a number of different domains, including prenatal/perinatal medical history, feeding history, demographics, home environment, school performance, and speech-language development.

From the entire WRRP sample, children were selected for the experimental group if they met either of two criteria: very low birthweight (<1500g) or prematurity (born at ≤32 weeks' gestation). These selection criteria yielded a group of 59 children (40 girls and 19 boys) with the perinatal risk factors (PRF) of prematurity and/or very low birthweight. A second group of children born at ≥37 weeks' gestation with no reported perinatal complications were selected from the WRRP database, matched for the following characteristics in order of priority: sex, age (within a 4-month range), highest level of parental education, and race. It was not possible to match all pairs exactly for the two latter criteria; one PRF twin pair whose race was identified as "other" was matched with a white twin pair, and in some cases there was a one-level discrepancy in parental education between matched twin pairs (details appear in Table 3.15). After coding but prior to data analysis, four children were removed from the sample: two female controls, along with their PRF matches. In this case the control children had been prenatally exposed to phenobarbital, a medication which has been linked to cognitive impairment in children exposed gestationally.

The sample was almost entirely made up of twin pairs, but one child who met the weight criterion had a co-twin who was too large at birth for inclusion in this study. For that reason, each group contained 28 twin pairs and half of a 29th twin pair. Of the 114 children, 41 were monozygotic, 69 were dizygotic, and 2 were undetermined. Zygosity profiles differed between the two groups, with 27 MZ children and 30 DZ children in the PRF group, compared to 14 MZ children, 41 DZ children, and 2 children whose zygosity was undetermined in the control group. In the PRF group, 51 children were white, 4 were African-American, and 2 were classified as "other"; in the control group there were 53 white children and 4 African-American children. At the first visit considered in the present study, the mean age was 7.13 years for the PRF children (SD = 0.69) and 7.21 years for the control children (SD = 0.71), with a standard deviation of 0.70 in both groups.

Because the PRF children were selected using an either/or criterion (either ≤32 weeks' completed gestation or <1500g at birth), there were some missing data for gestational age and birthweight. Of the 57 children in the PRF group, 53 had a gestational age at birth reported; these values ranged from 27 to 33 weeks with a mean gestational age of 29.8 weeks (SD = 1.7). In the control group, where the selection criteria were a gestational age ≥37 weeks and an absence of reported perinatal complications, gestational age ranged from 37 to 40 weeks with a mean of 38.3 weeks (SD = 0.99). Birthweights were reported for 53 of the 57 PRF children and 54 of 57 control children. In the PRF group they ranged from 880g to 2213g (1#15 to 4#14), with a mean of 1453g or 3.2 pounds (SD = 331g or .73 pounds), versus a range of 1816g to 3746g (4#0 to 8#4) in the control group (mean = 2703g or 6.0 pounds; SD = 391g or .86 pounds). Hospital stays ranged in length from 10 to 120 days for the PRF group with a mean of 45.3 days (SD = 27.7), and from 0-7 days for the control group with a mean of 3.0 days (SD = 1.5).[2]

On the intake questionnaire parents were given the opportunity to describe a number of events surrounding their children's conception, gestation, and delivery, but for those questions missing data were widespread. Only two sets of parents in the entire WRRP database reported that they had conceived via in vitro fertilization (IVF), for instance. (One of those twin pairs was included in the PRF group.) For questions about NICU complications, it was unclear whether an empty cell in the database meant that a child's NICU course was uncomplicated or that the information was not provided. Because of these inconsistencies in reporting it was not possible to obtain detailed information for the PRF cohort about events during the postnatal period, prompting the decision to emphasize degree of prematurity in the analyses rather than presence or absence of complications.

Parents also responded to questions about their children's history of speech-language concerns, the results of which were described by DeThorne et al. (2006). Parents chose one of

---

[2] At the conclusion of data analysis, new information was integrated from the larger WRRP database which cast doubt on the classification of one twin pair in the PRF group. Their gestational age was entered as 32 weeks, a value which was corroborated with the original questionnaire completed by the parents, but the children's birthweights and length of hospital stay were more consistent with a near-term gestation (i.e., 35-36 weeks). The family could not be contacted to verify these responses prior to the deadline for reporting these data, but every effort will be made to ensure the accuracy of their classification prior to publication of these results. Because of this uncertainty, the birthweight data for the twin pair were treated as missing in regression modeling.

three responses to questions about their children's hearing and speech-language development: no concerns, ongoing concerns, or past concerns, now resolved. They used a similar format to report their children's history of speech-language therapy: no therapy, ongoing therapy, or a history of therapy in the past, now completed. No hearing concerns were reported for children in the control group. In the PRF group, two parents reported that they had been concerned about their children's hearing in the past but that those concerns were now resolved. When parents were asked about concerns regarding speech-language development, 11 PRF parents reported past concerns and 12 reported present concerns, versus 4 control group parents with past concerns and 9 with present concerns. Of the PRF children, 12 had received speech-language therapy in the past and 5 were receiving therapy at the time of the questionnaire; in the control group, 6 children had previously received therapy and 6 others were continuing to receive speech-language services.

General Procedures

Once their questionnaires were returned, families were visited in their homes by a pair of WRRP examiners. Co-twins were simultaneously evaluated in separate rooms, each by a trained research assistant. The evaluators obtained a variety of measures; the emphasis here will be on those included in the present study. At the first home visit (HV1), assessments focused on early reading skills and the children's home environment; no conversational language measures were obtained at this time. Approximately a year later, evaluators returned for a second home visit (HV2) to collect a variety of data including a conversational language sample. Home visit 3 (HV3) followed approximately a year after HV2 and also included conversational language sampling along with other measures of language and reading ability. Home visit 4 (HV4), which was not included in the present study, was a mid-year visit emphasizing mathematics skills. At home visit 5 (HV5), approximately a year after HV3, the measures administered included the Test of Narrative Language (TNL; Gillam & Pearson, 2004), which yields both a narrative language sample for analysis and a standardized score for narrative ability. The TNL is described further in the section of this chapter headed "Norm-Referenced Measures." For a summary of the relevant data collected at each of the home visits, see Table 2.1.

Language Sample Procedures

During HV2 and HV3, 15-minute conversational language samples were collected while a child and an examiner were manipulating modeling clay. Leadholm and Miller's guidelines (1992) for obtaining language samples were used in training, with an emphasis on open-ended questions and child-directed choice of conversational topics. Examiners were instructed to limit their use of requests, directions, and questions requiring only one-word answers. Additionally, they were adjured to exercise patience in order to encourage children to respond, and to leave quiet interludes in their exchanges with less talkative children rather than peppering them with questions. Sample topics of conversation, such as sports teams, favorite movies, and family activities, were provided to them during training along with model questions and responses. Further details on examiner training are available in DeThorne and Hart (2009). All samples were audio-recorded for later transcription.

Trained research assistants transcribed the language samples in Laura DeThorne's laboratories at the Pennsylvania State University and the University of Illinois according to protocols developed for use with Systematic Analysis of Language Transcripts (SALT, Version 8.0; Miller, 2004). Within every twin pair, the two language samples were transcribed by different research assistants in order to avoid inflating estimates of twin similarity. Utterance boundaries were marked using Nippold's 1998 guidelines for communication units (C-units), which require that independent clauses joined by "and," "but," and "or" be segmented into separate utterances. As an example, a child might say "I like to eat pickles but my sister thinks they're disgusting"; a transcriber would divide that into two utterances, "I like to eat pickles" and "But my sister thinks they're disgusting." The use of C-units limits inflation of MLU in a population where concatenation of multiple utterances is common (e.g., "I went to the park and I saw my friend and then she invited me to her house but I said I wasn't sure and so she....").

After transcription, each sample was reviewed by a second trained research assistant to ensure that it conformed to the established guidelines and was free of conspicuous errors in spelling or punctuation. Reliability comparisons were completed on 8% of HV2 transcripts, with a mean agreement of 90% for utterance boundaries and 91% for grammatical morphemes. For HV3, reliability numbers were available for 11% of transcripts; utterance boundary agreement averaged 93% while morpheme agreement averaged 92%. For HV5, reliability results from 11% of the TNL narrative transcripts showed a mean value for utterance boundary agreement of 92% and for morpheme agreement of 95%. TNL reliability values were also available for 11% of HV5 transcripts and averaged 86%.

Measures

The purpose of this study was to assess the morphosyntactic and semantic skills of these school-aged children; to that end, a number of analyses were completed for each transcript. A listing of these analyses, along with the rationale for their use, appears below. While the remaining measures are organized into semantic and morphosyntactic categories, it is recognized that in some cases these designations are somewhat arbitrary; there is a degree of unavoidable overlap in the abilities being assessed. Unless otherwise noted, measures were derived from a child's entire sample of complete utterances, meaning all utterances that are not interrupted or abandoned. In most cases, results were obtained using Computerized Language Analysis software (CLAN; MacWhinney, 2008). In contrast to SALT standard procedures, CLAN frequency counts included partially unintelligible utterances. Given the age of the children in the WRRP cohort and the typically high quality of the recordings obtained by the examiners, the number of partially unintelligible utterances was small and this different definition of the analysis set resulted in only minimal discrepancies.

Semantic Measures

Number of Different Words/Number of Total Words (NDW/NTW)

Both the number of different words (NDW) and the number of total words (NTW) used in each transcript were calculated in SALT in the UIUC Child Language and Literacy Laboratory. These two measures were both derived from the first 100 utterances produced by a child at HV2 and HV3, and from the first 50 utterances at HV5. (Only a few HV5 transcripts reach 100 utterances.) NTW is a straightforward tally of every word used; NDW counts the number of different root words in the analysis set (e.g., “go+ing” and “go+es” would not be counted as different words because they share the same root and are distinguished by their affixes). NDW and NTW both correlate with age in typically developing school-aged children (Leadholm &

Miller, 1992). NDW reflects vocabulary size and linguistic diversity, and is useful in distinguishing between typical and impaired speakers (Watkins, Kelly, Harbers, & Hollis, 1995). NTW reflects volubility and overall verbal proficiency, and is highly correlated with utterance length and NDW (DeThorne, Johnson, & Loeb, 2005). In a comparison of the utility of automated analyses within the HV5 data, NTW was the single best predictor of overall narrative skill (Mahurin Smith & DeThorne, 2008).

Measure D

Measure D is an alternative to the more traditional type-token ratio (TTR; equal to NDW divided by NTW), both of which are intended to be relatively independent of sample length. While D is highly correlated with NDW, NTW, and TTR, it may explain a unique portion of variance in semantic skill for the subset of children who tend to be terse in their use of expressive language but who nonetheless employ a rich vocabulary when they speak. Measure D is derived using the VOCD command within CLAN, which calculates an average of hundreds of TTRs calculated from random subsamples of a transcript. It reflects developmental change and can differentiate typical and impaired groups (Duran, Malvern, Richards, & Chipere, 2004; Owen & Leonard, 2002). In an earlier analysis of the WRRP data, D served as a highly significant predictor of narrative performance (Mahurin Smith & DeThorne, 2008). D values were obtained for all participants in the present study using the procedures outlined in Appendix C.

Word-frequency Analysis

The likelihood that a child will use a given word is related to the frequency with which that word occurs in spoken English. Virtually any typically-developing two-year-old will be able to say "bird"; few will come up with "cockatiel" or "merlin." In structured tasks children show less facility with low-frequency words (Weizman & Snow, 2001; Marinellie & Chan, 2006; Beals, 1997, Beals & Tabors, 1995); appropriate use of uncommon words can add color and sophistication to children's conversations and their narratives. Many children will describe an object as "green"; far fewer will call it "seafoam" or "emerald." The layers of nuance added by diverse vocabulary are an important component of mature language use. For children with subtle deficits, the processing demands imposed by fluent production of connected speech may make it more difficult to retrieve less familiar vocabulary elements. For this reason, the present study assessed children's use of low-frequency vocabulary.

Much of the available information on word frequency comes from large existing corpora, such as the London-Lund corpus (Brown, 1984) or the Kucera-Francis corpus (Kucera & Francis, 1967). Such corpora present three obstacles to use with the WRRP dataset: first, they were developed from the language of adults, including speakers of British English; second, they frequently draw on written language samples; third, they may be decades old and will necessarily misrepresent words heard frequently in the speech of children today, such as "cellphone," "internet," and even "soccer." Consequently, they can give a skewed result when used to evaluate children's spoken language. As an example, "alien" is a low-frequency word in the Kucera-Francis corpus and does not even appear in Brown's study of the London-Lund corpus, but nearly all of the HV5 narratives include the word "alien" because aliens appear in one of the TNL picture stimuli.

In lieu of these corpora, then, the larger WRRP database was used to generate a concordance of the 1.2 million words used in the the 1437 unique WRRP transcripts available in December 2008. (Reliability transcripts were eliminated from consideration.) This inclusive concordance, in which word frequency ranged from 50,329 instances (for "and") all the way down to 1 instance (for "ziggurats"), was pared down to a low-frequency concordance, containing only the words used 15 or fewer times across the entire WRRP sample. The low-frequency concordance was edited to remove typographical errors, movie names and other proper nouns, numbers, and sound effects (e.g., zzzz). In addition, kinship terms were removed from consideration regardless of frequency of use; if a child called his grandparents "my oma and my opa," for instance, those words were excluded despite their rarity in the corpus. Morphemic variations were collapsed (i.e., "borrowing" and "borrowed" were counted together with "borrow").

At this point the low-frequency list included many words which could serve as a useful index of expressive vocabulary development, but it also included a number of early-acquired words which seldom surfaced in the conversations between children and examiners (e.g., beans, soap, helicopter). A number of solutions to this problem were considered and discarded. When Beals and Tabors (1995) assessed maternal conversational language for low-frequency vocabulary, they removed items on the Dale-Chall list of words recognized by most 4th-grade readers (Dale & Chall, 1948). There are, however, significant discrepancies between lists of words that fourth-graders are able to read and lists of words that first-graders are likely to say. Some of the Dale-Chall words seldom occur in children's conversational speech but may bring added color and precision when they are used: words such as "savage," "steamboat," and "postage." In the end, the low-frequency word list derived from the WRRP data was cross-checked against the MacArthur-Bates database of words produced by toddlers, a list which includes the words produced by at least 15% of 2.5-year-olds in their normative sample. These words, 61 of which appeared on the edited WRRP low-frequency list, were eliminated from consideration. This approach to tallying low-frequency vocabulary might be called the Rule of 15s: low-frequency words are those that occurred no more than 15 times in the large corpus, and were acquired by <15% of young children in the MacArthur-Bates sample.

Once the concordance had been edited, the FREQ command was used to tally occurrences of these low-frequency words (types, rather than tokens) in the 301 transcripts under consideration. For each transcript, a listing was generated of the low-frequency words as they appeared in context, and this listing was reviewed carefully. Some words appeared solely because of typographical errors, or exploratory and semantically inappropriate uses (e.g., "My mother fileted my leg"). Such instances were removed from a child's tally of low-frequency vocabulary.

Because of the frequently subjective nature of the decisions involved, reliability comparisons were undertaken. A research assistant from the UIUC Child Language and Literacy Laboratory reviewed the tallies for 10% of the transcripts and recorded her judgments about retaining or rejecting the words in each tally based on their semantic appropriateness. Point-to-point reliability was above 90%.

Finally, a density measure for low-frequency vocabulary was calculated for each transcript. The number of low-frequency word types in each sample was divided by the number of utterances in that sample in order to reduce confounding with sample length.

Morphologically Complex Word Analysis in CLAN (MOR)

Vocabulary acquisition during the school years is driven in large part by children's burgeoning understanding of the role of derivational morphemes. Anglin (1993) studied this phenomenon cross-sectionally and found significant growth in children's ability to decode these morphemes between first and third grades, with an even more pronounced increase between third and fifth grades. Production of morphologically complex words was assessed by means of the CLAN programs called MOR and POST. MOR assigns possible parts of speech to each word of a transcript using its sizable lexicon of English words. When a word contains one of the target affixes, listed in Appendix A, it is flagged by MOR. "Unfriendly," becomes un#adj:n|friend-LY, allowing both the un- prefix and the -ly suffix to be tallied. When a word can serve more than one grammatical function, MOR lists all the available possibilities. The second program, POST, attempts to disambiguate each word in the transcript using information from neighboring words in the sentence. The word “her,” for instance, might be either the objective case pronoun (e.g., “He asked *her*”) or the possessive pronoun (e.g., “He asked *her* mother”). Because the possessive pronoun “her” must always be followed by a noun, POST would recognize that the word could not serve that function in a sentence where it occurred in final position.

As one might imagine, the many ambiguities of English coupled with the vagaries of young children's conversational efforts mean that any automated disambiguator will meet with limited success. Consequently, each line of each transcript was reviewed to correct the inevitable errors. Each word containing any of the affixes under consideration was tallied by means of CLAN's FREQ program.

Literate Language Elements: Adverbs and Metacognitive Verbs

Greenhalgh & Strong (2001) describe two sensitive markers for children's literate language ability: their use of adverbs and of metalinguistic/metacognitive verbs (e.g., “decide,” “wish”). They contend that when children use adverbs to elaborate on events, it indicates a growing understanding of nuance and subtlety. For narrative samples in particular, a single adverb can add layers of colorful detail. As an example, when the TNL aliens shout, "Greetings, earthlings," do they do so menacingly or enthusiastically? Similarly, when children employ metalinguistic/metacognitive verbs, it provides a demonstration of their ability to take another's perspective. CLAN's FREQ program was used to tally adverbs (previously identified in all transcripts via CLAN's MOR and POST programs) as well as metalinguistic/metacognitive verbs. The verbs of interest are listed in Appendix B. Both tallies were converted to densities to limit confounding with sample length.

Syntactic Measures

Mean Length of Utterance in C Units (MLU-C)

Mean length of utterance in C-units (MLU-C) was calculated in SALT in Laura DeThorne’s labs at PSU and UIUC for all language samples. MLU-C is obtained by dividing the total number of morphemes by the total number of utterances, using all of a child's complete and intelligible utterances. MLU-C shows developmental change across the school years (Leadholm & Miller, 1992; Rice, 2004; Rice, Redmond, & Hoffman, 2006) and serves to differentiate among children of varying language abilities (Klee, Schaffer, May, Membrino, & Mougey, 1989).

Conjunction Analyses

Previous work on the WRRP language data has included the Total Number of Conjunctions (TNC), a frequency count of twelve commonly used conjunctions. The present study did not examine TNC, focusing instead on two alternate analyses of conjunction use. First, the list of conjunctions to be counted was expanded using a comprehensive dictionary; occurrences of conjunctions were tallied using CLAN's FREQ command. Since any straightforward conjunction count is tightly bound up with sample length, a conjunction density measure was derived by calculating the ratio of total conjunctions to total number of utterances. Conjunction density can provide a useful indicator of language prowess and has been shown to be a significant predictor of listener judgment of narratives in the WRRP cohort (Mahurin Smith & DeThorne, 2008).

The second conjunction analysis was driven by the observation that one feature of literate language among school-aged children is reliance on conjunctions other than "and" or "then" (Greenhalgh & Strong, 2001). While use of conjunctions such as "however" is limited in school-aged children, it does occur and represents a potentially sensitive indicator of facility with literate language. Thus, instances of conjunctions other than "and" and "then" were tallied separately for all transcripts. This tally incorporated results from the MOR analysis, detailed below, to ensure that words with multiple potential functions were not counted as conjunctions in instances where they served as adverbs or prepositions (e.g., "Sally hadn't seen the aliens *yet*" would be excluded, while "Sally was sad the aliens had blown up her house, *yet* she resolved to carry on" would be counted).

Elaborated Noun Phrases (ENPs)

As children develop skill with language, their use of complex phrases grows. Their descriptions expand from "dog" to "black dog" to "the loud slobbery black dog that dug up my grandma's tomato plants." Children's use of elaborated noun phrases (ENPs) was assessed in the present study because it is a marker for literate language use (Nippold, 2007; Greenhalgh & Strong, 2001; Eisenberg, Ukrainetz, Hsu, Kaderavek, Justice, & Gillam, 2008): increasing density of elaborated noun phrases indicates increasingly sophisticated language use. ENPs were coded using guidelines adapted from Greenhalgh and Strong's 2001 paper, spelled out in Appendix D.

Developmental Sentence Scoring (DSS)

Developmental Sentence Scoring (DSS) was devised by Lee (1974) as a measure of syntactic complexity. Sentences are scored on the sophistication of eight different elements, with a final additional point awarded or withheld based on their overall grammatical appropriateness. DSS can distinguish between language-impaired children and typically-developing children, and has been used to illustrate the effects of language therapy on syntactic development (Rice, Redmond, & Hoffman, 2006; Hughes, Fey, & Long, 1992). DSS results were reported for HV2 in DeThorne, Petrill, Hart, Channell, Campbell, Deater-Deckard, et al. (2008); they were derived by the fourth author using his own gcSALT program (Channell, 2006) with subsequent manual corrections. He used the same program to obtain DSS results for the HV3 transcripts in the present study. For HV5, the author of the present study used Computerized Profiling software (Long, 2008) to produce an automated analysis, which was subsequently hand-corrected. In addition to the differences in software, slight differences in scoring protocols should be mentioned: the HV5 DSS results hewed closely to Lee's 1974 rules, while the HV2 and HV3 results incorporated the strategy for scoring the word “like” that was proposed in Hughes, Fey, and Long (1992). This difference could result in slightly higher values for the HV5 samples, so comparisons should be undertaken with caution.

Norm-Referenced Measures

Test of Narrative Language (TNL)

Gillam and Pearson (2004) devised the Test of Narrative Language (TNL), a norm-referenced measure of narrative ability in school-aged children. It consists of two subtests, one which measures comprehension of stories presented by the examiner, and one which assesses the characteristics of three stories the children produce: first, a re-tell task in which they are instructed to reproduce the story they just heard from the examiner; second, a task in which they tell a story based on a sequence of five pictures; and third, a story prompted by a single picture of picnicking aliens. Scoring is based on criteria both objective (e.g., number of causal conjunctions used in a story) and subjective (e.g., "Does the story make sense?"). The expressive subtest is scored in the UIUC Child Language and Literacy Laboratory after the sample is transcribed; only those results are reported here.

Clinical Evaluation of Language Fundamentals -4 (CELF-4)

Selected subtests of the Clinical Evaluation of Language Fundamentals - 4 (CELF-4; Semel, Wiig, & Secord, 2003), a norm-referenced assessment of global language ability, were administered at HV5. The four subtests given to WRRP participants included Recalling Sentences, Word Classes (expressive), Word Classes (receptive), and Understanding Paragraphs.

Cognitive Measures

Outcomes for two cognitive measures were recorded for both groups. Since previous studies have suggested that subtle language deficits observed in VLBW survivors may be associated with cognitive skills in the low-normal range, it is important to document the cognitive function of the children in the study (Singer, Siegel, Lewis, Hawkins, Yamashita, & Baley, 2001; Lewis et al., 2002; Short et al., 2003). The short form of the Stanford-Binet Intelligence Scale (Thorndike, Hagen, & Sattler, 1986) was administered to all children in the WRRP sample at HV2 and HV3. They also completed a digit span task, which prior research suggests may tap relevant skills such as short-term auditory memory (Bristol, Gathercole, & Marlow, 1998). Both IQ and digit span results were reviewed for this study.

Summary

The present study compared results for a number of measures designed to assess morphosyntactic and semantic abilities along with global language and cognitive skill. They are listed in Table 2.2.

Analyses

Factor Analysis

The large number of dependent variables assessed in this study resulted an unwieldy dataset containing thousands of individual data points. Regression modeling considering each of the variables in isolation would raise the familywise α to unacceptably high levels. Factor analysis was used to condense the available information into three factors; further details are presented in chapter 3.

Hierarchical Linear Modeling (HLM)

Hierarchical linear modeling (HLM) was used to assess group differences, with individual factors serving as dependent variables. HLM decreases the likelihood of a Type I error in a twin sample by allowing a researcher to control for nesting (Gelman & Hill, 2006; Bryk & Raudenbush, 1992). Individual children (considered to be level 1) were grouped into families (considered to be level 2), and the intercepts in regression models were allowed to vary by family. This multilevel approach was employed to consider all four research questions. All analyses were completed in the statistical software program R, version 2.9 (R Development Core Team, 2009).

Table 2.1

Summary of relevant measures completed at WRRP home visits

Visit	Children's grade level	Relevant data collected
Home visit 1 (HV1)	K or 1	Early reading measures, home environment assessment, parent questionnaire
Home visit 2 (HV2)	**1 or 2**	**Conversational language sample, IQ, digit span**
Home visit 3 (HV3)	**2 or 3**	**Conversational language sample, IQ, digit span**
Home visit 4 (HV4)	variable (mid-year visit)	None (assessed mathematical skills)
Home visit 5 (HV5)	**3 or 4**	**Test of Narrative Language, CELF-4 subtests**

Note: The three home visits under consideration in this study are shown in boldface type.

