

Early Indicators of Language Delay in Preterm Infants

By Glenda C. Molina

Sc.M., Brown University, 2012

B.S., Tufts University, 2004

A dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the Department of Cognitive, Linguistic & Psychological
Sciences at Brown University

Providence, Rhode Island

May 2014

© Copyright 2014 by Glenda C. Molina

This dissertation by Glenda C. Molina is accepted in its present form by the Department
of Cognitive, Linguistic & Psychological Sciences as satisfying the dissertation
requirement for the degree of Doctor of Philosophy.

Date _____

James Morgan, Director

Recommended to the Graduate Council

Date _____

Dima Amso, Reader

Date _____

Sheila Blumstein, Reader

Approved by the Graduate Council

Date _____

Peter Weber, Dean of the Graduate School

GLEND A C. MOLINA

Department of Cognitive, Linguistic & Psychological Sciences (CLPS)
Box 1821 • Brown University • Providence, RI 02912
Email: Glenda_Molina@brown.edu • Phone: (401) 863-5182

EDUCATION

- Brown University**; Providence, RI, USA
Ph.D., Cognitive Science 2007–2013
Early Indicators of Language Delay in Preterm Infants
Advisor: James L. Morgan, Ph.D.
- Sc.M., Cognitive Science** 2012
Reconsidering the Case for Acquired Distinctiveness in the Acquisition of Native-Language Phonetic Categories
Advisor: James L. Morgan
- Tufts University**; Medford, MA, USA 2000–2004
B.S., Biology; Spanish with a focus on literature
Honors in Biology, *Summa Cum Laude*
- Universidad de Alcalá**; Alcalá de Henares, Madrid, Spain Fall 2003
Study Abroad Program

GRANTS & FELLOWSHIPS

- National Research Service Award**, Eunice Kennedy Shriver NICHD 2010–2013
Brain Science Graduate Research Award, Brown University 2008
Gates Millennium Scholarship 2000–2004

AWARDS & HONORS

- Constantine Ghikas Prize in Romance Languages**, Tufts University 2004
Paula Frazier Poskitt Scholarship in Biology, Tufts University 2003
Louis Levin Scholarship, Tufts University 2001
National Society of Collegiate Scholars 2000

RESEARCH EXPERIENCE

- Brown University**, Dept. of CLPS, Infant Research Laboratory
Primary Advisor: Dr. James L. Morgan
- Postdoctoral Research Associate* 2013–present
- Assisting with a longitudinal study assessing impact of receptive language skills, language input, and cognitive abilities on language outcomes in toddlerhood and early school years

Graduate Student 2007–2013

- PI for longitudinal study investigating early indicators of language delay in preterm and full-term children
- Assessed potential sensitivities to phonotactics at 6 months
- Explored roles of acquired distinctiveness *versus* acquired equivalence, and of frequency *versus* markedness in the acquisition of phonetic categories
- Characterized voicing distinctions in production of stops across dialects of Spanish

Research Technician 2006–2007

- Contributed to efforts to create ties between RI's Latino Community and the Infant Lab to begin studies with Spanish-English bilingual infants
- Assisted with design, set-up and implementation of studies as well as infant subject recruitment

Tufts University, Dept. of Child Development
Supervisor: Dr. Calvin L. Gidney

Research Assistant 2005–2006

- Aided with research on children's television programming, observing and coding dialects used in various animated series to investigate how these are used to personify characters

Tufts University, Dept. of Biology, Neurobiology Research Laboratory
Supervisor: Dr. Barry A. Trimmer

Undergraduate Researcher 2001–2003

- Investigated effects of nitric oxide on the CNS of the tobacco hornworm, *M. sexta*, using a variety of methods, including immunocytochemistry and electrophysiological recordings

PUBLICATIONS

Molina Onario, G., & Morgan, J. (In preparation). Infant Sensitivities to Native Language Lexical Stress and Phonotactics Predict Comprehension in Full-term and Preterm 2-Year-Olds. Paper to be submitted to *Child Development* (2014).

Molina Onario, G., & Morgan, J. (In preparation). Sensitivity to Mispronunciations in Full-term and Preterm Toddlers. Paper to be submitted to *Journal of Memory and Language* (2014).

PRESENTATIONS

Molina, G., & Morgan, J. (2012). Infant Sensitivities to Native-Language Phonotactics as a Predictor of Later Language Outcomes in Preterm and Full-term Children. XVIII Biennial International Conference on Infant Studies, Minneapolis, MN.