Table 2.2

Summary of measures in the present study

Skill to be assessed	Measure
Semantic ability	NDW, NTW, Measure D, low-frequency word density, morphologically complex word density, metalinguistic/metecognitive verb density.
Syntactic ability	MLU, total conjunction density, density of low-frequency conjunctions, DSS, elaborated noun phrase density
Global language/cognition	Stanford-Binet short form, digit span, TNL standard score, CELF-4 standard score

CHAPTER THREE: RESULTS

This chapter will first review the descriptive results obtained for all variables, followed by an examination of the relationships observed among those variables via correlation analysis and factor analysis. The research questions presented in Chapter One are considered in the subsequent sections of this chapter with a review of the evidence regarding (1) group differences across home visits, (2) the correlation between degree of prematurity and language skills at school age, (3) changes over time in the relationship between groups, and (4) the impact of potential moderating variables.

Descriptive statistics were obtained for all measures and are presented in the tables that follow this chapter. First, utterance counts at each visit were compared for the PRF and control samples as a measure of general volubility; those results appear in Table 3.1. At HV2 and HV3, when the children participated in a 15-minute conversational sample, the utterance counts were normally distributed. Utterance counts ranged from 10 to 287 at HV2, and from 42 to 241 at HV3. At HV5 the children were instructed to complete a structured narrative task with no time limit. The resulting distribution of utterance counts had a long right tail, illustrating the tendency of a few children to tell lengthy stories; values ranged from 19-120. For HV2 only, examiner utterance counts were also available. These values were normally distributed and ranged from 38 to 226.

Descriptive statistics for the other HV2 measures are presented in Table 3.2 as density measures. As discussed in chapter 2, all count results were converted to densities to limit confounding with sample length. Because these density values were almost always less than 1, they have been multiplied by 100 in the tables that follow in order to facilitate comparisons.

Using adverb density as an example, the mean value of 23.36 in the PRF group corresponds to an actual density value of 0.2336. In other words, the PRF group used an adverb an average of once every 4 or 5 utterances.

Table 3.2 also includes the same descriptive statistics presented as z scores to provide an estimate of effect size. Readers will observe that for all variables assessed at HV2, the PRF group scored below the overall mean while the full-term group scored above it, although in some cases the difference is very small; these group differences ranged from .06 SD for DSS to .55 SD for IQ. These discrepancies do not reflect a clinically significant difference in the PRF group, since group means fall within the normal range in all cases where a normal range has been defined. One element of the z score comparison which may initially be puzzling is its symmetry: if the PRF group scored .04 SD below the mean, the full-term group scored .08 SD higher, or exactly .04 SD above the mean. This symmetry is to be expected because the groups are the same size: if one group mean falls a quarter of a standard deviation below the grand mean, the other group mean must fall a quarter of a standard deviation above the grand mean. For most of the variables, the standard deviations are smaller for the PRF group than for the group as whole, suggesting that their performance falls within a narrower range. The largest discrepancies between the two groups at HV2 were for IQ and digit span, closely followed by density of low-frequency word use and by NDW.

Because of the anticipated importance of sex differences, the HV2 z scores were further broken down by sex and group (see Table 3.3). For a number of the measures, the lowest mean z scores were observed among the PRF girls while the highest were seen in the FT boys. In some cases the difference between these group reached one standard deviation. The direction of the group difference is somewhat surprising, since the literature indicates a higher degree of resilience among girls born prematurely.

Descriptive statistics for HV3 appear in Table 3.4. As for HV2, the PRF group performed below the grand mean for all measures except for a small positive difference on D. The spread between groups ranged from .01 SD to .80 SD. Here, too, the standard deviations tended to be smaller for the PRF group, suggesting a narrower range of abilities. These values are less symmetrical than the HV2 values because attrition between HV2 and HV3 affected the two groups asymmetrically. At HV3 there were 50 children in the control group and 42 in the experimental group, a discrepancy that corrected itself between HV3 and HV5 as shown below. The largest difference between group means was observed for IQ, followed by low-frequency word density and digit span. The HV3 results are separated by sex and group in Table 3.5. The PRF girls scored below the grand mean on all but one of the measures; for IQ and low-frequency word density, their results fell .36 SD below the grand mean. The strongest performance was observed in the FT boys, whose scores ranged from .18 to .90 SD above the grand mean.

At HV5, there were complete data on 43 children in the experimental group and 42 in the control group, and for the most part the descriptive statistics presented in Table 3.6 as z scores reflect this restored symmetry in group sizes. Where asymmetry is observed for z scores in Table 3.6, it indicates asymmetry in the numbers of missing values for a variable. For instance, the asymmetry evident for NDW-50 and NTW-50 reflects the fact that fewer of the PRF children than the FT children reached the 50-utterance mark. As with HV2 and HV5, there was a trend toward lower performance in the PRF group. On all measures except for metalinguistic verb density and TNL standard score, where the group means were respectively .01 and .13 SD above the grand mean, the PRF scores fell below the grand mean. Group differences ranged from .02 SD to .64 SD. HV5 data did not include IQ or digit span, but scores are available for four subtests of the Clinical Evaluation of Language Fundamentals-4 (CELF-4; Semel, Wiig, & Secord, 2003) and for the Test of Narrative Language (TNL; Gillam & Pearson, 2004). The three largest differences between groups, ranging from .45 to .64 SD, were observed on three of these four CELF-4 subtests. Sex contrasts for HV5 are presented in Table 3.7. Patterns similar to those seen at HV2 and HV3 are evident: the PRF girls scored below the grand mean on all measures, with group means ranging from -.01 SD to -.29 SD; the full-term boys scored above the grand mean on all measures except elaborated noun phrase density (-.01 SD). For two measures, MLU and morphologically complex word density, the full-term boys' scores were more than half a standard deviation above the grand mean.

Correlation Matrices

To facilitate understanding of the interrelationships among variables, correlation matrices were obtained for all variables at each of the three home visits; they are presented in Tables 3.8, 3.9, and 3.10. All values are Pearson correlation coefficients, and missing values were dealt with via pairwise deletion. Some brief observations about these matrices suggest themselves. At HV2, 61 of 91 possible correlations (67.0%) among the measures are statistically significant and 50 of 91 are highly significant (54.9%); many of these values range from moderate to large in magnitude. The values of the statistically significant correlations at HV2 range from 0.18 to 0.88. A similar trend is evident again at HV3, where 62 of 91 possible correlations are significant (68.2%) and 54 of 91 are highly significant (59.3%). Coefficients of even greater magnitude occur; the range of significant correlations runs from 0.21 to 0.97. For the narrative task completed at HV5, however, the language measures correlate less often and the magnitude of the correlation coefficients is generally smaller. Of the 136 possible correlations, 76, or 55.9%, are statistically significant and 46 are highly significant (33.8%). Significant correlations range from 0.21 to 0.80, with most correlations falling toward the lower end of that range. For an illustration of this shift, the reader might compare the first column of Table 3.9, where 10 of the 13 correlations are highly significant and 9 of the 13 are >.5, and the first column of Table 3.10, where there are no highly significant coefficients and the largest *r* value is <.25. This trend might reflect the change in the nature of the task at HV5 (structured narratives as opposed to the open-ended conversations of HV2 and HV3), or it might be an artifact of the decreased sample length seen at HV5. While density measurements are more impervious than raw count data to changes related to sample length, shorter samples offer a narrower window in which individual differences can emerge.

Factor Analysis

As discussed in chapter 2, the analyses conducted in this study yielded a sizable array of information, and it was necessary to condense the available data into a smaller number of dependent variables. Confirmatory factor analysis, a statistical approach that uses covariance matrices to assess underlying relationships among variables, was undertaken using the sem() package in R. This step of the data analysis process was intended to reduce the number of dependent variables based on their relationships to each other, minimizing the impact of collinearity among variables and reducing the risk of Type I error. Previous work on the WRRP data (DeThorne et al., 2008) indicated that standardized tests tended to load on one factor, labeled the “formal factor,” while conversational language measures tended to load on another.

In addition to this distinction, conversational language measures in the present study were further divided into semantic measures and syntactic measures to yield a total of three factors, as discussed in Chapter 2 and illustrated in Table 3.11. Much like correlation coefficients, factor loadings usually range between -1 and 1, with stronger loadings evident in more extreme values. Since factor loadings are derived from covariance matrices, it is unsurprising that trends in the factor loadings mirror trends observed in the correlation matrices (see Table 3.11). At HV3, for instance, the higher correlations among many variables were reflected in higher factor loadings. At HV5, when the semantic variables correlated poorly with each other, the factor loadings for variables other than NDW and NTW were low, ranging from .10 to .25. The previously mentioned example of shifting correlations for adverb density provides a useful illustration here as well: at HV3, the factor loading for adverb density was .82, but it fell to .20 at HV5. Unexpectedly, measure D's factor loadings were quite low at all three home visits and dipped below zero at HV3, reflecting its negative correlations with other measures of semantic skill.

For the conversational language variables, factor loadings were then used to create weighted sums for use as dependent variables in regression models (see DiStefano, Zhu, & Mîndrilă, 2009). The loading for a given variable was multiplied by a participant's *z* score for that variable to yield a weighted component score. Variables with loadings <.20 were excluded, resulting in the omission of measure D at all three home visits and of metalinguistic verb density at HV2. For each of the three factors, which are conceptualized as a semantic factor, a syntactic factor, and a formal factor, component scores were summed across variables In this way the 12-14 dependent variables measured for each home visit were condensed into three, reducing the risk of Type 1 error in subsequent analyses and likely increasing the stability of measurement. The resulting sums were re-scaled as z scores to aid in interpretation of regression results. This same strategy was employed to derive formal factor scores at HV5, when data were available for both CELF subtests and the Test of Narrative Language. At HV2 and HV3, when IQ and digit span were the only two measures loading on the formal factor, the mean of those two z scores was used as the factor score. All factor scores were approximately normally distributed, with a slight positive skew for the semantic factor. This skewness ranged from 0.47-0.69 while kurtosis values ranged from -0.57 to 0.60; no transformation of the data was deemed necessary in preparation for regression modeling.

Regression Modeling

Regression modeling was employed to address the four research questions raised in Chapter One and reiterated below. Because the nested structure of the data violates the assumption of independence, the lmer() package in R was used to create multilevel models, with family ID as the grouping variable. Intercepts were allowed to vary by family as illustrated in Figure 3.1; slopes remained fixed. Multilevel modeling complicates the calculation of R^2; in the models that follow, this difficulty was addressed by splitting the sample in half and calculating R^2 separately for the group of firstborn twins and the group of secondborn twins, using ordinary least squares regression. Both R^2 values are reported for any model that yielded a result with a p value < .10.

Question 1: Group Differences

The first research question under consideration asked how the semantic and morphosyntactic skills of children in the PRF group compared with those of their peers in the control group, as assessed at each of three annual home visits. The initial regression models were very simple: does PRF status predict performance on Factor 1 (semantic), Factor 2 (syntactic), or Factor 3 (formal) at HV2, HV3, or HV5? The details of the regression models are shown in Table 3.12 to augment the narrative descriptions presented here.

At HV2, a trend toward lower performance across all measures was observed in the PRF group, but the differences only reached statistical significance for Factor 3, the formal factor. Prematurity was associated with a decrement of just over half of a standard deviation (-0.54) on IQ/digit span. For the semantic factor, a decrement of slightly less than a third of a standard deviation was noted (-0.31); for the syntactic factor, it was a quarter of a standard devation (-0.26). For both semantic and syntactic factors, however, the standard error was large enough that these trends were not statistically significant. *A priori* power calculations indicated that this study was adequately powered to detect a difference of half of a standard deviation, but a larger sample would be required to show statistically significant differences for a smaller effect size. For a difference of a third of a standard deviation, groups of 144 would be required for 80% power in a two-tailed test; for a difference of a quarter of a standard deviation, the necessary group size rises to 251 (see Gelman & Hill, 2006).

At HV3, a very similar pattern was observed. The estimated beta value for the formal factor was -0.72, representing a decrement of approximately three-quarters of a standard deviation for the PRF children. This result was highly significant. For the semantic and syntactic factors, decrements of -0.35 and -0.18, respectively, were observed. Both coefficients, however, had sizable standard errors and were not statistically significant. As noted in the previous paragraph, a larger sample size would be necessary for reliable assessment of these trends toward modest differences.

None of the results for HV5 reached statistical significance, though similar trends were evident. At this visit, the beta value for the formal factor was estimated as -0.43. The reader is reminded that HV5 measures did not include either IQ or digit span; these formal factor results are based instead on 4 subtests of the CELF-4 and the Test of Narrative Language. The semantic and syntactic factors indicated a decrement for the PRF group of a fifth and a tenth of a standard deviation, respectively.

These results raise additional questions about the possible mechanism driving the observed differences and about potential strategies for predicting long-term outcomes. Specifically, is it possible to predict outcomes more accurately given more information about a child's perinatal history? If so, which information appears to be most useful? These questions will be addressed in the following section.

Question 2: Degree of Prematurity

The second research question in the present study assessed the predictive power of degree of prematurity, focusing ultimately on whether outcomes within the PRF group could be predicted based on indicators of perinatal risk. In the literature on long-term effects of prematurity, a gradient has typically been observed among children born prematurely: babies born at 24 weeks tend to fare worse than babies born at 28 weeks, who are more likely to struggle at school age than babies born at 32 weeks, and so on (Bhutta, Cleves, Casey, Cradock, & Anand, 2002). A series of regression models tested the explanatory power of four continuous measures of a child's degree of prematurity: gestational age, birthweight, length of hospital stay, and an index variable whose construction is explained below. This section will first describe the models that considered the full sample, both PRF and control groups, since trends in regression modeling are often clearer with a larger *n* and a wider range of predictor values. Results for models confined to the

PRF group will follow.

Continuous Predictor Models: Full Sample

Models built using these continuous variables yielded similar results to those obtained for Question 1 using the binary classification of premature vs. full-term, although birthweight performed poorly as a predictor. Coefficients for these models appear in Table 3.13. In all cases, the results paralleled those of the binary models described in the preceding pages and detailed in Table 3.12; differences in coefficients can largely be attributed to the variables' different scales. For gestational age, for example, each advancing week of gestation was associated with an improvement of .04 SD in the semantic domain. Using this model, one would expect a baby born at 38 weeks to score .40SD higher, on average, on the semantic factor than a baby born at 28 weeks (10 weeks' difference in gestation, multiplied by .04 for each week). This finding was marginally significant ($p < .10$); gestational age was a statistically significant predictor for the formal factor at all 3 home visits. For birthweight, the beta value indicates that each one-pound increase in birthweight predicted a small increase in a child's score; this model resulted in large standard errors, however, and only one of the nine results, the HV3 formal factor, reached statistical significance. For hospital stay, the beta was negative and much smaller, indicating that each day's increase in hospital stay was associated with a very small decrement in ability as measured in this study. Predictive power for hospital stay fell in between that of birthweight and that of gestational age, as illustrated in Table 3.13.

Gestational age and length of hospital stay were combined into an index variable intended to represent degree of prematurity, a continuous variable designed to reflect more specific information about each participant's perinatal course. This variable was constructed by obtaining z scores for both variables and deriving a mean for each child using the scaled values for gestational age and the opposite of the scaled value for hospital stay. (In contrast to birthweight and gestational age, where larger is better for premies, a briefer hospital stay suggests a less complicated neonatal course.) Results for the resulting regression models are also presented in Table 3.13. The meaning of this coefficient is less intuitively evident than that of the other continuous predictors: each unit represents an increase of one standard deviation in the mean of gestational age and a decrease of one standard deviation in length of hospital stay. Despite incorporating more information about each individual, this model did not improve on the predictive power afforded by gestational age.

To summarize, all of the approaches to predicting outcomes using continuous variables in the full sample yielded results similar to those obtained using binary predictors, with clear between-groups differences for the formal factor at HV2 and HV3, and more modest between-groups differences for the semantic factor at HV2 and the formal factor at HV5. While a trend toward poorer performance among the PRF children was clear, the results of the language sample measures were noisy enough that the differences never reached statistical significance. R^2 values never exceeded 0.16, indicating a small effect size.

Of the explanatory variables tested, gestational age proved the most useful at clarifying trends and reducing noise in the data. For the remainder of this chapter, then, models will use gestational age as an explanatory variable. For the four children who were missing responses for gestational age, values were imputed based on their birthweight and length of hospital stay.

Continuous Predictor Models: PRF Group

A second series of models was built to look more specifically at the impact of degree of prematurity within the PRF group. These models were designed to address the question of whether, among the children born prematurely, a trend toward poorer performance on the outcomes measures was associated with shorter gestation. Gestational age was used to predict factor scores for the PRF group; results are detailed in Table 3.14. The resulting models generally yielded the expected results, in that the intercepts were usually negative (indicating a disadvantage with diminishing gestational age) while the slopes were usually positive (indicating growth across the period of gestation), but the standard errors were quite large and no significant findings emerged. Possible reasons for these findings are discussed in Chapter 4.

Question 3: Changes Across Time in Relationships Between Groups

The third research question to be considered was whether the relationship between the two groups shifted from HV2 to HV5. In other words, did the differences between the two groups at HV2 stay constant, diminish, or increase at HV3 and HV5? When the study was originally designed, this question was to be addressed quantitatively, using a statistical approach such as a mixed-design MANOVA to model the within-group as well as the between-group effects. Inspection of the descriptive results, however, makes plain that this strategy is difficult to defend in view of the discrepancies between the first two home visits, where conversational language samples were collected, and the third home visit, where a highly structured narrative task was completed. It was deemed inappropriate to tackle the question using repeated-measures methodology, because the same measure was not repeated across all three home visits and it is impossible to be certain that maturation alone is driving the changes. Since the assumptions of repeated-measures testing were not met, a descriptive summary will be presented here in lieu of a dubious F statistic.

In looking at the question of change over time, one consideration was the relative attrition rates of higher and lower performers. One could imagine that families where children found the assessment tasks more taxing might be less apt to continue year after year. To examine this possibility, a global performance score was obtained by averaging z scores across all HV2 measures, and comparing attrition rates for roughly the highest and lowest quintiles in the PRF group. Of the 12 lowest-performing children within the PRF group, 4 (33%) had dropped out by HV5 while 8 remained in the study. Of the 12 highest-performing PRF children, 5 (42%) dropped out by HV5 while 7 continued to participate. These very comparable rates suggest that it is reasonable to view changes in between-groups relationships as at least partially a function of maturation rather than as an artifact of asymmetrical attrition.

At all three home visits, scores on the formal factor illustrated most clearly the differences between the children born prematurely and their full-term counterparts. This difference increased from half to three-quarters of a standard deviation between HV2 and HV3. At HV5, when different standardized tests were substituted, the difference shrank to two-fifths of a standard deviation.

Figures 3.3 and 3.4 illustrate the trajectory of means for the semantic and syntactic variables. Measure D is not shown because it correlated so poorly with the other variables that it was not included in the semantic factor scores. On all semantic and syntactic measures shown, the control children outperformed the PRF children, though these differences were largely narrower by HV5. Growth was evident for most of the measures, although downturns at HV5 were noted for NDW, NTW, and low-frequency vocabulary density. In general there was a trend for the gap between the PRF group and the control group to be narrowest at HV5; the two exceptions to this trend toward narrower between-groups divergence were noted for density of morphologically complex words (shown in the right third of Figure 3.3) and for DSS (see the right third of Figure 3.4).

Interpretation of patterns across the language factors is complicated by the large standard errors observed for these variables throughout the process of regression modeling, but the explanatory power of prematurity appears to diminish across the three home visits. The trends described in the remainder of this paragraph are drawn from beta values presented in the top third of Table 3.13, in the section headed "Gestational Age." For the semantic factor, the coefficient associated with each week of advancing gestational age shrank from .04 (HV2) to .03 (HV3) to .02 (HV5), and its initial marginal significance evaporated. For the syntactic factor, the same trend was evident: a beta of .04 (HV2) dropped to .01 at HV3 and remained there for HV5. While the control children outscored the PRF children on every syntactic measure, at no point did the difference reach statistical significance.

Question 4. Moderating Variables

The final research question under consideration was the impact of selected moderating variables. Specifically, are there protective effects associated with increasing parental education, increased breastfeeding duration, or female sex? These variables will be addressed individually first to provide background information on their distribution across the sample and their explanatory power in simple multilevel models, summarized in Table 3.16. Review of the models combining perinatal history and these explanatory variables will follow; these results appear in Tables 3.17 and 3.18.

Parental Education

In the WRRP sample as a whole, parents with higher education are overrepresented compared to US census data. A 2004 census publication, for instance, indicates that approximately 57% of adults between the ages of 20 and 50 have completed at least some college (United States Census Bureau, 2004). Within the WRRP sample, however, 87% of parents have completed at least some college; only 8% reported having a high school diploma or less education. (The remaining 5% of WRRP parents described their educational level as "other.") Given this trend in the WRRP sample as a whole, it is noteworthy that a different educational profile is evident in the subsample assessed in the current study. In contrast to the 8% of parents in the full sample with only a high school diploma, 21% of parents in the PRF group reported that their highest level of education was a high school diploma. In addition to this discrepancy in parent education between the PRF subsample and the wider WRRP sample, differences in SES emerged in the current study's subsamples based on sex. To wit, 26% of the PRF girls' parents fell into the "high school diploma" category, versus 11% of the PRF boys' parents. Further details are presented in Table 3.15, which breaks down parental educational achievement by group and sex. Table 3.15 also includes an expected value for each category, based on the breakdown in the WRRP sample as a whole. This difference in group composition was assessed via chi-square testing with highly significant results, χ^2 (5, n = 57) = 20.1, p = .001), indicating that the parents of PRF children diverge markedly in their educational background from the patterns seen in the WRRP sample as a whole.

The results of regression models using parental education as the sole explanatory variable are shown in Table 3.16. Nine models appear in the table, one for each of the three factors at each of the three home visits. The effects of parental education were variable, with larger betas observed, along with a trend toward statistical significance, as the children grew older. In the models with $p < .10$, split-half R^2 results ranged did not exceed 0.17, indicating a small and inconsistent effect.

Breastfeeding

At intake, all parents in the WRRP sample reported the duration of breastfeeding and/or formula feeding for each twin. Within the full sample, the breastfeeding initiation rate was 71% with a median duration of 1.4 months and a mean duration of 3.6 months. In the subsample under consideration, the breastfeeding initiation rate was 73.7%, with very similar rates for the PRF children (75.5%) and the control children (72.1%). The median breastfeeding duration was 2 months for the PRF group and 2.6 months for the controls. The mean for both groups was 4 months, although this figure was inflated by the presence of two pairs of twins who nursed until they were 36 months old. When those values were removed from consideration, the mean age of weaning dropped to 2.9 months for both groups. To correct the long tail observed in the distribution of breastfeeding duration (skewness = 3.36), these values were log-transformed. The large majority of parents in the sample (108 out of 114) reported using some infant formula; no information on the ratio of formula-feeding to breastfeeding was available.

In the US at this time, breastfeeding rates vary with parental education (Taveras, Capra, Braveman, Jensvold, Escobar, & Lieu, 2003); consequently, research on the impact of breastfeeding must always consider the possibility of confounding with parental education. In the present study, however, the correlation between breastfeeding duration and parental education was a very modest and non-significant 0.12, indicating that both variables could be used in multiple regression equations for these data without concerns about collinearity.

Results for models fitted using breastfeeding as the single explanatory variable appear in Table 3.16. Of the nine betas, seven were weakly positive, suggesting a very small positive effect of breastfeeding. These coefficients are more difficult to interpret than those of previous models because of the natural log transformation. As an example, consider the results for the semantic factor at HV2: a unit increase in breastfeeding duration, which corresponds to 2.7 months of breastfeeding because of the natural log scale, would be associated with an increment of .01SD for the semantic factor. It is important to note, however, that no beta value exceeded .05SD, and two of them were weakly negative. In addition, the associated standard errors are many times larger than the coefficients themselves, so that none of these models even approached statistical significance.

Sex

The WRRP sample as a whole included more girls than boys: at HV2 it was 58.5% female, 41.5% male. This asymmetry is more pronounced in the subsample, which is 66.7% female and 33.3% male. Parental education profiles were quite different for the girls and the boys in the subsample, as illustrated in Figure 3.2.

The utility of sex as an explanatory variable was probed in a series of regression models summarized in the bottom third of Table 3.16, in which sex was used as the sole predictor of factor scores at each home visit. Its effects were inconsistent: at HV2, for instance, being male was associated with a small positive beta (0.13) for semantic skill and a small negative beta (-0.08) for syntactic skill. For all models, the standard errors were much larger than their coefficients, and sex never reached statistical significance as an explanatory variable for any of the measures under consideration.

Model-Fitting with the Moderating Variables

Despite the unimpressive performance of breastfeeding and sex in preliminary model-fitting, the literature offers substantial support for their inclusion in models of long-term effects (Lucas, Morley, Cole, Lister, & Leeson-Payne, 1992; Stevenson et al., 2000; Ingemarsson, 2003). There is support, furthermore, for a number of potential interactions. Breastfeeding may interact with parental education (Oddy, 2006) and with sex (Broad & Duganzich, 1983, Mahurin-Smith & Ambrose, 2008), as well as with degree of prematurity. The different patterns of parental education for boys and girls in this sample raise the possibility of an interaction between those two variables.

Tables 3.17 and 3.18 present two models which evaluate the explanatory power of these potential moderating variables in conjunction with gestational age: a full model in Table 3.18, which includes all three proposed moderating variables and the interactions highlighted above, and a more parsimonious model in Table 3.17, which incorporates only gestational age and parental education. In the full model, no protective effects emerged for female sex or for breastfeeding. The full model supported the effects seen in previous models for gestational age, which showed significant predictive effects for the formal factor at HV2 and HV3, and a marginally significant effect for the semantic factor at HV2.

The more parsimonious model offers a clearer picture than the full model, suggesting three observations which will serve as a précis of this chapter’s findings. First, these results corroborate the models derived using perinatal risk factors as their sole predictors. Gestational age serves as a statistically significant predictor for lower formal factor scores at HV2 and HV3, and as a marginally significant predictor for lower semantic factor scores at HV2 and lower formal factor scores at HV5. Second, increasing parental education was generally associated with better performance, with a clearer effect noted at the later home visits. Third, the split-half R^2 values associated with these regression models never exceeded 0.18 and were often much lower, indicating a small effect size. In sum, then, these results indicate that prematurity is associated with a modest but measurable decrement in outcomes for the variables observed in the present study. It is evident that many other factors come into play in determining a child's performance on these measures, as reflected in the sizable standard errors and variable statistical significance seen here.

Table 3.1

Number of utterances in language samples at each home visit (HV2, HV3, and HV5) for both the perinatal risk factor (PRF) and full-term (FT) groups

	PRF mean (*SD*)	PRF range	FT mean (*SD*)	FT range
HV2	145.4 (50.5)	41-287	132.6 (45.2)	10-224
HV2 (examiner)	124.2 (40.9)	38-226	117.6 (33.1)	50-201
HV3	156.1 (49.2)	42-241	134.8 (45.3)	53-219
HV5	58.6 (24.6)	27-120	49.6 (17.3)	19-102

Table 3.2

Descriptive statistics for home visit 2 for both the perinatal risk factor (PRF) and full-term (FT) groups

	PRF mean (*SD*)	PRF mean z-score (*SD*)	FT mean (*SD*)	PRF - FT	PRFz - FTz
Adverb density	23.36 (8.60)	-0.04 (0.82)	24.13 (12.07)	-0.76	-0.08
Conjunction density	30.39 (14.99)	-0.15 (0.96)	35.16 (16.08)	-4.77	-0.30
Complex conjunction density	11.97 (6.45)	-0.07 (0.92)	12.89 (7.54)	-0.92	-0.13
Metalinguistic verb density	6.59 (3.07)	-0.12 (0.91)	7.40 (3.64)	-0.81	-0.24
Morphologically complex word density	4.28 (2.98)	-0.05 (0.93)	4.62 (3.43)	-0.33	-0.10
Low-frequency word density	4.80 (2.75)	-0.26 (0.80)	6.55 (3.85)	-1.75	-0.51
Elaborated noun phrase density	9.90 (5.02)	-0.19 (0.86)	12.13 (6.40)	-2.22	-0.38
Measure D	69.65 (11.57)	-0.09 (0.99)	71.72 (11.77)	-2.07	-0.18
MLU	5.60 (1.21)	-0.11 (0.96)	5.87 (1.30)	-0.27	-0.21
NDW-100	185.07 (29.98)	-0.22 (0.96)	199.19 (31.38)	-14.12	-0.45
NTW-100	512.80 (108.09)	-0.17 (0.91)	555.21 (126.50)	-42.41	-0.35
DSS	9.90 (1.75)	-0.03 (0.98)	10.00 (1.85)	-0.11	-0.06
IQ	97.82 (13.70)	-0.28 (0.99)	105.40 (13.11)	-7.59	-0.55
Digit span	50.67 (4.78)	-0.27 (0.89)	53.49 (5.53)	-2.82	-0.53

Note: All density values represent the raw count data divided by the number of utterances, and then multiplied by 100 for ease of interpretation. A density value of 25 would thus indicate that a target was observed once in every four utterances, on average, while a density value of 10 would correspond to an average frequency of one occurrence in every 10 utterances. The population mean for IQ as assessed via the Stanford-Binet short form is 100 (*SD* = 15); for digit span as assessed via the Stanford-Binet Memory for Digits subtest it is 50 (*SD* = 8). DSS: Developmental Sentence Scoring; MLU: mean length of utterance; NDW: number of different words; NTW: number of total words.

Table 3.3

Home visit 2 z *scores, separated by sex and group, for both the perinatal risk factor (PRF) and full-term (FT) groups*

	PRF girl means	PRF boy means	FT girl means	FT boy means
Adverb density	-0.15	0.18	-0.13	0.36
Conjunction density	-0.26	0.04	0.04	0.37
Complex conjunction density	-0.17	0.13	-0.03	0.25
Metalinguistic verb density	-0.15	-0.08	0.15	0.04
Morphologically complex word density	0.03	-0.21	-0.05	0.25
Low-frequency vocabulary density	-0.30	-0.18	0.02	0.71
Elaborated noun phrase density	-0.20	-0.18	0.10	0.35
Measure D	-0.20	0.12	0.01	0.24
IQ	-0.24	-0.35	0.31	0.19
Digit span	-0.12	-0.54	0.39	0.00
MLU	-0.17	0.02	0.01	0.29
NDW-100	-0.20	-0.26	-0.04	0.79
NTW-100	-0.13	-0.25	0.02	0.52
DSS	0.03	-0.16	0.05	-0.02

Note: All density values represent the raw count data divided by the number of utterances, and then multiplied by 100 for ease of interpretation. A density value of 25 would thus indicate that a target was observed once in every four utterances, on average, while a density value of 10 would correspond to an average frequency of one occurrence in every 10 utterances. The population mean for IQ as assessed via the Stanford-Binet short form is 100 (*SD* = 15); for digit span as assessed via the Stanford-Binet Memory for Digits subtest it is 50 (*SD* = 8). DSS: Developmental Sentence Scoring; MLU: mean length of utterance; NDW: number of different words; NTW: number of total words.

Table 3.4

Descriptive statistics for home visit 3 for both the perinatal risk factor (PRF) and full-term (FT) groups

	PRF mean (*SD*)	PRF mean z score (*SD*)	FT mean (*SD*)	PRF - FT	PRFz – FTz
Adverb density	23.62 (9.76)	-0.13 (0.86)	26.42 (12.41)	-2.81	-0.24
Conjunction density	33.79 (17.00)	-0.10 (0.94)	37.05 (18.96)	-3.26	-0.18
Complex conjunction density	14.05 (7.52)	-0.01 (0.89)	14.22 (9.28)	-0.17	-0.02
Metalinguistic verb density	6.82 (3.17)	-0.13 (0.92)	7.62 (3.65)	-0.81	-0.24
Morphologically complex word density	4.62 (3.10)	-0.005 (0.98)	4.64 (3.24)	-0.03	-0.01
Low-frequency word density	6.89 (3.17)	-0.27 (0.61)	9.48 (6.22)	-2.58	-0.50
Elaborated noun phrase density	12.88 (6.14)	-0.13 (0.85)	14.57 (8.07)	-1.68	-0.24
DSS	10.21 (1.80)	-0.16 (0.99)	10.76 (1.82)	-0.55	-0.30
IQ	96.12 (11.14)	-0.43 (0.81)	107.16 (13.89)	-11.04	-0.80
Digit span	50.93 (5.05)	-0.26 (0.91)	53.57 (5.70)	-2.64	-0.48
MLU	5.83 (1.41)	-0.05 (0.92)	5.96 (1.63)	-0.14	-0.09
NDW-100	191.53 (28.41)	-0.16 (0.91)	200.81 (33.06)	-9.27	-0.30
NTW-100	538.72 (112.90)	-0.08 (0.87)	558.31 (143.84)	-19.59	-0.15
Measure D	73.07 (14.53)	0.03 (1.06)	72.35 (13.23)	0.72	0.05

Note: All density values represent the raw count data divided by the number of utterances, and then multiplied by 100 for ease of interpretation. A density value of 25 would thus indicate that a target was observed once in every four utterances, on average, while a density value of 10 would correspond to an average frequency of one occurrence in every 10 utterances. The population mean for IQ as assessed via the Stanford-Binet short form is 100 (*SD* = 15); for digit span as assessed via the Stanford-Binet Memory for Digits subtest it is 50 (*SD* = 8). DSS: Developmental Sentence Scoring; MLU: mean length of utterance; NDW: number of different words; NTW: number of total words.

Table 3.5

Home visit 3 z *scores, separated by sex and group for both the perinatal risk factor (PRF) and full-term (FT) groups*

	PRF girl means	PRF boy means	FT girl means	FT boy means
Adverb density	-0.15	-0.11	-0.17	0.77
Conjunction density	-0.11	-0.08	-0.13	0.58
Complex conjunction density	-0.04	0.04	-0.17	0.42
Metalinguistic verb density	-0.13	-0.11	0.07	0.18
Morphologically complex word density	0.15	-0.26	-0.15	0.37
Low-frequency word density	-0.36	-0.12	-0.06	0.90
Elaborated noun phrase density	-0.10	-0.17	-0.09	0.57
Measure D	-0.10	0.24	-0.11	0.18
MLU	-0.05	-0.04	-0.19	0.55
NDW-100	-0.25	-0.03	-0.20	0.82
NTW-100	-0.14	0.00	-0.14	0.50
DSS	-0.17	-0.16	0.03	0.38
IQ	-0.36	-0.56	0.28	0.56
Digit span	-0.15	-0.42	0.17	0.33

Note: All density values represent the raw count data divided by the number of utterances, and then multiplied by 100 for ease of interpretation. A density value of 25 would thus indicate that a target was observed once in every four utterances, on average, while a density value of 10 would correspond to an average frequency of one occurrence in every 10 utterances. The population mean for IQ as assessed via the Stanford-Binet short form is 100 (*SD* = 15); for digit span as assessed via the Stanford-Binet Memory for Digits subtest it is 50 (*SD* = 8). DSS: Developmental Sentence Scoring; MLU: mean length of utterance; NDW: number of different words; NTW: number of total words.

Table 3.6.

Descriptive statistics for home visit 5 for the perinatal risk factor (PRF) and full-term (FT) groups

	PRF mean (*SD*)	PRF mean z score (*SD*)	FT mean (*SD*)	PRF - FT	PRFz – FTz
Adverb density	30.73 (11.67)	-0.10 (0.98)	33.04 (12.07)	-2.31	-0.20
Conjunction density	77.86 (21.03)	-0.05 (1.07)	79.83 (18.22)	-1.97	-0.10
Complex conjunction density	27.38 (12.05)	-0.01 (0.95)	27.64 (13.45)	-0.26	-0.02
Metalinguistic verb density	19.31 (9.25)	0.005 (1.08)	19.23 (7.90)	0.08	0.01
Morphologically complex word density	6.43 (5.63)	-0.11 (1.08)	7.57 (4.78)	-1.14	-0.22
Low-frequency word density	4.95 (4.66)	-0.04 (1.08)	5.33 (3.99)	-0.38	-0.08
Elaborated noun phrase density	16.50 (6.96)	-0.02 (1.01)	16.71 (6.87)	-0.21	-0.04
D	62.30 (11.97)	-0.07 (1.00)	64.08 (12.09)	-1.78	-0.15
MLU	9.44 (1.32)	-0.03 (1.00)	9.52 (1.32)	-0.08	-0.06
NDW-50	165.00 (16.91)	-0.12 (0,85)	170.50 (23.77)	-5.50	-0.28
NTW-50	411.45 (52.51)	-0.11 (0.91)	426.94 (65.02)	-15.48	-0.27
DSS	12.65 (1.76)	-0.21 (0.89)	13.63 (2.17)	-0.98	-0.49
CELF Recalling Sentences subtest	9.97 (2.83)	-0.32 (0.93)	11.93 (2.99)	-1.95	-0.64
CELF Understanding Paragraphs subtest	8.33 (3.30)	-0.04 (1.00)	8.59 (3.38)	-0.25	-0.08
CELF Word Classes (expressive) subtest	8.68 (2.71)	-0.24 (0.93)	10.05 (2.99)	-1.37	-0.47
CELF Word Classes (receptive) subtest	10.45 (2.86)	-0.24 (0.94)	11.76 (3.09)	-1.31	-0.45
TNL standard score	9.12 (2.00)	0.06 (0.96)	8.85 (2.19)	0.27	0.13

Note: All density values represent the raw count data divided by the number of utterances, and then multiplied by 100 for ease of interpretation. A density value of 25 would thus indicate that a target was observed once in every four utterances, on average, while a density value of 10 would correspond to an average frequency of one occurrence in every 10 utterances. The population mean for IQ is 100 (*SD* = 15); for digit span it is 50 (*SD* = 8). Clinical Evaluation of Language Fundamentals subtests and the Test of Narrative Language (TNL) have a population mean of 10 (*SD* = 3).

Table 3.7

HV5 z *scores, separated by sex and group for both the perinatal risk factor (PRF) and full-term (FT) groups*

	PRF girl means	PRF boy means	FT girl means	FT boy means
Adverb density	-0.20	0.10	-0.05	0.36
Conjunction density	-0.08	0.00	-0.14	0.39
Complex conjunction density	-0.08	0.12	-0.20	0.39
Metalinguistic verb density	-0.09	0.19	-0.14	0.23
Morphologically complex word density	-0.15	-0.04	-0.12	0.53
Low-frequency word density	-0.01	-0.11	-0.06	0.23
Elaborated noun phrase density	-0.05	0.06	0.03	-0.01
DSS	-0.18	-0.32	0.23	0.40
Measure D	-0.01	-0.19	0.00	0.21
MLU	-0.15	0.20	-0.30	0.62
NDW 50	-0.12	-0.11	0.03	0.46
NTW 50	-0.07	-0.26	0.07	0.35
CELF - Recalling Sentences	-0.29	-0.38	0.35	0.25
CELF - Understanding Paragraphs	-0.11	0.09	-0.04	0.19
CELF - Word Classes (expressive)	-0.22	-0.26	0.20	0.28
CELF - Word Classes (receptive)	-0.10	-0.41	0.26	0.14

Note: All density values represent the raw count data divided by the number of utterances, and then multiplied by 100 for ease of interpretation. A density value of 25 would thus indicate that a target was observed once in every four utterances, on average, while a density value of 10 would correspond to an average frequency of one occurrence in every 10 utterances. Clinical Evaluation of Language Fundamentals subtests have a population mean of 10 (*SD* = 3).

Table 3.8.
Pearson correlation coefficients at home visit 2

	ADV	CNJ	CCNJ	ML	MOR	LF	ENP	D	DSS	MLU	NDW	NTW	IQ	Digit span
Adverb density (ADV)	1.00													
Conjunction density (CNJ)	0.57**	1.00												
Complex conjunction density (CCNJ)	0.62**	0.76**	1.00											
Metalinguistic verb density (ML)	0.18	0.17	0.16	1.00										
Morphologically complex word density (MOR)	0.47**	0.18	0.30**	0.21*	1.00									
Low-frequency word density (LF)	0.50**	0.48**	0.44**	0.19*	0.25**	1.00								
Elaborated noun phrase density (ENP)	0.48**	0.61**	0.52**	0.07	0.28**	0.34**	1.00							
D	0.19*	-0.06	0.16	-0.02	0.16	0.15	0.21*	1.00						
DSS	0.52**	0.61**	0.72**	0.21*	0.20*	0.32**	0.61**	0.25**	1.00					
MLU	0.73**	0.85**	0.76**	0.20*	0.34**	0.52**	0.71**	0.13	0.69**	1.00				
NDW-100	0.55**	0.69**	0.60**	0.08	0.15	0.50**	0.56**	0.28**	0.54**	0.78**	1.00			
NTW-100	0.57**	0.76**	0.65**	0.11	0.20*	0.39**	0.62**	0.00	0.59**	0.88**	0.87**	1.00		
IQ	0.23*	0.16	0.21*	0.14	0.22*	0.26**	0.27**	0.07	0.32**	0.26**	0.31**	0.31**	1.00	
Digit span	0.04	-0.02	0.01	0.11	0.10	0.00	0.14	0.06	0.16	0.07	0.15	0.17	0.65**	1.00

Note: Most abbreviations are spelled out in the column at far left. DSS: Developmental Sentence Scoring; MLU: mean length of utterance; NDW: number of different words; NTW: number of total words. IQ and digit span were assessed using the Stanford-Binet short form and the Stanford-Binet Memory for Digits subtest, respectively.

Table 3.9

Pearson correlation coefficients at home visit 3

	ADV	CNJ	CCNJ	ML	MOR	LF	ENP	D	DSS	MLU	NDW-100	NTW-100	IQ	Digit span
Adverb density (ADV)	1.00													
Conjunction density (CNJ)	0.66**	1.00												
Complex conjuntion density (CCNJ)	0.66**	0.77**	1.00											
Metalinguistic verb density (ML)	0.12	0.15	0.21*	1.00										
Morphologically complex word density (MOR)	0.55**	0.33**	0.38**	-0.22*	1.00									
Low-frequency word density (LF)	0.55**	0.52**	0.37**	-0.02	0.49**	1.00								
Elaborated noun phrase density (ENP)	0.69**	0.70**	0.48**	0.14	0.41**	0.65**	1.00							
D	-0.05	-0.16	-0.01	-0.14	0.07	0.00	-0.03	1.00						
DSS	0.58**	0.69**	0.78**	0.20	0.36**	0.38**	0.59**	0.09	1.00					
MLU	0.83**	0.85**	0.76**	0.18	0.49**	0.57**	0.81**	-0.07	0.73**	1.00				
NDW-100	0.66**	0.78**	0.63**	0.26*	0.44**	0.70**	0.69**	0.13	0.70**	0.82**	1.00			
NTW-100	0.75**	0.85**	0.78**	0.33**	0.40**	0.55**	0.75**	-0.22*	0.79**	0.97**	0.85**	1.00		
IQ	0.34**	0.24*	0.17	0.06	0.30**	0.46**	0.36**	-0.07	0.22*	0.32**	0.33**	0.35**	1.00	
Digit span	0.06	0.17	0.15	-0.07	0.21*	0.08	0.11	0.02	0.16	0.13	0.21*	0.14	0.54**	1.00

Note: Most abbreviations are spelled out in the column at far left. DSS: Developmental Sentence Scoring; MLU: mean length of utterance; NDW: number of different words; NTW: number of total words. IQ and digit span were assessed using the Stanford-Binet short form and the Stanford-Binet Memory for Digits subtest, respectively.

Table 3.10

Pearson correlation coefficients at home visit 5

	ADV	CNJ	CCNJ	ML	MOR	LF	ENP	D	DSS	MLU	NDW	NTW	C1	C2	C3	C4	TNL
Adverb density (ADV)	1.00																
Conjunction density (CNJ)	0.23*	1.00															
Complex conjunction density (CCNJ)	0.18	0.67**	1.00														
Metalinguistic verb density (ML)	0.03	0.12	0.19	1.00													
Morphologically complex word density (MOR)	-0.01	0.08	0.25*	0.45**	1.00												
Low-frequency word density (LF)	0.21*	0.00	0.19	0.14	0.22*	1.00											
Elaborated noun phrase density (ENP)	0.16	0.06	0.25*	0.09	0.13	0.25*	1.00										
D	0.06	-0.22*	0.10	0.25*	0.29**	0.41**	0.20	1.00									
DSS	0.05	0.41**	0.52**	0.17	0.11	0.27*	0.50**	0.17	1.00								
MLU	0.24*	0.69**	0.65**	0.32**	0.26*	0.29**	0.38**	0.15	0.81**	1.00							
NDW-50	0.21	0.16	0.30**	0.21	0.30**	0.38**	0.48**	0.58**	0.63**	0.67**	1.00						
NTW-50	0.22*	0.60**	0.53**	0.22*	0.21*	0.16	0.42**	0.00	0.80**	0.95**	0.72**	1.00					
CELF Recalling Sentences (C1)	-0.16	-0.03	0.18	0.03	0.20	0.02	0.21*	0.18	0.22*	0.16	0.27*	0.17	1.00				
CELF Understanding Paragraphs (C2)	-0.11	-0.02	0.16	0.08	0.19	0.17	0.25*	0.37**	0.31**	0.21	0.18	0.14	0.57**	1.00			
CELF Word Classes – exp (C3)	-0.01	0.08	0.35**	0.09	0.25*	0.03	0.23*	0.40**	0.30**	0.30**	0.28**	0.26*	0.71**	0.57**	1.00		
CELF word classes – rec (C4)	-0.04	-0.10	0.20	0.05	0.16	0.11	0.25*	0.45**	0.14	0.15	0.27*	0.09	0.66**	0.50**	0.80**	1.00	
Test of Narrative Language (TNL)	0.21*	0.11	0.22*	0.29**	0.22*	0.26*	0.06	0.45**	0.05	0.29**	0.25*	0.21*	0.25*	0.39**	0.35**	0.39**	1.00

Note: Most abbreviations are spelled out in the column at far left. DSS: Developmental Sentence Scoring; MLU: mean length of utterance; NDW: number of different words; NTW: number of total words; CELF: Clinical Evaluation of Language Fundamentals.

Table 3.11

Factor loadings for the three factors at home visit 2 (HV2), home visit 3 (HV3), and home visit 5 (HV5)

	semantic HV2	syntactic HV2	formal HV2	semantic HV3	syntactic HV3	formal HV3	semantic HV5	syntactic HV5	formal HV5
Adverb density	.68			.82			.20		
Metalinguistic verb density	.16			.24			.25		
Morphologically complex word density	.28			.48			.22		
Low-frequency word density	.53			.61			.20		
Measure D	.13			-.09			.10		
NDW	.88			.86			.73		
NTW	.93			.96			.96		
Conjunction density		.87			.88			.67	
Complex conjunction density		.80			.79			.64	
Elaborated noun phrase density		.72			.81			.47	
DSS		.97			.80			.81	
MLU		.76			.95			.91	
Digit span			.39			.52			
IQ			1.67			1.03			
CELF – Recalling Sentences									.77
CELF – Word Classes (Rec)									.85
CELF – Word Classes (Exp)									.92
CELF – Understanding Paragraphs									.68
Test of Narrative Language									.41

Note: All density values represent the raw count data divided by the number of utterances, and then multiplied by 100 for ease of interpretation. The population mean for IQ is 100 (*SD* = 15); for digit span it is 50 (*SD* = 8). Clinical Evaluation of Language Fundamentals subtests and the Test of Narrative Language (TNL) have a population mean of 10 (*SD* = 3). DSS: Developmental Sentence Scoring; MLU: mean length of utterance; NDW: number of different words; NTW: number of total words.

Table 3.12

Regression coefficients for binary models at each home visit (HV)

	Intercept (*SE*)	*t* value	Slope (*SE*)	*t* value	Split-half adjusted R^2
HV2-semantic	0.16 (0.15)	1.06 (NS)	-0.31 (0.21)	-1.48 (NS)	NA
HV2-syntactic	0.13 (0.15)	-1.22 (NS)	-0.26 (0.21)	-1.22 (NS)	NA
HV2-formal	0.27 (0.16)	1.64 (NS)	-0.54 (0.23)	-2.32*	.05, .06
HV3-semantic	0.17 (0.16)	1.06 (NS)	-0.35 (0.24)	-1.46 (NS)	NA
HV3-syntactic	0.09 (0.16)	0.58 (NS)	-0.18 (0.25)	-0.73 (NS)	NA
HV3-formal	0.33 (0.16)	2.13*	-0.72 (0.23)	-3.12**	.14, .10
HV5-semantic	0.10 (0.16)	0.64 (NS)	-0.21 (0.23)	-0.91 (NS)	NA
HV5-syntactic	0.05 (0.17)	0.28 (NS)	-0.10 (0.24)	-0.40 (NS)	NA
HV5-formal	0.22 (0.19)	1.13 (NS)	-0.43 (0.27)	-1.61 (NS)	NA

Note: Split-half adjusted R^2 values are reported for all models in which the multilevel approach yielded one or more *p* values <.10. The value obtained for firstborn twins appears first in the cell, followed by the value for secondborn twins.