- Molina, G., & Morgan, J. (2011).** Preterm and Full-term Infants' Sensitivity to Permissible Native-Language Sound Sequences. 37th New England Conference on Perinatal Research, Chatham, MA.
- Molina, G., & Morgan, J. (2011).** Sensitivities to Native-Language Phonotactics at 6 Months of Age. Society for Research in Child Development 2011 Biennial Meeting, Montreal, Canada.
- Molina, G. (2008).** Reconsidering the case for acquired distinctiveness in the acquisition of native-language phonetic categories. First year project presentation, Dept. of Cognitive & Linguistic Sciences, Brown University.
- Molina, G., & Morgan, J. (2008).** The voicing distinction in Spanish: A prelude to what Spanish-learning infants can reveal about native-language phonetic category acquisition. International Conference on Literature, Language & Linguistics, Athens, Greece.
- Molina, G. (2003).** The localization of NO-synthesizing cells in the frontal ganglion of *M. sexta* larvae. Independent research presentation, Department of Biology, Tufts University.
- Molina, G., Zayas, R., Siddiqui, A. & Trimmer, B. (2003).** NO-induced behavior in *M. sexta* larvae. 29th Annual East Coast Nerve Net, Woods Hole Science Center.
- Molina, G. (2002).** Nitric oxide signaling in the central nervous system of *M. sexta* larvae. New England Science Symposium, sponsored by Harvard Medical School, Boston, MA.

TEACHING CERTIFICATES & EXPERIENCE

Brown University, Sheridan Center Teaching Certificate Programs

- | | |
|--|-----------|
| • <i>Certificate I: Teaching Seminar</i> | 2009 |
| • <i>Certificate II: Classroom Tools Seminar</i> | 2010 |
| • <i>Certificate IV: Teaching Consultant Program</i> | 2010–2013 |

Brown University, Dept. of CLPS

Teaching Assistant

- | | |
|--|-------------|
| • Introduction to Cognitive Neuroscience | Fall 2010 |
| • Human Cognition | Spring 2010 |
| • Introduction to Cognitive Science | Fall 2009 |
| • Children's Thinking | Spring 2009 |

Tufts University, Dept. of Community Health

Supervisors: Dr. Edith Balbach and Carol Baffi-Dugan

Program Coordinator and Tutor for the Health Careers Opportunity Program

2004–2006

University of Massachusetts Boston, Health Careers Opportunity Summer Program

Supervisor: Shirley Smith

High School Science Co-Instructor

Summer 2005

ACTIVITIES

Member, Applied Methodology in Psychiatry, Human Behavior, and Neurosciences Reading Group (AMPHIBIAN)	2013–present
Coordinator, Cognitive Development Reading Group	2010–2012
Co-organizer, CLPS Ling Lang Lunch	2011
Member, Cognitive Development Reading Group	2008–2010
Graduate Student Organizer for Student Recruitment Weekend for Dept. of Cognitive & Linguistic Sciences	2008–2009
Graduate Student Representative for Dept. of Cognitive & Linguistic Sciences	2008–2009
Mentor, African American, Latino, Asian/Asian American, and Native American (ALANA) Student Mentoring Program	2006–2008

RELEVANT COURSEWORK

Developmental Cognitive Neuroscience	Fall 2011
The Developing Brain	Fall 2010
Individual Differences and Longitudinal Studies of Language Development, Cognition Seminar	Spring 2010
Language Processing	Fall 2009
Multivariate Statistics, Independent Study on the Development of Hearing	Spring 2009
Experimental Design, Child Language Acquisition, Language Seminar	Fall 2008
Acoustic Phonetics, Psychological Acoustics, Perception Seminar	Spring 2008
Phonological Theory, Topics in Child Language Acquisition	Fall 2007
Speech Prosody	Spring 2007
Neuroimaging and Language, Language and the Mind	Fall 2006
Development of Language	Fall 2005
Neurobiology, Biostatistics, Translation	2000–2004

Acknowledgements

First and foremost, I would like to thank the families and children who committed their time and energy to participate in the many studies making up this research project. Seeing participants grow from tiny adorable babies into little toddlers with unique personalities was a reward on its own, and without them, there would be nothing for you to read today.

Thank you to my advisor, James Morgan, for giving me the confidence to take on such an ambitious project as a graduate student. Without his encouragement and support, I would have never received the National Research Service Award that made this work possible. Above all, I thank him for teaching me independence and self-motivation and for always getting me to think critically about the “big picture” when designing experiments and analyzing data.

I would also like to thank Dima Amso, who has always been a great supporter of my work. She has been an incredible and inspirational addition to our department, and I feel lucky to have her as one of my mentors. I thank her for all the advice she has given me throughout the duration of this project. As a fellow Tufts’ Jumbo, how could she be anything but awesome?