Table 3.13

Regression models using continuous predictors at each home visit (HV)

	Intercept (*SE*)	*t* value	Slope (*SE*)	*t* value	Split half adjusted R^2
Gestational age as a predictor					
HV2 semantic	-1.46 (0.85)	-1.72-	0.04 (0.02)	1.74-	.02, .02
HV2 syntactic	-1.29 (0.85)	-1.51 (NS)	0.04 (0.02)	1.51 (NS)	NA
HV2 formal	-2.43 (0.92)	-2.63*	0.07 (0.03)	2.63*	.07, .09
HV3 semantic	-0.99 (1.05)	-0.94 (NS)	0.03 (0.03)	0.97 (NS)	NA
HV3 syntactic	-0.43 (1.06)	-0.41 (NS)	0.01 (0.03)	0.42 (NS)	NA
HV3 formal	-3.25 (0.98)	-3.31**	0.09 (0.03)	3.33**	.15, .14
HV5 semantic	-0.81 (0.94)	-0.86 (NS)	0.02 (0.03)	0.89 (NS)	NA
HV5 syntactic	-0.48 (0.98)	-0.49 (NS)	0.01 (0.03)	0.51 (NS)	NA
HV5 formal	-2.61 (1.07)	-2.44*	0.07 (0.03)	2.43*	.14, .01
Birthweight as a predictor					
HV2 semantic	-0.25 (0.34)	-0.72 (NS)	0.05 (0.07)	0.72 (NS)	NA
HV2 syntactic	-0.08 (0.34)	-0.23 (NS)	0.01 (0.07)	0.17 (NS)	NA
HV2 formal	-0.65 (0.36)	-1.82 -	0.14 (0.07)	1.90 -	.07, .02
HV3 semantic	-0.48 (0.38)	-1.26 (NS)	0.09 (0.08)	1.23 (NS)	NA
HV3 syntactic	-0.18 (0.39)	-0.46 (NS)	0.03 (0.08)	0.38 (NS)	NA
HV3 formal	-1.18 (0.36)	-3.24**	0.25 (0.07)	3.42**	.12, .16
HV5 semantic	-0.56 (0.35)	-1.58 (NS)	0.11 (0.07)	1.55 (NS)	NA
HV5 syntactic	-0.33 (0.38)	-0.88 (NS)	0.07 (0.08)	0.91 (NS)	NA
HV5 formal	-0.54 (0.41)	-1.32 (NS)	0.12 (0.08)	1.43 (NS)	NA
Hospital stay as a predictor					
HV2 semantic	0.18 (0.14)	1.25 (NS)	-0.006 (0.004)	-1.64 (NS)	NA
HV2 syntactic	0.15 (0.15)	1.03 (NS)	-0.005 (0.004)	-1.31 (NS)	NA
HV2 formal	0.35 (0.15)	2.35*	-0.010 (0.004)	-3.05**	.08, .05
HV3 semantic	0.12 (0.15)	0.81 (NS)	-0.005 (0.004)	-1.24 (NS)	NA
HV3 syntactic	0.05 (0.15)	0.35 (NS)	-0.002 (0.004)	-0.44 (NS)	NA
HV3 formal	0.22 (0.15)	1.44 (NS)	-0.010 (0.004)	-2.41*	.09, .00
HV5 semantic	0.03 (0.15)	0.20 (NS)	-0.003 (0.004)	-0.66 (NS)	NA
HV5 syntactic	0.15 (0.15)	0.98 (NS)	-0.006 (0.004)	-1.46 (NS)	NA
HV5 formal	0.20 (0.18)	1.12 (NS)	-0.006 (0.005)	-1.38 (NS)	NA
Degree of prematurity (mean of scaled gestational age and the opposite of scaled hospital stay)					
HV2 semantic	0.01 (0.10)	0.14 (NS)	0.20 (0.11)	1.86-	.02, .02
HV2 syntactic	0.01 (0.11)	0.08 (NS)	0.16 (0.11)	1.48 (NS)	NA
HV2 formal	0.02 (0.17)	0.17 (NS)	0.35 (0.12)	2.94*	.07, .07
HV3 semantic	0.01 (0.12)	0.06 (NS)	0.15 (0.13)	1.21 (NS)	NA
HV3 syntactic	0.01 (0.12)	0.10 (NS)	0.05 (0.12)	0.40 (NS)	NA
HV3 formal	-0.01 (0.12)	-0.07 (NS)	0.37 (0.12)	2.95*	.14, .08
HV5 semantic	0.01 (0.11)	0.05 (NS)	0.09 (0.12)	0.80 (NS)	NA
HV5 syntactic	0.01 (0.12)	0.08 (NS)	0.14 (0.12)	1.17 (NS)	NA
HV5 formal	0.01 (0.13)	0.10 (NS)	0.25 (0.14)	1.88-	.07, -.01

Note: Split-half adjusted R^2 values are reported for all models in which the multilevel approach yielded one or more *p* values <.10. The value obtained for firstborn twins appears first in the cell, followed by the value for secondborn twins.

Table 3.14

Gestational age as a predictor within the perinatal risk factor (PRF) group at each home visit (HV)

	Intercept (*SE*)	*t* value	Slope (*SE*)	*t* value	Split-half adjusted R^2
HV2 semantic	-3.26 (2.28)	-1.43 (NS)	0.10 (0.08)	1.36 (NS)	NA
HV2 syntactic	-2.33 (2.44)	-0.95 (NS)	0.07 (0.08)	0.90 (NS)	NA
HV2 formal	-0.83 (2.86)	-0.29 (NS)	0.02 (0.10)	0.19 (NS)	NA
HV3 semantic	3.48 (2.26)	1.54 (NS)	-0.12 (0.08)	-1.63 (NS)	NA
HV3 syntactic	2.49 (2.97)	0.84 (NS)	-0.09 (0.10)	-0.87 (NS)	NA
HV3 formal	-1.38 (2.58)	-0.53 (NS)	0.03 (0.09)	0.38 (NS)	NA
HV5 semantic	-0.58 (2.61)	-0.22 (NS)	0.02 (0.09)	0.18 (NS)	NA
HV5 syntactic	-0.33 (2.69)	-0.12 (NS)	0.01 (0.09)	0.11 (NS)	NA
HV5 formal	-2.11 (3.05)	-0.69 (NS)	0.06 (0.10)	0.62 (NS)	NA

Note: Split-half adjusted R^2 values are reported for all models in which the multilevel approach yielded one or more *p* values <.10. Since none of the coefficients in these models reached statistical significance, no values appear in the R^2 column.

Table 3.15

Highest level of education attained by the primary caregiver for full-term (FT) and perinatal risk factor (PRF) groups, reported separately by child sex

	1	2	3	4	5	6	7	8	9
PRF parent education	0	0	12 (10, 2)	10 (8, 2)	3 (2, 1)	12 (6, 6)	4 (4, 0)	16 (8, 8)	0
FT parent education	0	0	10 (7, 3)	10 (6, 4)	5 (3, 2)	12 (11, 1)	4 (0, 4)	16 (11, 5)	0
Expected frequency in the PRF group, based on full-sample proportions	<1	<1	4	9	6	19	4	13	<1

Note: Values in parentheses show the sex breakdown for results, with girls first and boys second. As an example, of the 12 PRF children whose parents reported receiving a high school diploma, 10 were girls and 2 were boys.

1. Grade 6 or less
2. Grade 7-12 (without graduating high school or equivalent)
3. Graduated high school or equivalent
4. Some college
5. Graduated from 2-year college
6. Graduated from 4-year college
7. Attended graduate or professional school without graduating
8. Completed graduate or professional school
9. Other

Table 3.16

Regression models using moderating variables as sole predictors at each home visit (HV)

	Intercept (*SE*)	*t* value	Slope (*SE*)	*t* value	Split half adjusted R^2
Parental education as a predictor					
HV2 semantic	-0.32 (0.33)	-0.94 (NS)	0.06 (0.06)	1.01 (NS)	NA
HV2 syntactic	0.23 (0.34)	0.70 (NS)	-0.04 (0.06)	-0.74 (NS)	NA
HV2 formal	-0.65 (0.37)	-1.74-	0.17 (0.06)	1.85-	.08, .00
HV3 semantic	-0.78 (0.37)	-2.12*	0.14 (0.06)	2.27*	.17, -.01
HV3 syntactic	-0.45 (0.38)	-1.19 (NS)	0.08 (0.06)	1.30 (NS)	NA
HV3 formal	-0.48 (0.39)	-1.24 (NS)	0.09 (0.07)	1.13 (NS)	NA
HV5 semantic	-0.56 (0.35)	-1.60 (NS)	0.10 (0.06)	1.69-	.05, -.02
HV5 syntactic	-0.76 (0.35)	-2.19*	0.14 (0.06)	2.31*	.07, .02
HV5 formal	-0.88 (0.41)	-2.16*	0.16 (0.07)	2.27*	.12, .02
Breastfeeding duration (log-transformed) as a predictor					
HV2 semantic	0.01 (0.11)	0.07 (NS)	0.01 (0.04)	0.22 (NS)	NA
HV2 syntactic	-0.01 (0.11)	-0.04 (NS)	-0.02 (0.04)	-0.42 (NS)	NA
HV2 formal	0.02 (0.12)	0.14 (NS)	0.05 (0.04)	1.24 (NS)	NA
HV3 semantic	0.02 (0.12)	0.14 (NS)	0.03 (0.05)	0.69 (NS)	NA
HV3 syntactic	0.02 (0.12)	0.16 (NS)	0.04 (0.05)	0.91 (NS)	NA
HV3 formal	0.002 (0.13)	0.02 (NS)	-0.01 (0.05)	-0.30 (NS)	NA
HV5 semantic	0.01 (0.12)	0.10 (NS)	0.02 (0.04)	0.45 (NS)	NA
HV5 syntactic	0.01 (0.12)	0.07 (NS)	0.01 (0.04)	0.30 (NS)	NA
HV5 formal	0.02 (0.14)	0.12 (NS)	0.03 (0.05)	0.68 (NS)	NA
Sex as a predictor					
HV2 semantic	-0.04 (0.13)	-0.31 (NS)	0.13 (0.23)	0.59 (NS)	NA
HV2 syntactic	0.03 (0.13)	0.22 (NS)	-0.08 (0.23)	-0.37 (NS)	NA
HV2 formal	-0.03 (0.15)	-0.17 (NS)	0.08 (0.26)	0.32 (NS)	NA
HV3 semantic	-0.08 (0.15)	-0.50 (NS)	0.24 (0.25)	0.96 (NS)	NA
HV3 syntactic	-0.02 (0.15)	-0.10 (NS)	0.08 (0.25)	0.31 (NS)	NA
HV3 formal	0.04 (0.16)	0.23 (NS)	-0.09 (0.26)	-0.33 (NS)	NA
HV5 semantic	-0.08 (0.14)	-0.58 (NS)	0.24 (0.24)	1.00 (NS)	NA
HV5 syntactic	-0.01 (0.14)	-0.06 (NS)	0.03 (0.25)	0.11 (NS)	NA
HV5 formal	0.03 (0.17)	0.16 (NS)	-.10 (0.29)	-0.32 (NS)	NA

Note: Split-half adjusted R^2 values are reported for all models in which the multilevel approach yielded one or more *p* values <.10. The value obtained for firstborn twins appears first in the cell, followed by the value for secondborn twins.

Table 3.17

Models fitted using gestational age (GA) and parental education at each home visit (HV)

	Intercept (*SE*)	*t* value	GA slope (*SE*)	*t* value	Parent ed slope (*SE*)	*t* value	Split-half adjusted R^2
HV2 semantic	-1.68 (0.86)	-1.96-	0.04 (0.02)	1.72-	0.06 (0.05)	1.04 (NS)	.04, .00
HV2 syntactic	-0.83 (0.88)	-0.94 (NS)	0.03 (0.02)	1.31 (NS)	-0.04 (0.06)	-0.73 (NS)	NA
HV2 formal	-2.65 (0.94)	-2.83**	0.06 (0.03)	2.31*	0.12 (0.06)	1.93-	.12, .06
HV3 semantic	-1.66 (0.98)	-1.69-	0.03 (0.03)	0.97 (NS)	0.14 (0.06)	2.22*	.18, -0.03
HV3 syntactic	-0.79 (1.02)	-0.77 (NS)	0.01 (0.03)	0.35 (NS)	0.08 (0.06)	1.27 (NS)	NA
HV3 formal	-3.20 (0.97)	-3.30**	0.08 (0.03)	3.02**	0.08 (0.06)	1.31 (NS)	.14, .09
HV5 semantic	-1.36 (0.90)	-1.51 (NS)	0.02 (0.03)	0.97 (NS)	0.09 (0.06)	1.59 (NS)	NA
HV5 syntactic	-1.18 (0.90)	-1.31 (NS)	0.01 (0.03)	0.51 (NS)	0.13 (0.06)	2.22*	.05, .00
HV5 formal	-2.57 (1.03)	-2.50*	0.05 (0.03)	1.79-	0.15 (0.07)	2.13*	.17, .01

Note: Split-half adjusted R^2 values are reported for all models in which the multilevel approach yielded one or more *p* values <.10. The value obtained for firstborn twins appears first in the cell, followed by the value for secondborn twins.

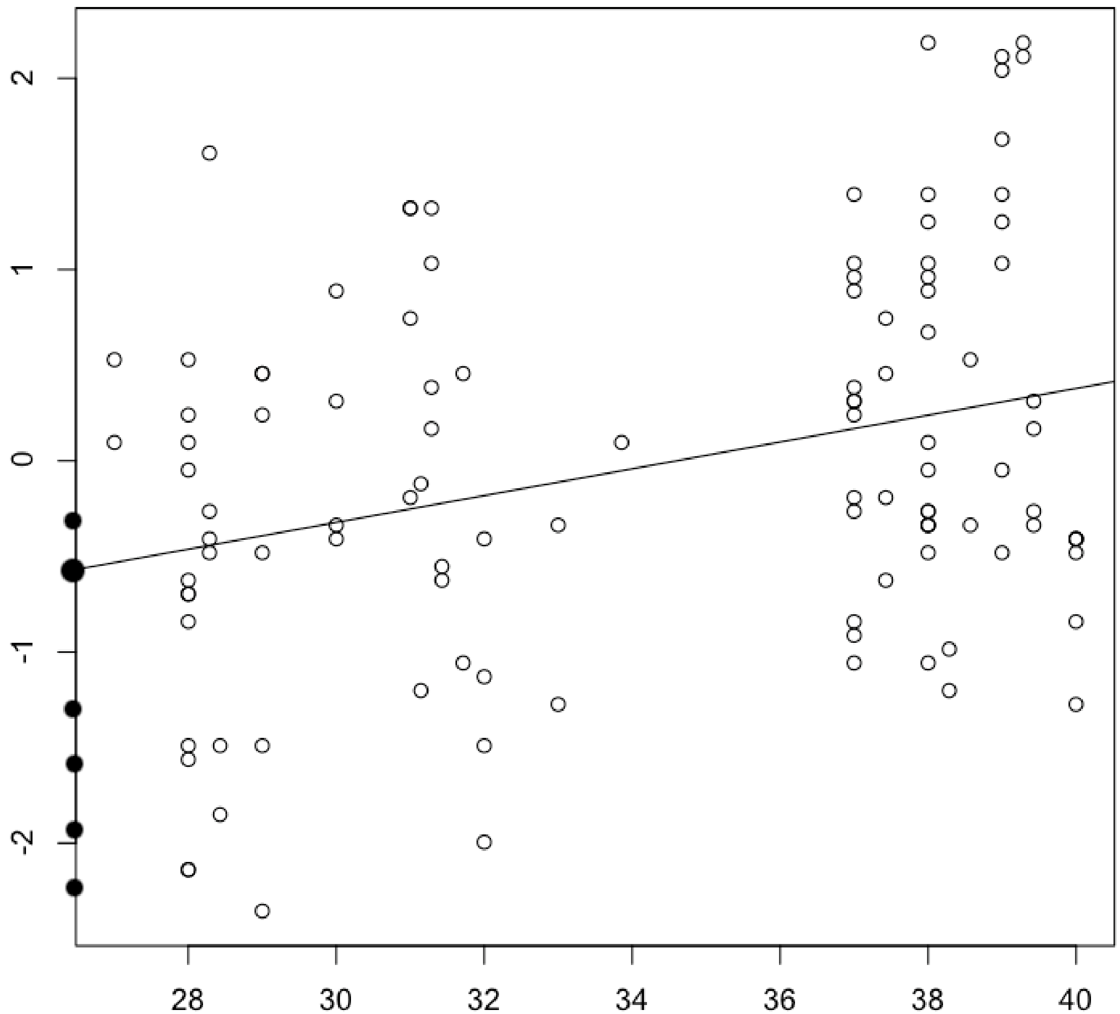
Table 3.18

Models fitted with all moderating variables and interactions of interest included at each home visit (HV)

	Intercept	GA	Parent edu	Breastfed	Sex	Bf*GA	Bf*Sex	Bf*Edu	Edu*Sex	Adjusted R^2
HV2 sem	-1.85*	0.04-	0.07 (NS)	-0.18 (NS)	0.41 (NS)	0.001 (NS)	0.03 (NS)	0.03 (NS)	-0.05 (NS)	-.01, -.01
HV2 syn	-0.90 (NS)	0.03 (NS)	-0.03 (NS)	0.01 (NS)	0.08 (NS)	-0.01 (NS)	-0.15 (NS)	0.04 (NS)	-0.02 (NS)	NA
HV2 form	-2.5*	0.06*	0.11 (NS)	0.30 (NS)	0.27 (NS)	-0.01 (NS)	-0.12 (NS)	0.01 (NS)	-0.05 (NS)	.04, .08
HV3 sem	-1.67 (NS)	0.04 (NS)	0.06 (NS)	-0.03 (NS)	-0.98 (NS)	-0.00 (NS)	0.00 (NS)	0.04 (NS)	0.20 (NS)	NA
HV3 syn	-0.88 (NS)	0.02 (NS)	0.02 (NS)	0.26 (NS)	-0.79 (NS)	-0.01 (NS)	0.02 (NS)	0.03 (NS)	0.15 (NS)	NA
HV3 form	-3.14**	0.08**	0.05 (NS)	0.73-	-0.53 (NS)	-0.02*	-0.10 (NS)	0.02 (NS)	0.07 (NS)	.13, .08
HV5 sem	-1.95*	0.04 (NS)	0.06 (NS)	-0.70*	-0.28 (NS)	0.02 (NS)	-0.01 (NS)	0.01 (NS)	0.10 (NS)	-.04, -.03
HV5 syn	-1.59 (NS)	0.02 (NS)	0.10 (NS)	-0.49 (NS)	-0.70 (NS)	0.01 (NS)	-0.02 (NS)	0.01 (NS)	0.12 (NS)	NA
HV5 form	-2.04-	0.03 (NS)	0.22*	0.40 (NS)	1.19 (NS)	-0.01 (NS)	-0.10 (NS)	0.00 (NS)	-0.26 (NS)	.13, .00

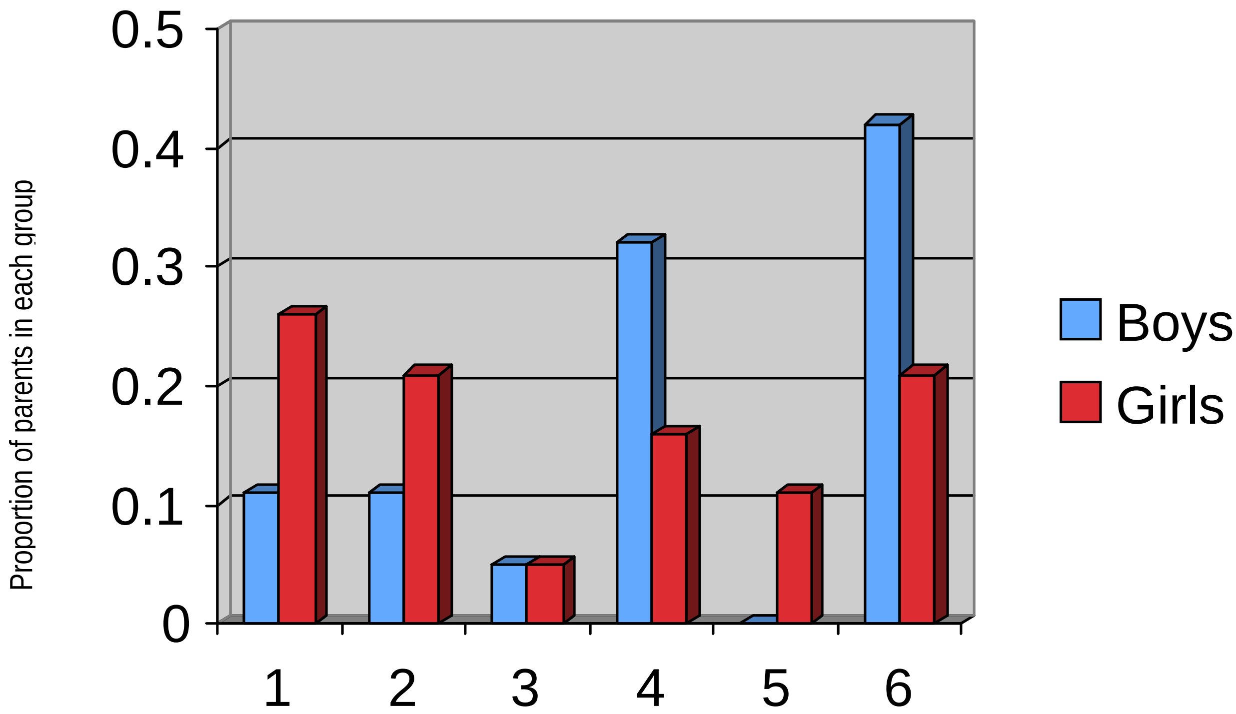
Note: Split-half adjusted R^2 values are reported for all models in which the multilevel approach yielded one or more p values <.10. The value obtained for firstborn twins appears first in the cell, followed by the value for secondborn twins.
GA: gestational age
Bf: breastfeeding
Sem: semantic factor
Syn: syntactic factor
Form: formal factor

Figure 3.1. A regression model with varying intercepts



Note: This figure illustrates the relationship between gestational age and formal factor scores at home visit 2. The large dot on the y axis shows the full-sample intercept; the smaller dots on the y axis show a sampling of the intercepts generated for each family using multilevel modeling. The slope was not allowed to vary, remaining the same across all of the families in the sample.

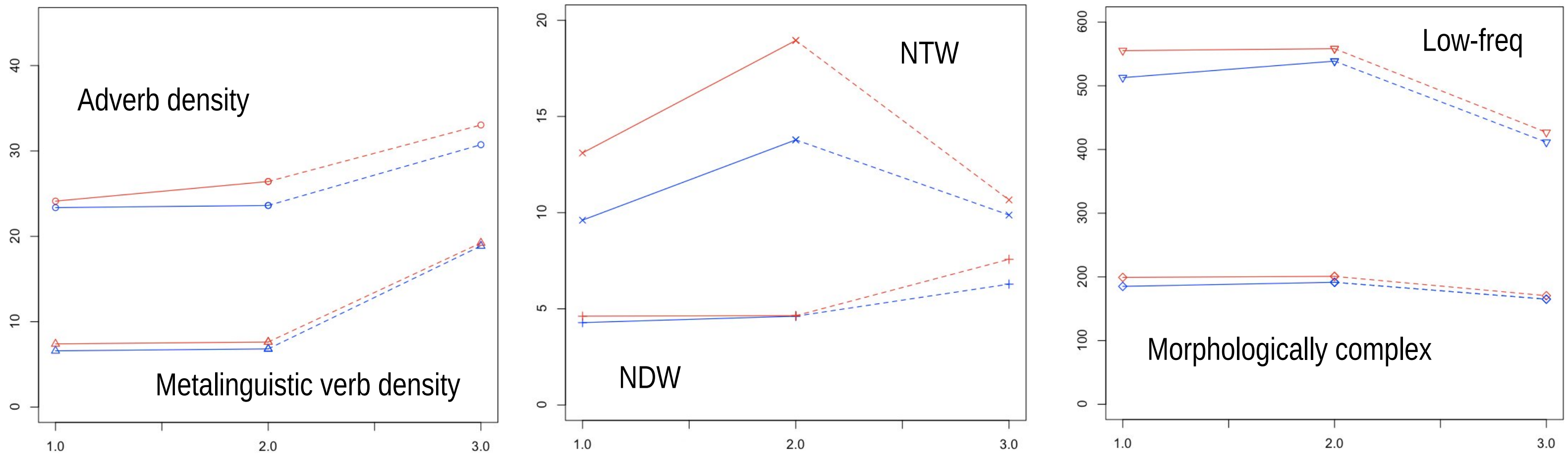
Figure 3.2. Parental education, separated by sex



Note: Figure shows the proportion of parents in each educational group; the corresponding numerical values are presented in Table 3.15.

1. Graduated high school or equivalent
2. Some college
3. Graduated from 2-year college
4. Graduated from 4-year college
5. Attended graduate or professional school without graduating
6. Completed graduate or professional school

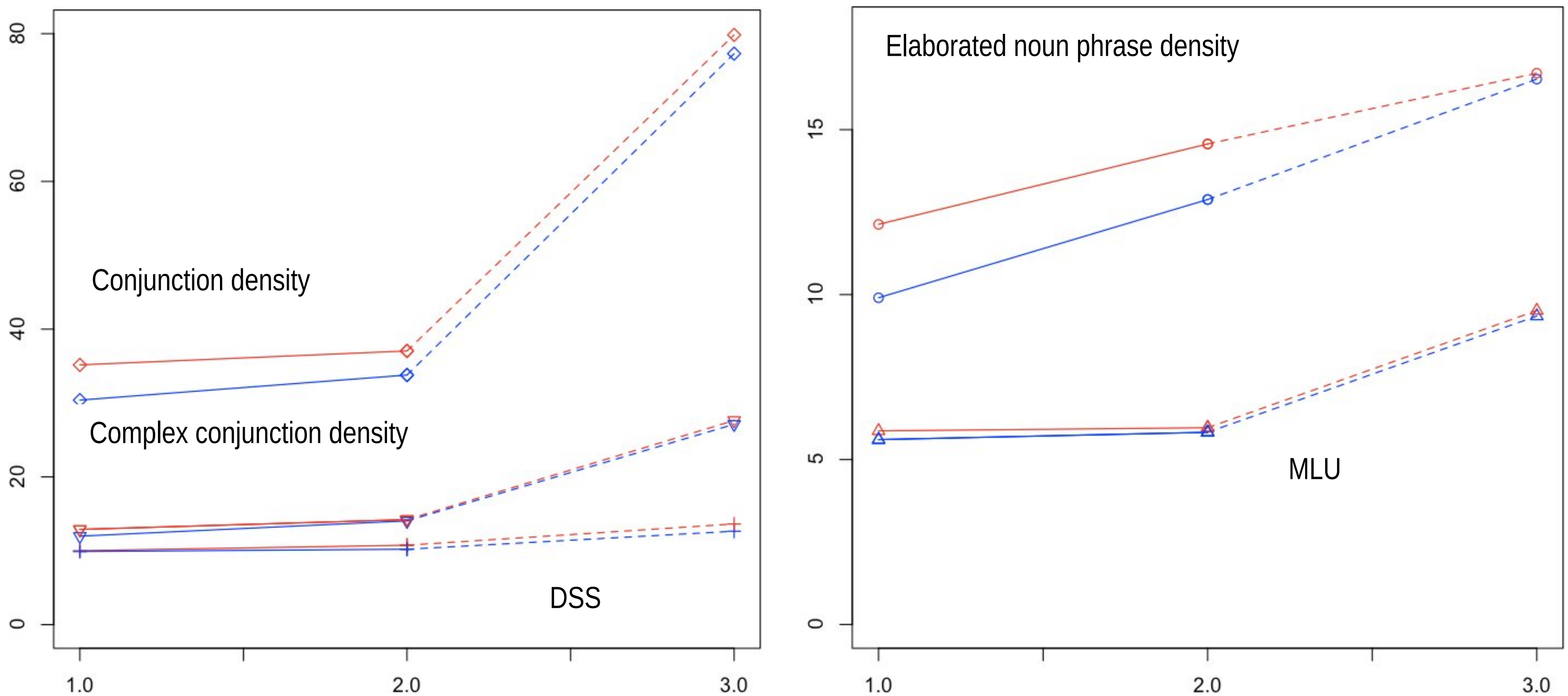
Figure 3.3. Growth in semantic measures



Note: In all three squares, red represents mean performance for control children and blue represents premature children. The square at the left shows growth in adverb density (top pair of lines) and metalinguistic verb density (bottom pair of lines). The center square shows NTW (top) and NDW (bottom). The square at the right illustrates changes in low-frequency vocabulary density (top) and morphologically complex word density (bottom). The x axis shows values at year 1 (home visit 2), year 2 (home visit 3), and year 3 (home visit 5); the y axis of each square is scaled to reflect the values shown in Tables 3.2, 3.4, and 3.6 with one exception: values for low-frequency word density were doubled to eliminate overlap between the pairs of lines. The dashed lines between values for year 2 and year 3 are intended to remind readers that year 3 results are drawn from a briefer narrative sample, rather than the 15-minute conversational sample obtained at year 1 and year 2. across home visits.

Readers are reminded that in all cases the confidence intervals around these means are wide and there is considerable overlap between the distributions. For adverb density at home visit 5, for instance, the 95% confidence interval would run from 29.3 to 36.8 for the full-term group, and from 27.1 to 34.3 for the premature group.

Figure 3.4. Growth in syntactic measures across home visits



Note: In both squares, red represents mean performance for control children and blue represents premature children. The square at the left shows growth in conjunction density (top pair of lines), complex conjunction density (middle pair of lines) and DSS (bottom pair of lines). The square at the right illustrates growth in elaborated noun phrase density (top pair of lines) and MLU (bottom pair of lines). The x axis shows values at year 1 (home visit 2), year 2 (home visit 3), and year 3 (home visit 5); the y axis of each square is scaled to reflect the values shown in Tables 3.2, 3.4, and 3.. The dashed lines between values for year 2 and year 3 are intended to remind readers that year 3 results are drawn from a briefer narrative sample, rather than the 15-minute conversational sample obtained at year 1 and year 2. across home visits.

Readers are reminded that in all cases the confidence intervals around these means are wide and there is considerable overlap between the distributions. For elaborated noun phrase density at home visit 5, for instance, the 95% confidence interval would run from 14.6 to 18.8 for the full-term group, and from 14.4 to 18.6 for the premature group.

CHAPTER FOUR: DISCUSSION

This chapter will first review the major findings of the present study and their possible interpretations. A summary of its limitations will follow, along with implications for future research and clinical practice.

Review of Primary Findings

The first research question asked whether between-group differences could be observed on the variables of interest. Together with many studies in the existing literature, these results indicate that school-aged children who were born prematurely tend to show deficits in performance on standardized tests in comparison with peers born at full term with no perinatal complications. This difference between groups was statistically significant at HV2 and HV3, when results were obtained for IQ and digit span, and at HV5, when the standardized tests included the CELF-4 and the TNL.

For the language sample measures assessed in the present study, group differences did not reach statistical significance. Despite the lack of statistical significance, they show a clear and consistent trend toward poorer language outcomes in the PRF group. Across all three home visits, the control children outscored the PRF children on more than 90% of measures assessed. Many of these differences were small, under a tenth of a standard deviation. At all three home visits, however, the difference was as large as half of a standard deviation for some measures. In all of the regression models fitted to explore the relationships between prematurity and language outcomes, the slopes associated with advancing gestational age were universally positive, with not a single exception. Although the substantial heterogeneity associated with children's language acquisition and use resulted in standard errors that were consistently large and thus in *p* values that were non-significant, it is proposed that these findings support the idea of modest decrements in association with prematurity for semantic and, to a lesser degree, syntactic skill.

This study also considered the hypothesis that there would be a performance gradient among PRF children, with a trend toward poorer outcomes in the children who were born earlier and smaller. This hypothesis is afforded some support by the success of continuous predictor variables in models that included both PRF and control children: if there were no such trend among the premature group, then specification of gestational age would not have improved the fit of the models obtained using a binary premature/full-term predictor variable. In regression models that looked solely at the PRF group, however, the expected gradient did not approach statistical significance; PRF children born at the later end of the range of gestational ages (32-33 weeks) did not consistently outperform PRF children born at the earlier end of the range of the gestational ages (27-28 weeks). Potential reasons for this finding are discussed in the following section.

The third research question under consideration focused on changes over time in the relationships between the two groups. A review of results across variables and home visits revealed two trends: growth for all children across a majority of measures, and a general narrowing of the gap between PRF and control children at HV5. These trends are illustrated in Figures 3.3 and 3.4.

The study's final research question concerned the importance of parental education, breastfeeding, and female sex as moderating variables. For both breastfeeding and female sex, no effect was found either alone or in combination with other variables. Increasing parental education was associated with a generally positive effect, a trend which became statistically significant more often in later home visits. When parental education and gestational age were entered together into regression equations, each retained its own share of predictive power. In other words, in this sample, where every parent had at least a high school diploma, it did not appear the putative effects of prematurity were actually attributable to lower levels of parental education.

These findings and their non-significant associated *p* values will inevitably raise the question of statistical power: was the study adequately powered to observe the effect it was designed to observe? Power calculations were based on a predicted difference of .5 SD between groups, an effect size observed for most of the standardized tests but for only a few of the language measures. Further discussion of this issue follows at the conclusion of the following section.

Interpretations

Standardized Tests and Language Sample Measures

This study was driven by the hypothesis that children born prematurely would use language with less semantic and syntactic complexity than that of their full-term counterparts due to differences in early neurodevelopment triggered by preterm birth and its medical sequelae. While the results lend some support to the idea of a modest decrement in those domains, it is smaller than hypothesized. For the standardized test results considered in the present study, the between-groups difference was of the expected magnitude based on previous studies, approximately .5 SD. This finding suggests that standardized tests may tap skills that are more vulnerable in children born prematurely, focused attention being a likely culprit (Elgen, Sommerfelt, & Markestad, 2002). An additional though not mutually exclusive interpretation is that the language tasks presented to the children were not as taxing as the standardized tests. Consider, for instance, a hypothetical conversation in which the examiner said, “Quick, let’s name all the kinds of dogs we can think of in one minute.” Such an exchange would likely prove more demanding than the meandering conversations analyzed here, and might reveal wider gaps between the two groups. Such an exchange, of course, would also be much lower in social validity than a more customary conversation.

While the examiners at HV2 and HV3 were instructed to facilitate conversation by following the child’s lead, this was not the case at HV5, when the TNL was administered. In the literature on language development in school-aged children, it is often proposed that narrative tasks are more difficult for children than simple conversations (Liles, 1993; Scott & Windsor, 2000). Consequently, one possible outcome for these data was a wider divergence between groups at HV5, when the task changed from a child-directed conversation to an examiner-directed narrative. Actually, however, the results showed the reverse effect: the between-groups differences across language measures became even smaller at HV5. In fact, as a group the PRF children scored higher than the full-term children on the Test of Narrative Language, though the difference was a modest .13 SD. One cautionary note must be sounded here, however; previous analysis of selected HV5 data indicated that the best predictor for TNL score was NTW (Mahurin-Smith & DeThorne, 2008). It is possible that the PRF group’s proclivity for telling slightly longer stories, illustrated in Table 3.1, may explain some of their higher scores although the scoring criteria are not directly tied to length.

One trend observed across all three home visits was a wider between-groups discrepancy for the semantic factor than for the syntactic factor. Three possible reasons for this finding are proposed. First, it might be that syntactic skills are robust enough that there is genuinely a narrower gap between the two groups in the syntactic domain. A second explanation is that the range of syntactic complexity in children's discourse-level language is relatively small, so that the two groups look fairly similar. In standardized tests such as the CELF-4, stimulus materials include sentences with dizzying assemblages of modal and auxiliary verbs or of nested passive constructions, the likes of which do not typically occur in conversations with young children. A third possibility is that the measures used in this study do not provide a sensitive index of differences between the two groups.

Potentially Useful Variables

The question of which indicators are most sensitive to group differences was not addressed directly by this study. There are disadvantages to conducting a study with multiple dependent variables, one of which is the risk that potentially valuable indicators may get lost in the resulting pile of information. Among the language measures, a few stood out as yielding the largest group differences; they may merit closer consideration in future research. First, the PRF group's use of low-frequency vocabulary was half of a standard deviation lower than the control group's at HV2 and HV3. While the difference was much smaller at HV5, this change is reasonable in view of the TNL requirement that children tell stories about prescribed topics. Second, use of elaborated noun phrases in conversation was also more common among full-term children than among children born prematurely, with effect sizes of .38 SD and .24 SD at HV2 and HV3 respectively. Both of these measures tap children's abilities to use elements of literate language and are known to be less prevalent among children who struggle with language (Greenhalgh & Strong, 2001; Ukrainetz & Gillam, 2009). Finally, NDW and NTW were consistently lower among the PRF children, sometimes by as much as .45 SD. An advantage to further investigations emphasizing these two measures is the existence of normative data and the ease of deriving them automatically using language analysis software.

In addition to highlighting useful outcome variables, it may also be helpful to touch on the question of useful predictor variables. In these data, the most useful predictor of performance on the measures of interest was gestational age in weeks. Birthweight performed poorly as a predictor of school-aged outcomes, perhaps because the two groups had overlapping distributions for birthweight. Twins tend to be lighter at birth than singletons, and in a small number of cases the full-term twins weighed less than their premature counterparts. It is also important to bear in mind that a higher birthweight may not signal an optimal uterine environment. Mothers with gestational diabetes, for instance, often deliver larger babies but those babies face increased risks of complications.

Unanswered Questions: Hypotheses and Possibilities

The Absence of a Gradient Effect in the PRF Group

Within the PRF cohort, it was somewhat surprising that there were no gradient effects observed for outcomes. When regression models were confined to the PRF group, the effects of gestational age on language outcomes were unpersuasive. Two explanations are proposed for this null finding. First, the range of prematurity observed in this study is compressed; there were no participants with gestational ages >34 weeks (due to the study's design) or <27 weeks (due to happenstance or selection bias). Second, regression models with only one of the present study's participant groups will likely have larger standard errors than both groups combined as a function of the decreased *n*, and are less likely to prove statistically meaningful. This point is illustrated in Figure 4.1, which contrasts scatterplots and regression lines for the full sample and the PRF group alone. While the regression line is slightly steeper in the half of the figure that shows the PRF group, the compressed range of the x axis and the smaller number of participants prevent the trend from reaching statistical significance.

Parental Education and Its Determinants

Of the moderating variables considered in this study, only parental education proved to have a statistically significant effect on the language and cognitive measures. One unexpected finding was the difference in parental education profiles between the PRF subgroup and the WRRP sample as a whole. A random sample of 57 WRRP families would be expected to include 4 families in which the primary caregiver held a high school diploma; in the PRF cohort there were 12. This disparity raises questions about the causal factors underlying prematurity. It is possible, for instance, that parents with less education wind up in jobs with fewer benefits, and that less access to preventative health care results in higher-risk pregnancies.

It is also possible, however, that another variable is driving the findings observed here: that the differences which are being ascribed to prematurity are in fact related to another factor, either environmental or genetic. For example, exposure to lead and certain pesticides can raise a woman’s risk of preterm delivery (Savitz, Whelan & Kleckner, 1989; Longnecker, Zhou, & Brock, 2001). It is not difficult to imagine a neighborhood in which dilapidated housing increases the risk of lead exposure for young children, or a farm community where agricultural toxins contaminate the water supply. For some children in those communities, educational careers may be cut short at least partially because of the cognitive impairment associated with toxin exposure. If those children grow up and start families in the same geographic location, they could face increased risks of both preterm labor and toxin exposure in their own children. In such a case, the exposure to toxins might play a causative role in all three variables: the parental educational level, the duration of pregnancy, and the children's later cognitive development. Another possible scenario is that genetic and environmental factors may work together to explain some of the differences seen in children born prematurely. Some evidence, for instance, suggests that genetics may explain some of the increased prevalence of preterm birth among African-American women (DeFranco, Teramo, & Muglia, 2007).

The two groups in this study were matched for parental education to reduce the risk of spurious differences, and the regression models which incorporated parental education did not diminish the effects of prematurity, decreasing the likelihood that a third variable is at work. Still, it is prudent to remember that the factors underlying both preterm labor and children's cognitive development are complex.

The Asymmetrical Sex Ratio

In addition to the puzzle of differing parental education profiles, these groups raise another question: where are the boys? The 2:1 ratio of girls to boys in this subsample does not reflect the ratio seen in NICUs, suggesting that fewer parents of premature boys elected to participate in WRRP. Those premature boys who did participate tended to have highly educated parents. One must wonder about the reasons for this sex discrepancy. One hypothesis stems from the fact that premature boys are more vulnerable to deficits in sustained attention than either premature girls or full-term boys (Reijneveld et al., 2006). It is possible that the parents of premature boys who might have considered participation in WRRP opted instead to avoid the difficulty of lengthy testing sessions with distractible boys. This is, of course, merely conjecture.

The Role of Breastfeeding

A third puzzle concerns the final moderating variable, breastfeeding, which did not have any significant effects. Two possible explanations for this finding are proposed. First, breastfeeding duration and exclusivity were relatively low in both groups; the median breastfeeding duration in the sample was 2 months, and 95% of mothers supplemented with infant formula. Since the benefits of breastfeeding are often dose-related, the low "doses" in this sample might explain the absence of an effect. This idea is supported by a study of the impact of breastfeeding on language development in which a threshold effect was observed at 3 months of breastfeeding duration (Dee, Li, Lee, & Grummer-Strawn, 2007). Second, breastfeeding may have a different effect in twin populations than in populations of singletons; Thorpe and colleagues (2003) found that breastfeeding had a significant effect on language development for singletons but not for their twin siblings. A possible explanation for this discrepancy is that much of the neurodevelopmental advantage associated with human milk has been tied to its long-chain fatty acids, particularly docosahexaenoic acid (DHA). Mothers of twins may produce milk lower in DHA than mothers of singletons, since they are drawing on their own fat reserves to produce milk for two babies rather than one (see McFadyen, Farquharson, & Cockburn, 2003). Whatever the reason, the present study did not show breastfeeding-related effects on any of the measures observed.

Standard Errors and Statistical Significance

Null findings, in fact, outnumber statistically significant findings in this study, though the reasons for this are unclear. It is certainly possible that a larger study might yield statistically significant results where the present study did not. The reader is invited to consider, however, that statistical significance is a relatively arbitrary construct that should be interpreted with care.

In the present study, for instance, the use of one-tailed tests would have been entirely justified based on the existing literature. Altering the significance criterion would have made these findings look more weighty on paper without altering their actual meaning. The more pressing question is what these trends might mean for families: where are these children likely to struggle? What kinds of support will they need as they progress through school? This question will be addressed further in the section on implications.

One final note on statistical significance: it is axiomatic that to achieve a statistically significant result, the standard error must be no more than half the size of the observed difference between groups. In contrast to standardized measures, which are explicitly designed to have small standard errors, conversational language measures are far more heterogeneous. This variability in children's use of language is part of what makes it a fascinating topic of study; at the same time, it increases the difficulty of labeling differences as statistically significant. Potential future studies are likely to be plagued by similar problems; regardless, it is this author's view that such research has a worthwhile role to play in a field of research where the emphasis has long been on standardized tests.

The Impact of Environmental Variables on Discourse-Level Language at School Age

As varied and unpredictable as children's language output may be, some prior research on conversational language use in the WRRP sample suggests that the factors undergirding it might be somewhat more homogeneous, with genetic effects accounting for a sizable fraction of the observed variability. When DeThorne and colleagues (2008) described their findings for language sample measures in the WRRP cohort as a whole, they reported that approximately half of the variance was heritable, with no significant shared environmental effects. These findings offered a pronounced contrast to those of Koeppen-Schomerus and colleagues (2000), who found that shared environmental variables explained 84% of the variability in language skills at age 2 for twins born before 32 weeks, while their impact was much diminished for children born closer to term. The present study was not designed to estimate heritability, an undertaking which requires a sizable *n*, but the modest split-half R^2 values observed here, in conjunction with the findings from DeThorne et al. (2008), suggest that prematurity explained much less of the variance in these children's language skills than it did in the sample assessed by Koeppen-Schomerus et al. This discrepancy might arise from changes and catch-up growth between preschool and school age. It might also be a function of differences in the gravity of the early medical challenges faced by children in the two studies, a possibility discussed further in the following section.

Limitations

This study's most significant limitation is its inability to address outcomes in the population most affected by prematurity: babies, particularly boys, born before 28 weeks' gestation and weighing <1000g. Only a handful of the children in this sample met those criteria, and thus it would be unwise to extrapolate from these results and postulate comparable language outcomes for smaller, sicker children. It is likely that these results understate the true differences between all children born prematurely and children born at full-term, since it includes no children born before 27 weeks' gestation, no children with frank neurological impairment, and furthermore includes twice as many girls, who are known to be more resilient in the face of prematurity, as it does boys. Additionally, since every child in the study was born to parents who had completed high school, it cannot speak to the potential compounding effects of lower levels of parental education.

Another possible limitation stems from this study's two inclusion criteria for the control group: (1) a gestational age ≥37 weeks, with (2) no reported perinatal complications. The absence of a birthweight selection criterion, coupled with differences in twin growth patterns as compared to singleton pregnancies, resulted in a small degree of overlap in the two groups' birthweights: 6 PRF children were born weighing between 4 and 5 pounds, as were 5 control children. Birthweight was the least useful of the continuous predictors in regression analyses confined to the PRF children as well as in the larger sample; still, it is possible that the imposition of a birthweight criterion would have resulted in a wider divergence between the two groups.

Further limitations were imposed by the design of the WRRP study from which these data are drawn. Information on perinatal outcomes was collected retrospectively, without corroboration by professionals of parent recall. It is easy to imagine that parents might misremember or misstate details such as exact gestational age or length of hospital stay, particularly in a questionnaire completed at least five years after the events in question. In all but one case, values reported by parents for gestational age were entirely consistent with the values reported for birthweight and hospital stay; even so, the possibility of errors in parental recall cannot be ruled out.

An additional concern is the discontinuity in measures utilized between HV3 and HV5, resulting from the change in WRRP assessment protocols. The switch from a 15-minute conversational task at HV2 and HV3 to the highly structured and typically briefer narrative task at HV5 obfuscates the meaning of the changes observed between HV3 and HV5. The trend toward fewer and smaller significant correlations at HV5 could reflect a maturational change for the PRF children, but this idea cannot be tested effectively without a contemporaneous conversational sample for contrast. Furthermore, the shift from assessment via IQ and digit span measures at HV2 and HV3 to assessment via standardized language instruments (the CELF and the TNL) precludes definitive statements about development across formal factor scores.

In addition to the limitations caused by selection bias and data collection protocols, a final limitation of this study arises from the nature of the language measures themselves: children's discourse-level language is complex, and an element of measurement error is thus probable. A child can toss off a ten-second utterance that takes an adult coder ten minutes or more to parse. Mistakes are inevitable in the process of coding more than 300 transcripts. Every effort has been made to ensure the reliability of the data evaluated in this study; reliability assessments for the measures, with the lone exception of the TNL, found interrater agreement above 90%. At the same time, it is likely that the large standard errors seen for the language measures reflect not only the heterogeneity familiar to any observer of children's language use, but also small inconsistencies in the coding process. Another likely contribution to the heterogeneity of the child language measures is the role of the examiner. While the WRRP examiners received training to help them elicit language samples, they varied in their interaction styles and their abilities to maintain the flow of conversation with a second-grader. Some of the variance in the children's language sample measures may arise from differences in their conversational partners (see DeThorne & Hart, 2009).

Implications

Its limitations notwithstanding, this study raises a number of intriguing questions and suggests directions for potential future research. One question concerns the best measures for future researchers to study in considering differences between premature and full-term children. As mentioned earlier, density of low-frequency vocabulary and of elaborated noun phrases seemed to distinguish between groups in conversation, as did, to a lesser extent, NDW and NTW. Low-frequency vocabulary density measures children's use of more colorful, more adult-like words in conversation and narrative, while elaborated noun phrase density captures their ability to provide compact, information-rich descriptions during the time that they are learning to compress a series of choppy sentences into a single pithy phrase (e.g., "The guy was little, the guy was a robot, and the guy is looking for a brain" can become "The little robot guy who is looking for a brain.") NDW and NTW are known to distinguish between children with typical language skills and children who struggle with language.

In contrast, these findings cast doubt on the validity and utility of measure D. Although it purports to measure vocabulary diversity while controlling for volubility, in the present study it correlated poorly with other measures. At HV3 its loading for the semantic factor was negative, a rather startling result for a supposed indicator of semantic skill. Other researchers have reported a smaller effect size for D than for other measures (Owen & Leonard, 2002; DeThorne et al., 2008); similar results were obtained in two other recent studies of the WRRP data (DeThorne, Petrill, Schatschneider, & Cutting, 2010; DeThorne, Deater-Deckard, Mahurin-Smith, & Petrill, under review), each of which used independently derived values for D. The difficulty, then, is unlikely to be faulty calculation; perhaps it is that the serial type-token ratios from which D is calculated are no more effective than aggregate TTRs in distinguishing between typical and atypical language use (Watkins, Kelly, Harbers, & Hollis, 1995). If a single TTR is not useful in discriminating between typical and impaired language abilities, it is unsurprising that an agglomeration of TTRs is no more so. Collectively these findings suggest that researchers interested in vocabulary diversity may be better off with a measure such as NDW-100, which builds in a measure of control for volubility, rather than measure D.

Future research into the impact of prematurity on discourse-level language should also consider additional moderating variables, an area in which opinions abound and hard data are scarce. It is widely assumed, for instance, that parents who read frequently to their children will have children with stronger language skills. Child effects, however, are less often considered: children who are predisposed to develop strong language skills are more likely to enjoy being read to (Scarr & McCartney, 1983; Scarborough & Dobrich, 1994; Johnson, 2006; see DeThorne & Hart, 2009 for additional background on evocative gene-environment correlation). In view of the modest split-half R^2 results, which indicate that many other variables besides prematurity and parental education serve to explain variance in discourse-level language skills, it could be very helpful to offer evidence-based information to NICU families about what those variables might be. Many possibilities are bandied about: parent-child interaction patterns, an early diet rich in long-chain essential fatty acids, early intervention programming, and exposure to literacy-promoting activities all might moderate the impact of prematurity on the developing brain. The author urges others who are conducting research in this area to bear in mind the complex interplay between children's genes and their environments. Genotype can influence, for instance, a child's ability to reap a neurodevelopmental benefit from long-chain fatty acids (Caspi, Williams, Kim-Cohen, Craig, Milne, Poulton, et al., 2007); genotype can affect scores on the widely used HOME scale and other similar measures (Plomin, Loehlin, & DeFries, 1985). Even vulnerability to neonatal sepsis may have a genetic component (Strunk & Burgner, 2006). Twin studies can offer a unique perspective on the true effects of moderating variables, all the more so given the frequency of preterm birth among twins.

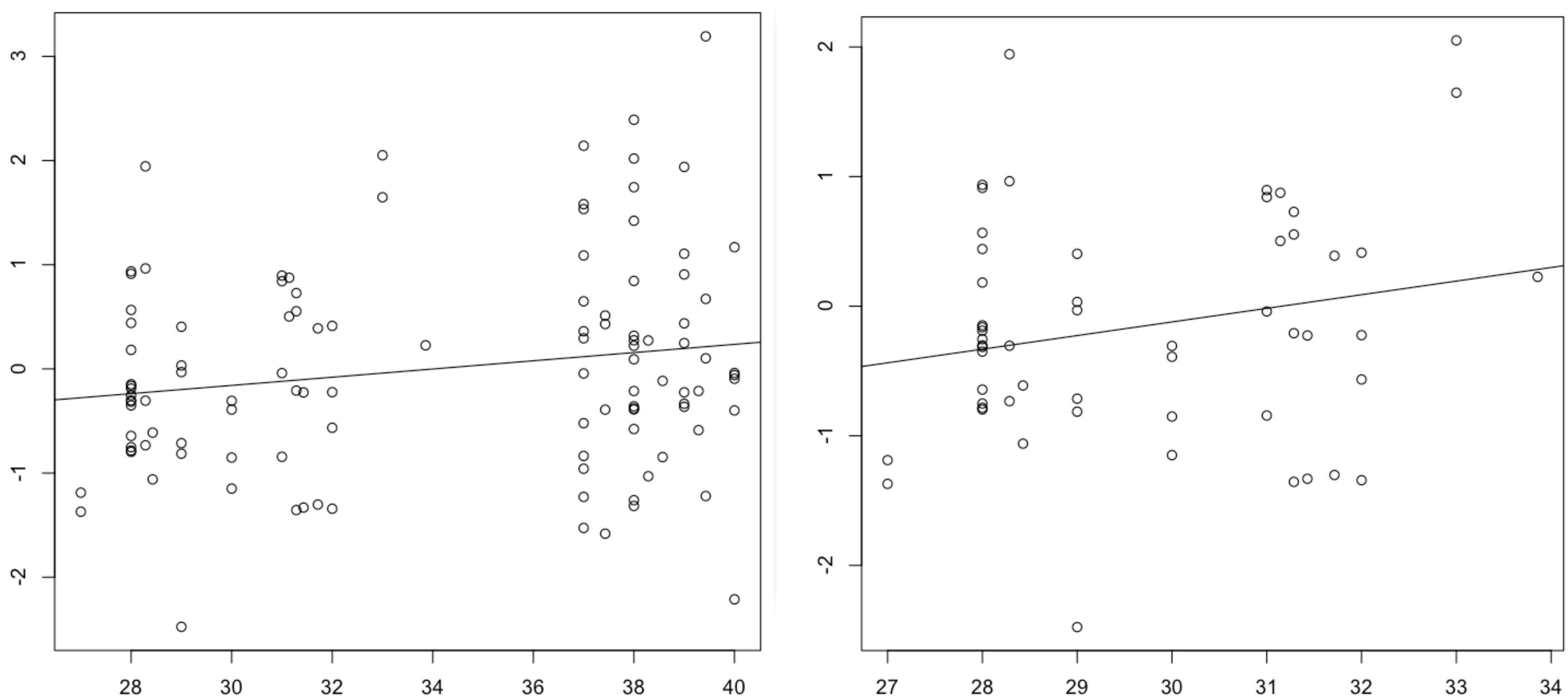
On a related note, future research should examine the question of which types of learning support are most beneficial to premature children when they reach school age as well as what forms of assessment are most meaningful. The discrepancy between PRF children's standardized test scores and their results for semantic and syntactic measures suggest that the PRF children are able to use a wide variety of literate language features in tasks where they are able to guide the flow of discourse, but that these skills break down when they are given traditional standardized tests such as the Stanford-Binet and the CELF-4.

It is probable that deficits in attention and executive function contribute to these performance profiles (Bhutta, Cleves, Casey, Cradock, & Anand, 2002; Aarnoudse-Moens, Smidts, Oosterlaan, Duivenvoorden, & Weisglas-Kuperus, 2009). Some authors attribute these difficulties to a neurodevelopmental trajectory triggered by premature birth and its resulting cascade of stressful events (Perlman, 2001); they suggest that changes in the NICU environment such as reduction of visual stimulation might promote healthier brain development in this population (Als & Butler, 2006). Whatever the mechanism, attention impairments are common among children born prematurely. They may benefit from strategies designed to compensate for attention deficits, even if they have not received a formal diagnosis of ADD. These strategies could include environmental modifications such as a quiet room or extra time for assignments and exams, sensitivity on the part of parents and teachers to their difficulties with transitions and distractions, or setting explicit goals such as ten minutes of focused effort toward an assigned task (U.S. Department of Education, 2004). To target deficits in executive function, planning and problem-solving strategies could be explicitly introduced and rehearsed. Given the large number of children born prematurely who will be entering the school system in years to come, it seems sensible to devote research to defining the areas in which they might benefit from assistance. Additional research to define deficit areas more sharply and to investigate support strategies has the potential to provide direct and substantial benefit to families.

Summary and Conclusion

After the urgent question of "Will my baby live?" has been answered affirmatively, the question asked most often by parents of premies is "Will my baby be all right?" This study corroborates existing findings that children born prematurely tend to perform more poorly on standardized tests at school age, but it also offers cause for optimism. These results suggest that their children born prematurely may struggle with aspects of discourse-level language, but that by late grade school they will be, by and large, statistically indistinguishable from their full-term peers in conversation and storytelling. The early risk factors studied here can cast a long shadow, but in the absence of extreme prematurity or overt neurological impairment they account for only a small fraction of a child's ability in the complex and dynamic domain of language learning. For the families who are anxious and uncertain, perhaps these results can offer a measure of hope.

Figure 4.1. Scatterplots and regression lines for the full sample versus the perinatal risk factor (PRF) group alone



Note: This figure shows performance on the semantic factor at home visit 2 plotted against weeks of gestational age. Full sample results are at left; the PRF group alone is shown at right.

REFERENCES

Aarnoudse-Moens, C.S.H., Smidts, D.P., Oosterlaan, J., Duivenvoorden, H.J., & Weisglas-Kuperus, N. (2009). Executive function in very preterm children at early school age. *Journal of Abnormal Child Psychology, 37*(7), 981-993.

Adams-Chapman, I., & Stoll, B.J. (2006). Neonatal infection and long-term neurodevelopmental outcome in the preterm infant. *Current Opinion in Infectious Disease, 19*(3), 290-297.

Aldenderfer, M.S., & Blashfield, R.K. (1984). *Cluster analysis.* Sage University Paper series on Quantitative Applications in the Social Sciences, 07-001. London: Sage Publications.

Als, H., & Butler, S. (2006). Neurobehavioral development of the preterm infant. In R.J. Martin, A.A. Fanaroff, & M.C. Walsh (Eds.), *Neonatal-Perinatal Medicine* (pp. 1051-1068). Philadelphia: Elsevier.

Anglin, J.M. (2000). Vocabulary development: A morphological analysis. *Monographs of the Society for Research in Child Development, 58*(10, Serial No. 238).

Aram, D.M., Hack, M., Hawkins, S., Weissman, B.M., & Borawski-Clark, E. (1991). Very-low-birthweight children and speech and language development. *Journal of Speech and Hearing Research, 34,* 1169-1179.

Baron, R.M., & Kenny, D.A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173-1182.

Barrington, K. (2001). The adverse neuro-developmental effects of postnatal steroids in the preterm infant: A systematic review of RCTs. *BioMedCentral Pediatrics, 1,* 1471.

Bayley, N. (1969). *Bayley Scales of Infant Development.* New York: Psychological Corporation.

Bayley, N. (1993). *Bayley Scales of Infant Development (*2nd ed.). San Antonio, TX: Psychological Corporation.

Beals, D.E. (1997). Sources of support for learning words in conversation: Evidence from mealtimes. *Journal of Child Language, 24,* 673-694.

Beals, D.E., & Tabors, P.O. (1995). Arboretum, bureaucratic, and carbohydrates: Preschoolers' exposure to rare vocabulary at home. *First Language, 15*(1), 57-76.

Bhutta, A.T., Cleves, M.A., Casey, P.H., Cradock, M.M., & Anand, K.J.S. (2002). Cognitive and behavioral outcomes of school-aged children who were born preterm: A meta-analysis. *Journal of the American Medical Association, 288,* 728-737.

Bishop, D.V.M. (1997). Pre- and perinatal hazards and family background in children with specific language impairments: A study of twins. *Brain and Language, 56,* 1-26.

Bishop, D.V.M., & Edmundson, A. (1987). Language-impaired 4-year-olds: Distinguishing transient from persistent impairment. *Journal of Speech and Hearing Disorders, 25,* 156-173.

Blake, J., Quartaro, G., & Onorati, S. (1993). Evaluating quantitative measures of grammatical complexity in spontaneous speech samples. *Journal of Child Language, 20*(1), 139-52.

Botvin, G.J., & Sutton-Smith, B. (1977). The development of structural complexity in children's fantasy narratives. *Developmental Psychology, 13*(4), 377-388.

Briscoe, J., Gathercole, S.E., & Marlow, N. (1998). Short-term memory and language outcomes after extreme prematurity at birth. *Journal of Speech, Language, and Hearing Research, 41,* 654-666.

Broad, F.E. & Duganzich, D.M. (1983). The effects of infant feeding, birth order, occupation, and socio-economic status on speech in six-year-old children. New Zealand Medical Journal, 96, 483-486.

Brown, G.D.A. (1984). A frequency count of 190,000 words in the London-Lund corpus of English conversation. *Behavioral Research Methods, Instruments, & Computers, 16*(6), 502-532.

Bruininks, R.H. (1978). *The Bruininks-Oseretsky Test of Motor Proficiency-Short Form.* Circle Pines, MN: American Guidance Service.

Bryan, E. (2003). The impact of multiple preterm births on the family. *BJOG: An International Journal of Obstetrics and Gynaecology, 110*(Suppl. 20), 24-28.

Bryk, A.S., & Raudenbush, S.W. (1992). *Hierarchical linear models: Applications and data analysis methods.* Newbury Park, CA: Sage Publications.

Caspi, A., Williams, B., Kim-Cohen, J., Craig, I.W., Milne, B.J., Poulton, et al. (2007). Moderation of breastfeeding effects on the IQ by genetic variation in fatty acid metabolism. *Proceedings of the National Academy of Sciences of the United States of America, 104*(47), 18860-18865.

Channell, R.W. (2006). gcSalt (Version 1.02) [Computer software]. Provo, UT: Brigham Young University.

Chapieski, M.L., & Evankovich, K.D. (1997). Behavioral effects of prematurity. *Seminars in*

Perinatology, 21(3), 221-239.

Cincotta, R.B., Gray, P.H., Phythian, G., Rogers, Y.M., Chan, F.Y. (2000). Long term outcome of twin-twin transfusion syndrome. *Archives of Disease in Childhood, Fetal and Neonatal Edition, 83,* F171-F176.

Curenton, S.M., & Justice, L.M. (2004). African American and Caucasian preschoolers' use of decontextualized language: Literate language features in oral narratives. *Language, Speech, and Hearing Services in the Schools, 35,* 240-253.

Dale, E., & Chall, J. (1948). A formula for predicting readability. *Educational Research Bulletin, 27*, 11-20; 37-54.

Daniels, M.C., & Adair, L.S. (2005). Breastfeeding influences cognitive development in Filipino children. *Journal of Nutrition, 135,* 2589-2595.

Dee D.L., Ruowei, L., Lee, L., & Grummer-Strawn, L.M. (2007). Associations between breastfeeding practices and young children's language and motor development. *Pediatrics, 119* (Suppl. 1), S92-S98.

DeFranco, E., Teramo, K., & DeMuglia, L. (2007). Genetic influences on preterm birth. *Seminars in Reproductive Medicine, 25*(1), 40-51.

Denckla, M.B., & Rudel, R.G. (1976). Rapid "automatized" naming (RAN): Dyslexia differentiated from other learning disabilities. *Neuropsychologia, 14,* 471-479.

DeThorne, L.S., Deater-Deckard, K., Mahurin-Smith, J., & Petrill. S. (Under review). Limited associations between conversational language measures and child temperament: Evidence of divergent validity.

DeThorne, L.S., & Hart, S.A. (2009). Use of the twin design to examine evocative gene-environment effects within a conversational context. *European Journal of Developmental Science, 3,* 175-194.

DeThorne, L.S., Hart, S.A., Petrill, S.A., Deater-Deckard, K., Thompson, L.A., Schatschneider, C., & Davison, M.D. (2006). Children's history of speech-language difficulties: Genetic influences and associations with reading. *Journal of Speech, Language, and Hearing Research, 49*, 1280-1293.

DeThorne, L.S., Johnson, B.W., & Loeb, J.W. (2005). A closer look at MLU: What does it really measure? *Clinical Linguistics and Phonetics, 19,* 635-548.

DeThorne, L.S., Petrill, S.A., Hart, S.A., Channell, R.W., Campbell, R.J., Deater-Deckard, K., Thompson, L.A., & Vandenbergh, D.J. (2008). Genetic effects on children's conversational language use. *Journal of Speech, Language, and Hearing Research, 51,* 423-435.

DeThorne, L.S., Petrill, S.A., Hayiou-Thomas, M.E., & Plomin, R. (2005). Low expressive vocabulary: Higher heritability as a function of more severe cases. *Journal of Speech, Language, and Hearing Research, 48,* 792-804.

DeThorne, L.S., Petrill, S.A., Schatschneider, C., & Cutting, L. (2010). Conversational language skills as a predictor of early reading development: Language history as a moderating variable. *Journal of Speech, Language, and Hearing Research, 53,* 209-223.

de Vries, L.S. (2006). Intracranial hemorrhage and vascular lesions. In R.J. Martin, A.A. Fanaroff, & M.C. Walsh (Eds.), *Neonatal-Perinatal Medicine* (pp. 924-937). Philadelphia: Elsevier.

DiSimoni, F. (1974). *The Token Test for Children.* New York: Teaching Resources Corporation.

DiStefano, C., Zhu, M., & Mîndrilă, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research, and Evaluation, 14*(20), 1-11.

Dunn, L.M., & Dunn, L.M. (1981). *Peabody Picture Vocabulary Test -- Revised.* New York: Teaching Resources Corporation.

Dunn, L.M., Dunn, L.M., Whetton, C.W., & Pintilie, D. (1982) *British Picture Vocabulary Scale.* Windsor, UK: NFER Nelson Publishing Company Ltd.

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

Eichenwald, E.C., & Stark, A.R. (2008). Management and outcomes of very low birth weight. *New England Journal of Medicine, 358,* 1700-1711.

Eisenberg, S.L., Ukrainetz, T.A., Hsu, J.R., Kaderavek, J.N., Justice, L.M., & Gillam, R.B. (2008). Noun phrase elaboration in children's spoken stories. *Language, Speech, and Hearing Services in the Schools, 39,* 145-157.

Elgen, I., Sommerfelt, K., & Markestad, T. (2002). Population based, controlled study of behavioural problems and psychiatric disorders in low birthweight children at 11 years of age. *Archives of Disease in Childhood, Fetal and Neonatal Edition, 87,* F128-F132.

Evans, D.M., & Martin, N.G. (2000). The validity of twin studies. *GeneScreen, 1,* 77-79.

Farquharson, J., Cockburn, F., Patrick, W.A. et al. (1992). Infant cerebral cortex phospholipid fatty-acid composition and diet. *Lancet, 340,* Feldman, H.M., Janosky, J.E., Scher, M.S., & Wareham, N.L. (1994). Language abilities following prematurity, periventricular brain injury, and cerebral palsy. *Journal of Communication Disorders, 27*(2), 71-90.

Fenson, L., Dale, P.S., Reznick, J.S., Bates, E., Thal, D.J., Pethick, S.J., et al. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59*(5), 1-185.

Fenson, L., Dale, P.S., Reznick, J.S., Thal, D.J., Bates, E., Hartung, J.P., et al. (1991). *MacArthur Communicative Development Inventories.* San Diego: San Diego State University.

Field, T. (2002). Maternal depression effects on infants and early interventions. *Preventive Medicine, 27*(2), 200-203.

Fletcher, S.G. (1978). *The Fletcher Time-by-Count Test of Diadochokinetic Syllable Rate.* Tigard, OR: CCC Publications.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilvel/hierarchical models.* Cambridge: Cambridge University Press.

Gillam, R.B., & Pearson, N.A. (2004). *The test of narrative language.* East Moline, IL: LinguiSystems.

Gipps, C. (1999). Socio-cultural aspects of assessment. *Review of Research in Education, 24,* 355-392.

Goldman, R., & Fristoe, M. (1986). *The Goldman-Fristoe Test of Articulation.* Circle Pines, MN: American Guidance Service.

Gray, P.H., Burns, Y.R., Mohay, H.A., O'Callaghan, M.J., & Tudehope, D.I. (1995). Neurodevelopmental outcome of preterm infants with bronchopulmonary dysplasia. *Archives of Disease in Childhood, Fetal and Neonatal Edition, 73,* F128-F134.

Greenhalgh, K.S., & Strong, C.J. (2001). Literate language features in spoken narratives of children with typical language and children with language impairments. *Language, Speech, and Hearing Services in the Schools, 32,* 114-125.

Griffiths, R. (1974). *The abilities of babies.* London: Association for Research in Child Development.

Hack, M., Taylor, H.G., Drotar, D., Schluchter, M., Cartar, L., Wilson-Costello, D., et al. (2005). Poor predictive validity of the Bayley Scales of Infant Development for cognitive function of extremely low birth weight children at school age. *Pediatrics, 116*(2), 333-341.

Holdgrafer, G. (1995). Comparison of two methods for scoring syntactic complexity. *Perceptual and Motor Skills, 81*(2), 498.

Holdgrafer, G. (1996). Syntactic abilities of neurological normal and suspect preterm children. *Perceptual and Motor Skills, 83*(2), 615-618.

Holloman, H.A., Dobbins, D.R., & Scott, K.G. (1998). The effects of biological and social risk factors on special education placement: Birth weight and maternal education as an example. *Research in Developmental Disabilities, 19*(3), 281-294.

Hughes, D., Fey, M., & Long, S. (1992). Developmental sentence scoring: Still useful after all these years. *Topics in Language Disorders, 12*(2), 1–12.

Hüppi, P.S., Warfield, S., Kikinis, R., Barnes, P.D., Zientara, G.P., Jolesz, F.A., et al. (1998). Quantitative magnetic resonance imaging of brain development in premature and mature newborns. *Annals of Neurology, 43*(2), 224-235.

Ingemarsson, I. (2003). Gender aspects of preterm birth. *BJOG: An International Journal of Obstetrics and Gynecology, 110*(Suppl. 20), 34-38.

Jennische, M., & Sedin, G. (2000). Linguistic skills at 6½ years of age in children who required neonatal intensive care in 1986–1989. *Acta Paediatrica, 90*(2), 199-212.

Johnson, A.R. (2006). *Genetic and environmental influences on children's interest in reading: A twin study*. Unpublished master's thesis, Pennsylvania State University, State College, Pennsylvania.

Johnson, S. (2007). Cognitive and behavioural outcomes following very preterm birth. *Seminars in Fetal & Neonatal Medicine, 12,* 363-373.

Kaufman, A., & Applegate, B. Short forms of K-ABC Mental Processing and Achievement Scales at age 4 to 12½ years for clinical and screening purposes. *Journal of Clinical Child and Adolescent Psychology, 17,* 359-369.

Kendler, K.S., & Eaves, L.J. (1986). Models for the joint effect of genotype and environment on liability to psychiatric illness. *American Journal of Psychiatry, 143*(3), 279-289.

Klee, T., Schaffer, M., May, S., Membrino, I., & Mougey, K. (1989). A comparison of the age–MLU relation in normal and specifically language-impaired preschool children. *Journal of Speech and Hearing Disorders*, *54*, 226–233.

Koeppen-Schomerus, G., Eley, T.C., Wolke, D., Gringras, P., & Plomin, R. (2000). The interaction of prematurity with genetic and environmental influences on cognitive development in twins. *Journal of Pediatrics, 137*(4), 527-533.

Koller, H., Lawson, K., Rose, S.A., Wallace, I., & McCarton, C. (1997). Patterns of cognitive development in very low birth weight children during the first six years of life. *Pediatrics, 99*(3), 383-389.

Kramer, M.S., Aboud, F., Mironova, E., Vanilovich, I., Platt, R.W., Matush, L., et al. (2008). Breastfeeding and child cognitive development: New evidence from a large randomized trial. *Archives of General Psychiatry, 65*(5), 578-585.

Kramer, M.S., Platt, R.W., Wen, S.W., Joseph, K.S., Allen, A., Abrahamowicz, M., et al. (2001). A new and improved population-based Canadian reference for birth weight for gestational age. *Pediatrics, 108*(2), e35.

Kucera, H., & Francis, W.N. (1967). *Frequency analysis of English usage: Lexicon and grammar.* Boston: Houghton-Mifflin.

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

Lee, L.L. (1974). *Developmental sentence analysis: A grammatical assessment procedure for speech and language clinicians.* Evanston, IL: Northwestern University Press.

Leonard, L.B. (1998). *Children with specific language impairment.* Cambridge, MA: MIT Press.

Lewis, B.A., Singer, L.T., Fulton, S., Salvator, A., Short, E.J., Klein, N., et al. (2002). Speech and language outcomes of children with bronchopulmonary dysplasia. *Journal of Communication Disorders, 35,* 393-406.

Liles, B.Z. (1993). Narrative Discourse in Children with Language Disorders and Children with Normal Language: a Critical Review of the Literature JSLHR 38 868-882

Lilienfeld, A.M., Pasamanick, B., & Rogers, M. (1955). Relationship between pregnancy experience and development of certain neuropsychiatric disorders in childhood. *American Journal of Public Health, 45,* 637-643.

Long, S.H. (2008). Computerized Profiling (Version 9.7) [Computer software]. Arcata, CA: Author.

Long, S.H., & Channell, R.W. (2001). Accuracy of four language analysis procedures performed automatically. *American Journal of Speech-Language Pathology, 10,* 180-188.

Longnecker, M.P., Zhou, H., & Brock, J.W. (2001.) Association between maternal serum concentration of the DDT metabolite DDE and preterm and small-for-gestational-age babies at birth. *The Lancet, 358,* 110-114.

Lucas, A., Morley, R., Cole, T.J., Lister, G., & Leeson-Payne, C. (1992). Breast milk and subsequent intelligence quotient in children born preterm. *Lancet, 339,* 261-264.

Luke, B., & Keith, L.G. (1992). The contribution of singletons, twins, and triplets to low birth weight, infant mortality, and handicap in the United States. *Journal of Reproductive Medicine, 37,* 661-666.

Lumley, J. (2003). Defining the problem: The epidemiology of preterm birth. *BJOG: An International Journal of Obstetrics and Gynecology, 110*(Suppl. 20), 3-7.

Luoma, L., Herrgård, E., Martikainen, A., & Ahonen, T. (1998). Speech and language development of children born at ≤32 weeks' gestation: A 5-year prospective follow-up study. *Developmental Medicine and Child Neurology, 40,* 380-387.

MacWhinney, B. (2000). The CHILDES project: Tools for analyzing talk (3rd edition). Mahwah, NJ: Lawrence Erlbaum Associates.

MacWhinney, B., & Spektor, L. (2009). Computerized Language Analysis (Version 7.28.09) [Computer software]. Pittsburgh, PA: Author.

Mahurin Smith, J., & Ambrose, N.G. (2009, November). Stuttering persistence and infant diet: Breastfeeding may protect against persistent stuttering. Poster session presented at the American Speech-Language-Hearing Association Convention, Chicago, IL.

Mahurin Smith, J., & DeThorne, L.S. (2008, June). Can a computer recognize a good story? Predicting outcomes on the Test of Narrative Language. Poster session presented at the Symposium on Research in Child Language Disorders, Madison, WI.

Marinellie, S.A., & Chan, Y.L. (2006). The effect of word frequency on noun and verb definitions: A developmental study. *Journal of Speech, Language, and Hearing Research, 49,* 1001-1021.

Marlow, N. (2004). Neurocognitive outcome after very preterm birth. *Archives of Disease in Childhood, Fetal and Neonatal Edition, 89,* F224-F228.

Mather, P. L. and K. N. Black (1984). Hereditary and environmental influences on preschool twins' language skills. *Developmental Psychology, 20,* 303-308.

McCarthy, D. (1972). *McCarthy Scales of Children's Abilities.* New York: Psychological Corporation.

McFadyen, M., Farquharson, J., & Cockburn, F. (2003). Maternal and umbilical cord erythocyte omega-3 and omega-6 fatty acids and haemorrheology in singleton and twin pregnancies.

Archives of Disease in Childhood, Fetal and Neonatal Edition, 88(2), F134-F138.

Ment, L.R., Vohr, B., Allan, W., Katz, K.H., Schneider, K.C., Westerveld, M., et al. (2003). Change in cognitive function over time in very low-birth-weight infants. *Journal of the American Medical Association, 289*(6), 705-711.

Mentis, M. (1994). Topic management in discourse: Assessment and intervention. *Topics in Language Disorders, 14,* 29-54.

Miller, J. (2004). Systematic Analysis of Language Transcripts (Research Version 8.0) [Computer software]. Madison, WI: University of Wisconsin.

Moutquin, J.-M. (2003). Classification and heterogeneity of preterm birth. *BJOG: An International Journal of Obstetrics and Gynecology, 110*(Suppl. 20), 30-33.

Msall, M.E., Bier, J.-A., LaGasse, L., Tremont, M., & Lester, B. (1998). The vulnerable preschool child: The impact of biomedical and social risks on neurodevelopmental function. *Seminars in Pediatric Neurology, 5*(1), 52-61.

Msall, M.E., Phelps, D.L., DiGaudio, K.M., Dobson, V., Tung, B., McClead, R.E., et al. (2000). Severity of neonatal retinopathy of prematurity is predictive of neurodevelopmental functional outcome at age 5.5 years. *Pediatrics, 106*(5), 998-1005.

Mullineaux, P.Y., Deater-Deckard, K., Brown, S., Petrill, S.A., Thompson, L.A., DeThorne, L.S., & Schatschneider, C. (in press). Observed parenting and child behavioral problems: A longitudinal analysis of nonshared environment.

Nadeau, L., Boivin, M., Tessier, R., Lefebvre, F., & Robaey, P. (2001). Mediators of behavioral problems in 7-year-old children born after 24 to 28 weeks of gestation. *Developmental and Behavioral Pediatrics, 22*(1), 1-10.

Nagy, W.E., Herman, P.A., & Anderson, R.C. (1985). Learning words from context. *Reading Research Quarterly, 22,* 233-253.

Newcomer, P., & Hammill, D. (1988). *Test of Language Development-2: Primary.* Austin, TX: Pro-Ed.

Nippold, M.A. (1998). *Later language development: The school-age and adolescent years.* Austin, TX: Pro-Ed.

Nippold, M.A. (2007). *Later language development: The school-age and adolescent years* (3rd ed.). Austin, TX: Pro-Ed.

Nippold. M.A., & Schwarz, I.E. (1996). Children with slow expressive language development:

What is the forecast for school achievement? *American Journal of Speech-Language Pathology, 5,* 29-30.

Nippold, M.A., Schwarz, I.E., & Undlin, R.A. (1992). Use and understanding of adverbial conjuncts: A developmental study of adolescents and young adults. *Journal of Speech and Hearing Research, 35,* 108-118. 303-308.

Nippold, M.A., & Sun, L. (2008). Knowledge of morphologically complex words: A developmental study of older children and young adolescents. *Language, Speech, and Hearing Services in the Schools, 39,* 365-373.

Oddy, W.H. (2006). Fatty Acid Nutrition, Immune and Mental Health Development from Infancy through Childhood. In J.D. Huang, Ed., Frontiers in Nutrition Research, (pp. 177-211). New York: Nova Science.

Ogden, C.L., Kuczmarski, R.J., Flegal, K.M., Mei, Z., Guo, S., Wei, R., et al. (2002). Centers for Disease Control and Prevention 2000 growth charts for the United States: Improvements to the 1977 National Center for Health Statistics version. *Pediatrics, 109*(1), 45-60.

Owen, A.J., Leonard, L.B. (2002). Lexical diversity in the spontaneous speech of children with specific language impairment: Application of D. *Journal of Speech, Language, and Hearing Research, 45,* 927-937.

Patra, K., Wilson-Costello, D., Taylor, H., Mercuri-Minich, N., & Hack, M. (2006). Grades I-II intraventricular hemorrhage in extremely low birth weight infants: Effects on neurodevelopment. *The Journal of Pediatrics, 149*(2), 169-173.

Paul, R. (1996). Clinical implications of the natural history of slow expressive language development. *American Journal of Speech-Language Pathology, 5,* 5-21.

Pendergast, K., Dickey, S., Selmar, J., & Soder, A. (1984). *Photo Articulation Test.* Danville, IL: Interstate.

Peterson, B.S. (2003). Brain imaging studies of the anatomical and functional consequences of preterm birth for human brain development. *Annals of the New York Academy of Sciences, 1008,* 219-237.

Perlman, J.M. (2001). Neurobehavioral deficits in premature graduates of intensive care: Potential medical and neonatal environmental risk factors. *Pediatrics, 108*(6), 1339-1348.

Petrill, S.A., Deater-Deckard, K., Thompson, L.A., DeThorne, L.S., & Schatschneider, C. (2006). Genetic and environmental effects of serial naming and phonological awareness on early reading outcomes. *Journal of Educational Psychology, 98*(1), 112-121.

Petrill, S.A., Deater-Deckard, K., Thompson, L.A., DeThorne, L.S., & Schatschneider, C. (2006). Reading skills in early readers: Genetic and shared environmental influences. *Journal of Learning Disabilities, 39*(1), 48-55.

Phelps, D.L. (2006). Retinopathy of prematurity. In R.J. Martin, A.A. Fanaroff, & M.C. Walsh (Eds.), *Neonatal-Perinatal Medicine* (pp. 1747-1754). Philadelphia: Elsevier.

Plomin, R., & Bergeman, C.S. (1991). The nature of nurture: Genetic influence on "environmental" measures. *Behavioral and Brain Sciences, 14,* 373-427.

Plomin, R., Loehlin, J.C., & DeFries, J.C. (1985). Genetic and environmental components of “environmental” influences. *Developmental Psychology, 21*(3), 391-402.

Prada, J.A., & Tsang, R.C. (1998). Biological mechanisms of environmentally induced causes of IUGR. *European Journal of Clinical Nutrition, 52,* S21-S27.

R Development Core Team. (2009). R: A Language and Environment for Statistical Computing (Version 2.10) [Computer software]. Vienna: R foundation for Statistical Computing.

Raven, J. (1977). *Raven's Coloured Progressive Matrices.* London: H.K. Lewis & Co. Ltd.

Reijneveld, S.A., de Kleine, M.J.K., van Baar, A.L., Kollee, L.A.A., Verhaak, C.M., Verhulst, F.C., et al. (2006). Behavioural and emotional problems in very preterm and very low birthweight infants at age 5 years. *Archives of Disease in Childhood, Fetal and Neonatal Edition, 91,* F423-F428.

Renfrew, C.E. (1985). *The Bus Story: A Test of Continuous Speech* (2nd ed.). Oxford: Author.

Rice, M. (2004). Growth models of developmental language disorders. In M. L. Rice & S. F. Warren (Eds.), *Developmental language disorders: From phenotypes to etiologies* (pp. 207–240). Mahwah, NJ: Erlbaum.

Rice, M. L., Redmond, S. M., & Hoffman, L. (2006). Mean length of utterance in children with specific language impairment and in younger control children shows concurrent validity and stable and parallel growth trajectories. *Journal of Speech, Language, and Hearing Research*, *49*, 793–808.

Riordan, J. (2005). The biological specificity of breastmilk. In J. Riordan (Ed.), *Breastfeeding and Human Lactation* (3rd ed., pp. 97-136). Sudbury, MA: Jones and Bartlett.

Rodgers, J.L., Rowe, D.C., & Li, C. (1994). Beyond nature versus nurture: DF analysis of nonshared influences on problem behaviors. *Developmental Psychology, 30*(3), 374-384.

Roth, F.P., & Spekman, N.J. (1986). Narrative discourse: Spontaneously generated stories of learning-disabled and normally achieving students. *Journal of Speech and Hearing Disorders, 51,* 8-23.

Rowe, D.C., Jacobson, K.C., & Van den Oord, E.J.C.G. (1999). Genetic and environmental influences on vocabulary IQ: Parental education level as moderator. *Child Development, 70*(5), 1151-1162.

Rutter, M., Thorpe, K., Greenwood, R., Northstone, K., & Golding, J. (2003). Twins as a natural experiment to study the causes of mild language delay I: Design; twin-singleton differences in language, and obstetric risks. *Journal of Child Psychology and Psychiatry, 44*(3), 326-341.

Saudino, K.J., Dale, P.S., Oliver, B., Petrill, S.A., Richardson, V., Rutter, M., et al. (1998). The validity of parent-based assessment of the cognitive abilities of two-year-old. *British Journal of Developmental Psychology, 16,* 349-363.

Savitz, D.A., Whelan, E.A., & Kleckner, R.C. (1989). Effects of parents' occupational exposures on risk of stillbirth, preterm delivery, and small-for-gestational-age infants. *American Journal of epidemiology, 129*(6), 1201-1218.

Scarborough, H.S., & Dobrich, W. (1994). On the efficacy of reading to preschoolers. *Developmental Review, 14*(3), 245-302.

Scarr, S. (1992). Developmental theories for the 1990s: Development and individual differences. *Child Development, 63,* 1-19.

Scarr, S., & McCartney, K. (1983). How people make their own environments: A theory of genotype -> environment effects. *Child Development, 54,* 424-435.

Schwartz, R.M., Luby, A.M., Scanlon, J.W., & Kellogg, R.J. (1994). Effect of surfactant on morbidity, mortality, and resource use in newborn infants weighing 500 to 1500 g. *New England Journal of Medicine, 330,* 1476-1480.

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

Scott, C.M., & Windsor, J. (2000). General language performance measures in spoken and written narrative and expository discourse of school-age children with language learning disabilities. *Journal of Speech, Language, and Hearing Research, 43,* 324-339.

Semel, E., Wiig, E., & Secord, W. (1995). *Clinical Evaluation of Language Fundamentals* (3rd ed.). New York: Psychological Corporation.

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical Evaluation of Language Fundamentals*

(4th ed.). Toronto, Canada: The Psychological Corporation/A Harcourt Assessment Company.

Semel-Mintz, E., & Wiig, E.H. (1982). *Clinical Evaluation of Language Functions.* Columbus, OH: C.E. Merrill.

Senat, M.-V., Ancel, P.-Y., Bouvier-Colle, M.-H., & Bréart, G. (1998). How does multiple pregnancy affect maternal mortality and morbidity? *Clinical Obstetrics and Gynecology, 41*(1), 79-83.

Shah, D.M. (2006). Hypertensive disorders of pregnancy. In R.J. Martin, A.A. Fanaroff, & M.C. Walsh (Eds.), *Neonatal-Perinatal Medicine* (pp. 307-320). Philadelphia: Elsevier.

Shevell, T., Malone, F.D., Vidaver, J., Porter, T.F., Luthy, D.A, Comstock, C.H., et al. (2005). Assisted reproductive technology and pregnancy outcome. *Obstetrics and Gynecology, 106,* 1039-1045.

Short, E.J., Klein, N.K., Lewis, B.A., Fulton, S., Eisengart, S., Kercsmar, C., et al. (2003). Cognitive and academic consequences of bronchopulmonary dysplasia and very low birth weight: 8-year-old outcomes. *Pediatrics, 112*(5), e359-e366.

Singer, L.T., Siegel, A.C., Lewis, B., Hawkins, S., Yamashita, T., & Baley, J. (2001). Preschool language outcomes of children with history of bronchopulmonary dysplasia and very low birth weight. *Developmental and Behavioral Pediatrics, 22*(1), 19-26.

Sonntag, J., Grimmer, I., Scholz, T., Metze, B., Wit, J., & Obladen, M. (2000). Growth and neurodevelopmental outcome of very low birthweight children with necrotizing enterocolitis. *Acta Paediatrica, 89,* 528-532.

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

Speltz, M.L., DeKlyen, M., Calderon, R., Greenberg, M.T., & Fisher, P.A. (1999) Neuropsychological characteristics and test behaviors of boys with early onset conduct problems. *Journal of Abnormal Psychology, 108*(2), 315-325.

Stevenson, D.K., Verter, J., Fanaroff, A.A., Oh, W., Ehrenkranz, R.A., Shankaran, S., et al. (2000). Sex differences in outcomes of very low birthweight infants: the newborn male disadvantage. *Archives of Disease in Childhood, Fetal and Neonatal Edition, 83,* F182-F185.

Strömberg, B., Dahlquist, G., Ericson, A., Finnström, O., Köster, M., & Stjernqvist, K. (2002). Neurological sequelae in children born after in-vitro fertilisation: A population-based study.

The Lancet, 359, 461-465.

Stromswold, K. (2006). Why aren't identical twins linguistically identical? Genetic, prenatal, and postnatal factors. *Cognition, 101,* 333-384.

Strunk, T., & Burgner, D. Genetic susceptibility to neonatal infection. *Current Opinion in Infectious Disease, 19*(3), 259-263.

Taveras, E.M., Capra, A.M., Braveman, P.A., Jensvold, N.G., Escobar, G.J., & Lieu, T.A. (2003). Clinician support and psychosocial risk factors associated with breastfeeding discontinuation. *Pediatrics, 112*(1), 108-115.

Thompson, D.K., Warfield, S.K., Carlin, J.B., Pavlovic, M., Wang, H.X., Bear, M., et al. (2007). Perinatal risk factors altering regional brain structure in the preterm infant. *Brain, 130*(3), 667-677.

Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). Stanford-Binet Intelligence Scale (4th ed.). Itasca, IL: Riverside Publishing.

Thorpe, K., Rutter, M., & Greenwood, R. (2003). Twins as a natural experiment to study the causes of mild language delay II: Family interaction risk factors. *Journal of Child Psychology and Psychiatry, 44*(3), 342-355.

Tomblin, J.B., Smith, E., & Zhang, X. (1997). Epidemiology of specific language impairment: Prenatal and perinatal risk factors. *Journal of Communication Disorders, 30,* 325-344.

Turkheimer, E., Haley, A., Waldron, M., D'Onofrio, B., & Gottesman, I.I. (2003). Socioeconomic status modifies heritability of IQ in young children. *Psychological Science, 14*(6), 623-628.

Tyson, J.E., Parikh, N.A., Langer, J., Green, C., & Higgins, R.D. (2008). Intensive care for extreme prematurity: Moving beyond gestational age. *New England Journal of Medicine, 358*(16), 1672-1681.

Ukrainetz, T.A., & Gillam, R.B. (2009). The expressive elaboration of imaginative narratives by children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 52,* 883-898.

United States Census Bureau. (2004). *Educational Attainment in the United States: 2003.* Washington, D.C.: U.S. Census Bureau.

United States Department of Education. (2004). *Teaching Children with Attention Deficit Hyperactivity Disorder: Instructional Strategies and Practices.* Washington, D.C.: U.S. Department of Education.

Updegrove, K. (2004). Necrotizing enterocolitis: The evidence for use of human milk in prevention and treatment. *Journal of Human Lactation, 20*(3), 335-339.

Vitali, G.J. (1986). *Test of Oral Structures and Functions.* East Aurora, NY: Slossan Educational Publications.

Vohr, B.R. (2007). How should we report early childhood outcomes of very low birth weight infants? *Seminars in Fetal & Neonatal Medicine, 12,* 355-362.

Ward, R.M., & Beechy, J.C. (2003). Neonatal complications following preterm birth. *BJOG: An International Journal of Obstetrics and Gynecology, 110*(Suppl. 20), 8-16.

Watkins, R.V., Kelly, D.J., Harbers, H.M., & Hollis, W. (1995). Measuring children's lexical diversity: Differentiating typical and impaired language learners. *Journal of Speech, Language, and Hearing Research, 38,* 1349-1355.

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

Wechsler, D. (1974). *Wechsler Intelligence Scale for Children -- Revised.* New York: Psychological Corporation.

Wechsler, D. (1989). *Wechsler Preschool and Primary Scale of Intelligence -- Revised.* New York: Psychological Corporation.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children* (3rd ed.). San Antonio: Psychological Corporation.

Weizman, Z.O., & Snow, C.E. (2001). Lexical input as related to children's vocabulary acquisition: Effects of sophisticated exposure and support for meaning. *Developmental Psychology, 37*(2), 265-279.

Wilson-Costello, D.E., & Hack, M. (2006). Follow-up for high-risk neonates. In R.J. Martin, A.A. Fanaroff, & M.C. Walsh (Eds.), *Neonatal-Perinatal Medicine* (pp. 1035-1044). Philadelphia: Elsevier.

Zahka, K.G., & Erenberg, F. (2006). The cardiovascular system: Congenital defects. In R.J. Martin, A.A. Fanaroff, & M.C. Walsh (Eds.), *Neonatal-Perinatal Medicine* (pp. 1222-1241). Philadelphia: Elsevier.

Zimmerman, I.L., Steiner, V.G., & Pond, R.E. (1992). *Preschool Language Scales - 3.* San Antonio: Psychological Corporation.

Zubrick, S.R., Taylor, C.L., Rice, M.L., & Slegers, D.W. (2007). Late language emergence at 24 months: An epidemiological study of prevalence, predictors, and covariates. *Journal of Speech, Language, and Hearing Research, 50,* 1562-1592.

Appendix A. Target affixes for MOR analysis (MacWhinney, 2000)

<u>Prefixes</u>

anti-
co-
de-
dis-
mega-
mini-
mis-
multi-
non-
out-
over-
pre-
re-
semi-
super-
un-
under-
up-

<u>Suffixes</u>

-able
-al
-er (both agentive, as in teacher, and comparative, as in uglier)
-est
-ful
-ie (diminutive)
-ier
-iest
-ing
-ish
-less
-lier
-liest
-like
-ly
-ness
-wise
-y

Appendix B. Metalinguistic/metacognitive verbs (Greenhalgh & Strong, 2001)

ask
believe
call
decide
forget
forgot
greet
knew
know
promise
said
say
scold
shout
tell
think
thought
told
wish
wished
yell

Appendix C. Commands used to obtain data from CLAN.

To derive, Measure D, the following command string was used:

```
> vocd +t*CHI +s*^%% +r6 +f *.cex
```

The +t*CHI use flag indicates that only the child tier is to be considered. The second use flag, +s*^%%, tells CLAN to collapse across morphemic variants, so that "tiger" and "tigers" are counted as two instances of the same word and not two distinct words. To eliminate mazes, the +r6 use flag was included. Finally, the +f flag instructs CLAN to send output for each transcript to a file.

In contrast to the D results obtained for DeThorne, Deater-Deckard, Mahurin-Smith, and Petrill (in press), these results were obtained over a child's entire transcript, including abandoned, interrupted, and partially unintelligible utterances. When these values were compared to values obtained on a narrower analysis set, they are marginally different; on average, these values were 1% higher.

Appendix D. Guidelines for coding elaborated noun phrases (ENPs), with examples.

Based on Greenhalgh and Strong (2001).

Phrases were marked if they met one or more of the following criteria:

1. They included three or more modifiers before the noun (e.g., "the very hungry Rottweiler"), or two or more postnominal modifiers, (e.g., "a dragon, brave and fierce"). Examples from the samples include phrases such as “a large vanilla milkshake, ” “a nice pink one,” “a real caterpillar's head,” “little bitty teensy tinsy pieces,” and “a whole alien family.”

2. They were followed by a relative clause or a clause in which a relative pronoun was implicit. Some examples used by the children in the study: “a little robot guy who is looking for a brain,” “a little boy who has a mom,” “the guy who has the treasure from Treasure Planet” “little skaters that look like us,” “one that normally comes,” “a slide [that] you can have at home,” “a nice thing [that] you’re doing there.” In the two final examples, the relative pronoun was implied rather than explicit; such phrases were included in the tally.

3. They had clearly associated infinitival, participial or prepositional phrases. To be counted, a phrase had to be clearly tied to the noun and not merely positioned next to it in the sentence. For example, in the phrase "a place to live" there is an implied relative clause: "a place [in which] to live." If a phrase could be rearranged without altering the meaning of the sentence, it was not counted. To clarify, in the sentence "there was a cat in the house," the prepositional phrase could be repositioned without difficulty (i.e., "In the house there was a cat") and thus "a cat in the house" would not be considered an elaborated noun phrase. In the sentence "I saw that girl with the brown shoes," the prepositional phrase provides important information about which girl the speaker is referring to. Consequently, "that girl with the brown shoes" would be coded as an ENP. The following are actual examples from the transcripts: “all the colors of Play-Doh,” “a colorful ball of that stuff,” “a bunch of pancakes with colors,” “a teensy bit of purple,” “all the rest of them”

Appendix E: Low-frequency words

This appendix includes the low-frequency words analyzed in the study. The list that follows represents a subset of the larger low-frequency list derived for the entire WRRP corpus. It is an alphabetical list of the words used at all three of the home visits by the 114 children under consideration, recorded along with their frequencies in the 301 transcripts reviewed here. Multiple instances of the same low-freqeuncy word could have come from the same child, within or across transcripts, or from different children. Each occurrence of each word in the following list was assessed in context to confirm that it was used with at least a measure of semantic appropriateness, and was not a typographical error, a transcription artifact (as when a transcriber labeled a child's string of jargon as "nonsense," but typed the word in parentheses rather than brackets), or an idiosyncratic use inconsistent with an emerging understanding of the word. For a word such as "sprite," its inclusion here indicates that it was used as a common noun, and did not refer to the trademarked beverage name; proper nouns were excluded although capitalized adjectives (e.g., "Amish") were not. Words that include underscores were treated as frozen forms during transcribing and checking in accordance with SALT protocols. In determining low-frequency status, morphemic variants were combined for the early-developing grammatical morphemes (progressive -ing, third-person singular -s, past tense -ed, plural -s, and possessive -s) but not for later-developing derivational morphemes such as comparative -er and superlative -est. Consequently, all inflectional morphemic variants listed here (e.g., "explode," "explodes," "exploded," and "exploding,") did not reach a frequency count of 15 when combined. In contrast, derivational morphemes such as "smarter" and "smartest" would not have been combined when determining low-frequency status.

1 aardvark
1 abduct
1 abducted
2 aboard
1 abridged
1 absolutely
1 accelerated
1 accept
1 accessories
1 account
1 accuses
1 action
1 actions
3 actor
1 actual
1 addition
1 admit
1 admits
6 adopted
1 ads
1 adventure
2 adventures
1 afford
1 afro
5 afterwards
2 agent
1 aim
1 airborne
1 air-conditioners
3 air-conditioning
1 allergies
1 allosaurous
1 alligators
1 all-star
1 Amish
3 anaconda
2 anacondas
2 ancestor
1 ancestors
2 ancient
1 angers
1 angolosaurus
1 anniversary
1 antennas
2 antlers
2 apartment
1 ape
1 apologized
3 appeared
3 appointment
1 aqua
1 aquariums
1 arch-enemy
1 arch-enemies
2 areas
2 argue
2 armor
3 armrests
2 arrived
2 arriving
1 arrow
1 aside
2 assembly
1 assign
1 assigned
1 assistant
1 associates
3 attendance
3 attic
1 attitude
1 attractions
2 authorities
3 autographs
1 automatically
2 axe
1 babyish
1 backcountry
4 backgrounds
1 back_handsprings
1 backpacks
1 backstage
3 backstroke
2 bacon
1 badge
2 badges
1 badly
2 bait
1 baker
1 balance_beam
1 bald
1 bale
1 bamboo

1 banjo
1 bank
1 banker
1 bank_robbery
1 barbecue
1 bare
1 based
4 bass
1 batteries
1 battleship
1 batons
2 beach_house
1 beagles
1 beams
3 beaver
1 beef
1 beg
2 begin
1 begins
1 begging
1 behavior
1 bells
1 bellyache
1 belong
1 belongs
2 bendable
3 beside
1 beta
1 betafish
1 bin
1 birdcalls
1 birdseed
2 blackish
3 blackout
2 blacktop
2 black_widow
1 blank
1 blankets
1 blanks
2 blasted
1 bleach
1 bleachers
1 blended
1 blimp
1 blind
1 blinds
1 blizzard
5 blonde
2 bloody
3 blow-up
1 bluebird
1 bluish
1 blushing
1 boa_constrictor
1 boa_constrictors
1 boarded
1 boarding_school
3 bobsleigh
2 boil
1 bologna
1 bolt
1 bonus
2 booklet
1 booklets
1 bookmobile
2 boombox
1 boomerang
1 boulder
2 braid
1 brains
4 brand
1 breadsticks
1 breaststroke
2 breathe
1 breathing
1 breeze
1 brick
1 brighten
3 brighter
1 broccoli
4 brontosaurus
1 bronze
1 brownish
1 bruise
4 buck
2 bucked
1 budge
1 buggies
1 builder
2 bull
1 bulldog
1 bullfrog
3 bully
2 bumpy
2 bun
1 bungee-jumping
1 bunk
3 bunkbed
1 burglar
5 bus_stop
2 busted
1 butler
2 cable
1 cactus
1 calf
1 calms
2 camel
1 camouflaged
3 campfire
1 campground
6 canceled
2 cancer
1 cane
3 cannon
1 cannons
1 cans
2 cantaloupe
1 canteen
5 canter
1 canters
1 cantering
1 cape
1 capitalized
1 capitals
1 capri
1 capris
1 caps
1 captain
1 capture
2 captured
1 caramel
1 cardiologist
5 carnival
1 carousel
1 carriage
2 cartwheel
4 cartwheels
1 carved
3 cash
1 cashews
5 cast
2 categories
4 catfish
1 CD-ROMs
1 cell
1 cells
1 century
1 centuries
2 chairlift
1 chalkdust
4 chameleons
1 champ
5 chance
4 charge
1 cheap
1 cheeks
1 cheers
1 chess
1 chickenpox
1 chihuahua
1 child_support
2 chinning
1 choked
1 choking
2 chopsticks
2 chubby
1 cicada
1 cicadas
1 classic
1 classmate
1 classmates
3 claw
2 claws
2 cleaner
5 clear
1 clicks
1 climate
1 clipboard
1 cloaking
1 closely
1 closeups
1 clothing
1 cloudy
1 clownfish
2 clubhouse
1 cockapoo
1 cockatoo
1 cocoa
4 coconut
1 cocoon
4 code
1 coffin
1 coin
2 coins
3 colder
1 collarbones
1 colliding
3 collies
1 colorblind
3 combined
1 comfortable
1 comic
1 commence
1 common
1 communion
2 community
1 compared
1 compass
3 competition
1 competitions
2 complaints
3 complete

1 completed
1 completely
2 complimented
1 comprehension
1 comrade
1 concassa
1 concrete
1 condo
1 congratulations
1 connection
1 contained
1 continued
2 contractions
2 controller
1 controllers
1 convince
1 convinced
1 cooler
2 cord
2 correct
1 coughing
1 council
1 counselors
1 county
1 courage
2 courses
1 court
1 courts
1 cranky
1 craters
1 crawdaddy
1 creaky
1 creation
1 creations
3 creative
1 creep
3 crickets
1 crime
1 crossbow
3 crow
2 crowded
1 crowing
1 crumble
1 crumpled
6 crush
1 crushed
2 crust
1 crusty
1 crystal
2 crystals
7 cube
1 cuckoo
4 cuckoo_clock
5 cucumber
2 cucumbers
1 cuddle
1 cuddled
1 cuddly
1 cupboard
3 cupcake
2 cupcakes
1 cure
2 curious
1 curtain
1 curvy
1 cushion
1 cussing
1 customer
2 cuttlebone
3 cyclops
1 dachshund
1 daddylongleg
1 daddylonglegs
1 dancer
1 dancers
2 dandelions
1 danger
2 dare
1 daredevil
2 darker
1 darling
1 dart
2 data
2 dear
2 death
1 debating
1 debt
1 defender
3 degrees
2 deleted
2 delivering
1 demerit
1 den
1 dented
1 department
1 described
3 desert
1 designers
3 dessert
1 details
1 device
1 diamonds
1 diaries
1 dice
3 difficult
1 digesting
1 digit
1 dinging
2 direction
1 directions
1 dirtier
1 disaster
2 disco
1 discussed
1 disease
5 disgusting
1 disk
3 disks
1 distinguish
3 disturb
1 divers
1 divorce
1 dizzy
1 Doberman_ pinschers
2 dodge
2 dollhouse
1 dominoes
1 donate
1 doomed
1 doorway
1 dotcom
1 doubling
2 downhill
2 drain
3 dresser
4 dress-up
1 dribbling
1 drill
1 drizzle
1 droop
1 drop-off
1 drowns
1 drunk
2 duckbill
3 duckling
1 dumbest
4 dungeon
1 dungeons
1 dunk
3 dust
1 dwarf
1 dye
3 dyed
1 eager
1 eagle
1 earmuffs
1 earthlings
5 earthquake
3 easiest
1 eater
1 echo
1 echoed
1 effect
1 elbow
1 electrical
6 electricity
1 element
3 elements
5 elf
1 elk
5 embarrassed
1 embarrassing
6 emergency
1 emergencies
2 emperor
2 employee
1 encountered
1 encountering
1 endless
5 enemy
4 energy
3 engine
1 enjoyed
1 enjoying
1 enter
2 equals
1 equator
4 equipment
1 estimates
1 event
3 exact
1 example
1 excellent
1 except
1 exchange
2 exhibits
1 exoskeleton
1 expect
1 expecting
1 expensive
2 experience
1 experiment
2 expert
1 expires
1 exploded
1 exploding
1 explosions
1 extinct
1 extract
2 eyebrow
1 eyebrows
1 fabulous

1 factors
2 facts
1 faded
2 failure
1 faint
3 fainted
1 fame
1 familiar
3 famous
1 fang
2 fangs
3 farmer
1 fashion
1 fastened
1 fastest
2 fatter
1 fattest
1 fattish
2 feast
1 feathery
1 feelings
2 female
1 females
1 ferocious
1 ferret
1 festival
1 fiery
1 figured
1 file
2 film
1 filming
1 final
2 fingerprints
1 fireball
2 fireballs
1 firebombs
2 firecrackers
1 fired
1 firefighter
1 firehouse
1 fireplace
1 fire_station
2 fisherman
2 fish_food
6 flame
1 flamethrowers
1 flamingos
1 flashbacks
1 flashcards
2 flashlight
1 flashlights
1 flashlight_tag
1 flatter
1 flea
1 flick
1 flings
2 flippers
3 flood
1 flooded
1 floods
1 floors
1 flour
1 flowerbed
1 flower_girls
1 flowerpot
1 flush
3 flushed
2 flushes
1 flute
3 foam
1 foamy
3 foggy
1 foil
2 folks
2 footsteps
1 footstool
2 force_field
1 foreign
1 forked
1 formula
6 fortune
1 fortune_teller
1 fortune_tellers
1 fountain
2 four-legged
1 foursquare
2 fox
4 frame
2 frames
1 frazzled
2 freaky
1 freedom
1 frees
4 freestyle
1 freezer
1 freezetag
1 friendlier
5 frightened
1 frontwards
1 frosting
1 frowning
2 frustrating
1 fuel
1 fully
1 furniture
2 furry
2 further
2 future
1 gags
1 gasp
6 gecko
1 geese
6 gems
7 genius
1 geography
2 geometry
1 gigantic
1 ginger_ale
1 gingerbread
1 gladly
1 glancing
1 glassy
1 glitter
3 globe
2 goats
3 godmother
1 goo
2 goodies
1 goodnight
1 goofing
1 googly
4 goop
1 gown
2 graham_crackers
1 graph
1 grasp
1 grasping
1 grayish
1 greener
2 greenish
1 greyish
1 grill
1 grilled
1 grouchy
1 grounded
1 grunts
1 guardian
3 guardians
1 guests
1 guided
1 habitat
3 haircut
1 halftimes
1 hallucinating
3 hallway
2 halter
2 hammerhead
1 handsome
2 handwriting
2 hanger
1 happily
1 hardness
2 hard_drive
1 harm
4 harmless
2 harshest
1 hash_brown
3 hatch
2 hatched
1 hatching
3 hay
1 hayloft
1 hayride
2 haystack
4 headache
1 headaches
3 health
3 healthy
1 heat
1 heater
1 heats
3 heaven
1 heavens
1 heel
1 heels
1 heifer
2 helmet
1 helper
1 helpers
1 hens
1 herd
1 hero
1 heroes
1 herons
1 hijacked
1 highschooler
2 hip
1 hippopotamuses
1 hitchhiking
2 hobby
1 hog
1 holder
1 holy
2 homeless
2 homeroom
1 honey
2 honk
1 honking
1 honored
1 honor_roll
1 hood
2 hopefully

1 hopscotch
1 hornet
2 horror
1 horsefly
2 horseshoes
1 hot-air_balloon
3 hot_tub
4 howl
2 howls
1 howling
2 huff
1 hug
1 hugs
2 hula
3 hula_hoops
3 hunter
1 hunters
2 hunting
1 hurricanes
1 hut
2 hyper
1 hyperactive
2 icicle
1 icicles
1 icing
3 icky
1 igloo
1 ignored
1 ignoring
1 iguana
1 iguanas
1 illustrated
2 imaginary
1 immediately
1 immigrants
1 important
1 impossible
1 improved
5 inchworm
1 infield
4 information
1 initial
3 initials
1 inject
1 ink
2 insane
1 insists
3 instruments
4 internet
2 interrupters
1 introduce
1 introduced
1 invaders
1 inventions
2 itchy
1 item
1 items
1 jaguars
5 jam
4 jazz
1 jealous
1 jean
3 jellyfish
1 jerking
4 jersey
2 jewelry
3 jewels
1 jockey
1 jog
2 jogging
1 joke
1 joking
2 journey
1 juggle
2 juicy
1 jumper
2 jumpers
1 jumping_jacks
1 jumpy
6 karaoke
3 kennel
2 kidding
3 kidnap
2 kidnapped
3 kite
1 kites
1 knots
1 knuckles
8 komodo_dragon
1 kooky
1 labeled
1 laboratory
1 lambs
1 laminated
1 landmarks
1 laptops
1 lately
1 launched
1 laundry
1 lawbreakers
1 lawn
8 layer
1 layers
1 lean
1 leap
3 lemon
4 lemonade
1 lemonades
1 length
1 lengths
1 leopard
2 librarian
3 license
4 licorice
2 lifeguard
3 lifeguards
1 limb
1 linemen
2 links
1 lipstick
1 littler
1 llamas
1 loads
2 lobster
1 locations
3 lockers
1 lonely
1 longest
1 long-necked
1 look-alike
1 loop
1 lopsided
1 lord
1 loudly
1 lousy
1 lovey-dovey
4 low
1 lower
1 lowered
1 loyalty
1 luau
1 luck
4 luckily
4 luge
2 lump
1 lumpy
1 lunchbox
1 lunchroom
4 lunchtime
1 lungs
3 lurkers
2 magazine
2 magazines
1 magical
2 magically
5 magician
1 magicians
1 mail
3 mailbox
6 Maine_coon
1 major
4 make-believe
2 makeover
3 maker
2 male
2 mammals
3 manager
2 mane
1 manager
2 mansion
1 manta_rays
1 maracas
1 marbleized
1 marching
1 markings
1 marlins
5 mass
2 massage_oil
1 matchbox
1 mattress
1 mealworm
1 meaner
1 meanings
1 measured
1 measuring
1 measurements
1 meatballs
1 meat-eater
1 medallions
4 medley
2 members
1 membership
1 memorized
2 memory
1 menace
1 mentioned
1 merman
1 messaging
1 meter
2 mile
1 millet
1 minerals
2 miserable
2 misery
1 mode
1 modern-day
1 modules
1 mohawks
1 moist
1 moister
1 moisturizer
1 mold

1 moldy
2 mole
2 moment
1 monarch
2 mosquito
3 moth
1 moths
1 motorized
1 mower
2 mozzarella
1 muddy
1 mudpie
2 mulch
1 multiplying
1 mummy
1 mushy
1 musicians
1 musketeers
1 mustard
2 mutant
4 mysterious
1 mystery
2 naked
7 nanny
2 napkins
1 naps
1 narrow
1 nasty
1 national
2 navigator
3 navy
1 nay
2 nearby
1 necklaces
1 nectar
1 negotiate
1 nerves
2 nervous
1 newest
3 news
3 newspaper
2 newt
1 nicer
2 nickels
1 nickname
1 nieces
1 nightlight
3 nightmares
3 nighttime
1 nocturnal
1 nonfiction
1 noon
1 north
3 note
3 notebook
1 notes
2 nowhere
1 nuclear
1 nursery
3 oaf
1 oak
1 obsessed
3 obviously
1 octopus-like
6 odd
1 oddly
3 offer
1 offered
1 offers
1 offering
1 officer
1 officers
2 off-limits
1 ogre
1 oil_pastels
3 old-fashioned
2 onions
2 op_art
3 opponent
2 opposing
1 opposite
1 orange-ish
1 organs
1 orphan
1 ostrich
2 ostriches
2 otter
1 ought
2 outfield
1 outlines
1 out-of-bounds
1 outstanding
1 oval
1 overlap
1 overpower
1 overthrow
1 overthrowing
1 pacifier
1 pacifiers
1 package
1 packet
1 pain
1 pal
1 palm
3 pan
2 papier-mache
4 parachute
2 paragraphs
2 parcheesi
1 partly
1 passage
1 password
2 paste
1 pastor
2 patio
1 paused
1 pausing
1 payback
1 peace
2 peaceful
2 peach-ish
5 peacock
6 peak
1 peanut-sized
1 peck
5 pelican
1 pelicans
1 peppers
2 per
1 period
2 periods
1 pesty
2 pet_shop
1 pet_store
1 PG
4 PG-13
3 photo
1 photography
4 pi
1 picky
2 pigeon
1 pigeons
1 pillowcase
1 pinch_pot
1 pineapple
1 pineapples
2 pine_cone
1 pinky
1 pioneer
1 pioneers
2 pipe-cleaners
3 pit
2 planetarium
1 platform
1 playful
2 playroom
4 playset
1 pleased
1 plow
1 plows
1 plug
2 PM
1 poacher
1 poachers
1 poem
1 pogo_stick
8 poison
4 poison_ivy
3 poker
5 police_officer
1 polka-dot
1 pontoon
1 pooper-scooper
4 popular
1 portrait
1 pose
1 possible
3 post
2 posters
1 potty-trained
1 pounce
1 pounces
1 powerful
1 power_line
4 practically
1 prank
1 prayed
1 prayer
1 pregnant
2 prettiest
2 preview
1 previews
1 price
2 primary
1 print
3 prison
1 private
1 professionals
1 promise
2 promised
1 promotion
3 property
2 pterodactyl
2 public
1 puck
2 puddle
1 puddles
2 pueblo
2 puff
1 puffy
3 punished
1 punishment

7 pupil
4 pure
1 purplish
2 purpose
1 purr
2 puzzled
2 pyramid
1 quality
1 quicker
1 quicksand
1 quilt
1 quits
2 quiz
2 quizzes
1 rabies
1 racecourses
1 racehorse
1 racer
1 rack
1 radioactive
1 raft
1 railroad
1 rails
2 rainbowfish
1 raindrops
1 rainforests
1 rake
2 rally
1 ram
1 ranch
1 raptor
1 raptors
5 rare
1 raspberry
8 rated
3 realm
1 realms
2 rec
1 recipe
6 recital
1 recitals
6 recorder
1 recorders
1 recycle
1 redo
1 refurbish
1 refurbished
2 refused
4 register
1 registered
1 rein
2 reins
2 reindeer
3 reindeers
1 rejected
1 rejoiced
1 relaced
1 relaces
2 related
1 relative
1 relatives
2 relays
3 release
3 relief
2 religion
1 remind
1 reminded
1 reminds
1 reminding
2 removed
5 rent
3 repair
1 replaced
3 reptiles
1 reputation
1 research
1 researcher
1 reset
2 responsible
1 restroom
1 retired
2 reviewing
2 reward
2 rewind
1 rewinding
4 rhino
1 rhyme
2 rhyming
3 ribbons
2 rice
2 rich
1 riddle
1 right-hand
4 rim
1 rims
1 ringbearers
1 rink
1 rises
1 road_rash
2 roast
5 robber
2 robe
1 robotic
1 rock-and-roll
2 roller
2 rookie
1 roommate
1 root
1 rotate
1 rough
1 rowdy
3 rub
2 rubber
2 rubber_band
1 ruiner
1 ruling
3 runaway
1 runner
1 runt
1 sacs
1 saddle
2 safari
1 safer
1 saints
5 salamander
1 salmon
2 salsa
1 saltwater
1 samurais
1 sandals
1 sault
1 scanning
2 scarecrow
1 scarecrows
1 scarier
1 scent
1 scheduled
1 schoolmates
3 schoolroom
1 schoolwork
1 schoolyard
1 scientist
1 scientists
1 scolded
2 scoot
1 scrambled
4 scrap
2 scrapbook
1 scrapes
1 screening
1 scribbling
1 scrunch
2 scuba_diver
1 sculptures
1 seafood
1 seagull
1 seagulls
1 search
1 seashells
5 seaweed
1 secondary
1 secrets
1 seek
2 self-portrait
2 senior
1 seniors
1 sense
1 sensor
1 separately
1 series
3 serious
1 seriously
1 serving
2 service
1 settles
2 sewer
3 shall
3 shamrock
2 shank
2 shave
1 shaving
2 shed
2 sheepdog
3 shelf
4 shelter
1 shelty
2 shield
2 shields
2 shiny
1 shoe_prints
1 shore
2 shortcut
1 shorten
1 shortest
1 shortstop
1 shoulders
1 shoveling
5 shrimp
1 shrinks
1 shrinking
3 shuttle
1 siblings
1 sickest
1 sidekick
1 sideways
1 sidewinder
3 sight
1 sights
1 sight_words
1 silent
2 silliest
1 silverback

2 simple
1 sincerely
1 sir
1 siren
1 site
1 sized
1 skaters
3 sketch
1 sketching
1 sketches
2 ski_lift
1 slap
3 slapped
1 slapping
11 slash
1 slat
1 sleeper
1 sleepers
1 sleeve
1 slight
2 slime
1 slimy
3 slingshot
2 slinky
4 slippery
2 slithering
2 sloppy
1 slower
2 slug
1 sluggish
1 slurp
1 smacking
5 smallest
4 smarter
1 smartest
2 smeared
1 smelly
2 smiley
3 smooth
1 smoothie
2 smoothies
1 snatched
2 sneaky
1 sneezing
3 sniff
1 sniffs
2 snout
1 snowboarder
1 snowpants
1 snowy
4 softly
5 solar_system
1 someday
1 someplace
2 somersault
2 somersaults
1 soothing
1 sour_cream
1 space_shuttle
3 spanked
2 sparkly
1 sparks
3 speaker
1 speakers
1 speaks
1 species
4 speckled
1 speedboat
1 speed-skating
1 speedy
1 speller
2 spicy
1 spider-like
1 spiderweb
1 spikes
1 spinach
3 spinosaurus
1 spine
1 spinner
2 splattered
1 splatters
1 splices
1 sponge
1 sponginess
3 spooky
1 sportsmanship
1 spotty
1 spray
1 sprayed
1 sprinkling
1 sprite
1 sqeak
1 squad
1 squash
2 squat
3 squawk
1 squawks
3 squeak
1 squeaking
1 squeaks
1 squeaky
1 squirtgun
1 stabbed
2 stable
1 stacked
1 stadium
2 staff
1 stain
1 stairway
1 staple
1 starlight
1 starter
1 startled
1 stash
1 station
5 stations
2 steady
1 steep
1 steering
1 stegosaurus
1 stencils
1 stepmom
2 stepmother
2 stepsisters
1 stepstool
1 stew
1 stinkiest
2 stinky
2 stitches
2 stock
1 stocking
2 stomp
3 stool
1 storybook
1 storyteller
1 straightened
1 stranded
2 strap
1 stray
2 strict
1 strike
1 strikes
1 striking
1 stringy
1 strokes
2 stronger
3 strongest
1 studio
1 stuffing
1 stuffy
1 stump
1 stunned
1 sturdy
1 subtract
1 subtracting
8 sucker
1 suckers
1 suffered
4 suffering
1 suffixes
1 sugary
1 summertime
1 summon
1 sumo
1 sundae
1 sunfish
1 sunflower
3 sunglasses
1 sunrise
1 sunscreen
2 sunset
1 superpowers
2 supper
2 supposedly
1 surgery
1 survive
2 survived
1 surviving
3 suspended
1 suspending
1 swallow
1 swallows
2 swamp
1 sweaty
3 sweets
1 swept
1 swerve
4 swirly
1 switches
2 systems
3 tabby
1 tablecloths
1 tae-kwan-do
2 tale
1 tallish
2 tannish
1 tape-record
2 tape-recorder
4 tarantula
3 tarantulas
1 target
1 tattoo
1 taxi
5 tea
1 teacup
1 teammates
6 teapot
2 tease
1 teasing
1 teen
1 telescope
1 television

3 temperature
1 terrible
2 terrified
1 terrorized
1 texture
2 thankful
3 theaters
1 theme
1 thesaurus
3 thick
3 thicker
1 thingamabobber
1 thingamajigger
2 thongs
1 three-legged_race
3 throat
1 throughout
1 thunderstorm
2 ticklish
3 tic-tac-toe
1 tighter
1 timeout
1 timer
1 timpani
1 tiniest
1 tinted
1 tiring
2 title
5 toes
1 tofu
4 tokens
1 tomcat
2 tools
1 top_hat
1 topic
1 toppings
1 toppling
3 tortilla
1 tortoise
5 tortoises
3 total
4 touchdown
3 tour
1 trailer
1 trainer
2 trainers
1 trampled
2 transform
1 translate
1 transport
1 transporter
1 transports
1 trial
5 tricky
1 tripling
1 trivia
2 troll
4 trolls
7 trot
2 trotting
1 troublemaker
1 troublemakers
1 troupe
1 truly
1 trumpeter_swans
1 trust
3 trusted
2 trustworthiness
3 truth
2 T-shirt
3 tsunami
1 tuck
1 tucked
4 tug-of-war
2 tulip
1 tumbled
1 tumbling
2 tunafish
3 tune
3 turquoise
1 tutor
1 tweezers
1 twisting
1 two-ended
1 two-headed
1 uglier
2 umbrella
1 umpire
1 unbelievable
1 unclip
2 uncurl
2 underground
5 underneath
1 unfriendly
3 unicorn
1 uniforms
2 unique
1 unit
1 units
2 unknown
2 unlock
2 unlocked
1 unroll
1 unsafe
2 untied
1 unties
1 unusual
2 uphill
1 upstate
1 user
1 usual
4 vampire
2 variety
1 vegetarian
1 vein
1 veins
3 velcro
3 velociraptor
1 velvet
2 venom
1 ventriloquist
1 vertically
1 veterinarian
1 vice-president
1 vicious
5 videocamera
4 videotape
2 videotaped
1 view
1 villagers
2 villages
2 vines
1 violent
1 violin
1 viperfish
1 visions
1 volume
1 volunteer
1 volunteered
1 volunteers
1 vomit
1 vote
1 vowels
2 vulture
1 wad
1 wade
2 wagon
2 waiter
1 walkers
1 walkie-talkies
3 wallet
3 wand
1 wander
1 wanders
1 wands
1 warmer
1 warning
1 warnings
2 waste
2 wasted
1 waterbeds
1 waterfall
1 waterfalls
7 watermelon
1 weasel
1 weaving
1 weavings
2 webbed
2 website
3 weeds
1 weekdays
1 weight
1 wiener
1 welcomed
3 western
1 wheat
1 wheelchair
1 wheelchairs
1 whereabouts
1 whichever
2 whip
1 whips
5 whipped_cream
3 whirlpool
2 whispered
1 whitish
1 whiter
4 whose
1 wider
3 wig
2 wiggle
2 wiggly
1 winner
1 wipeout
2 wireless
7 wizard
1 wombat
5 wonderful
5 wooden
1 woodman
1 wool
1 workbook
1 worksheet
1 worse
1 worth
1 wrappers
1 wren
1 wrestlers
1 wrinkle
1 wrist
4 wristband
2 writer

1 x-ray
2 yarn
1 yearbook
2 yellowish
1 yellowjackets
1 yolk
2 yoyo
1 zap
1 zapped
1 zombie
1 ziggurats