Thank you to Sheila Blumstein for playing such an integral part of my graduate school career. She has seen me through this entire journey, first as a member of my first year project committee, and now in this final chapter as a member of my dissertation committee. Sheila truly embodies the type of researcher we all aspire to be and I feel incredibly grateful for all the wisdom she has imparted on me throughout the years.

I am very appreciative for the help and support I have received over the years from members of the Metcalf Infant Research Lab, especially my “lab sisters”, Jill Thorson, Jie Ren, Claire Paulson, Elena Tenenbaum, Naomi Feldman, and Erin Conwell. Special thanks go to Lori Rolfe, for leading this project’s full-term recruitment endeavors, and to Stephanie Tin, Meghan Wenzel, and Emily Davis, for helping me code all the data collected for this dissertation. I would also like to acknowledge Dr. Betty Vohr and members of the Women and Infants Follow-Up Clinic for providing me with the resources to recruit preterm participants, and Richard Jones and Ronald Seifer for making structural equation modeling a little less intimidating. Thank you to the Eunice Kennedy Shriver National Institute of Child Health and Human Development for awarding me grant 1F31HD065406-01 and making this work possible.

I am very fortunate to have amazing friends who have been an incredible support system during this time. Thank you Devorath Ruiz, Shioban Torres, Yakhira Encarnacion Patterson, Jenny (Scullin) McCarthy, Erika Alexander, Jill Thorson, Deanna Macris, and Mike Ferrier for the many fun dinners, entertaining conversations (especially about alligators and crocodiles), and scary movie-watching. A special thanks goes to Julie Markant, for always reminding me that the emotional rollercoasters experienced while dissertating are “developmentally appropriate”. She kept me grounded, making me realize that this seemingly impossible time in my life is but a necessary stepping-stone for things to come, and that in the grand scheme of life dissertating is but a small thing that would soon, too, pass.

Of course, I would not be here without my loving family. Thank you Mami and Papi for all the sacrifices you have made to ensure that my siblings and I have the

opportunities you never had. Thanks to my sisters, Melissa and Iliana, and to my brother, Daniel, for always believing in me. Terrell, Amaya, and Olivia, thanks for always making me smile- your Titi loves you very much and feels blessed to have you in her life!

Last, but definitely not least, I would like to thank my soon-to-be husband, Randy Onario. His unconditional love and support have been indispensable to me during the completion of this work and throughout my graduate school career. Bay, I love you and I look forward to our life together.

Table of Contents

Chapter 1: Premature Birth and Language Development

1.1	Introduction -----	1
1.1.1	<i>Word-Learning in Typically Developing Children</i> -----	1
1.1.2	<i>The Case for Linguistic Delay in Children Born Preterm</i> -----	4
1.2	Project Details -----	10
1.2.1	<i>Goals</i> -----	10
1.2.2	<i>Hypotheses and Outline of Receptive Language Experiments and Collection of Language Outcome Data</i> -----	11
1.2.3	<i>Participants</i> -----	18

Part I- Measuring Predictors Of Language Outcomes: Receptive Language Skills And Cognitive Abilities In The First And Second Years Of Life

Chapter 2: Sensitivities to Native-Language Sound Structures

2.1	Sensitivities to Lexical Stress -----	23
2.1.1	<i>Participants</i> -----	23
2.1.2	<i>Method</i> -----	24
2.1.3	<i>Results</i> -----	27
2.1.4	<i>Discussion</i> -----	29
2.2	Sensitivities to Phonotactics -----	31
2.2.1	<i>Participants</i> -----	34
2.2.2	<i>Method</i> -----	34

2.2.3 Results	38
2.2.4 Discussion	40
2.3 General Discussion	43

Chapter 3: Representations of Native-Language Word Forms

3.1 Sensitivities to Mispronunciations	45
3.1.1 Participants	46
3.1.2 Method	46
3.1.3 Results	51
3.1.4 Discussion	54
3.2 Speed and Efficacy of Word Processing	58
3.2.1 Participants	60
3.2.2 Method	60
3.2.3 Results	67
3.2.4 Discussion	77

Chapter 4: Cognitive Skills as Modulators of Language Development

4.1 Rose Task: Assessing Immediate Visual Recognition Memory at 12 Months	92
4.1.1 Participants	94
4.1.2 Method	94
4.1.3 Results	100
4.1.4 Discussion	107
4.2 Mullen Scales of Early Learning: Assessing Visual Reception and Fine Motor Skills at 26 Months	111
4.2.1 Participants	113

4.2.2 Method -----	114
4.2.3 Results -----	115
4.2.4 Discussion -----	116

Part II- Measures Of Language Outcomes In The Second And Third Years Of Life

Chapter 5: Measurements of Child Language Outcomes

5.1 Language Outcome Assessment Tools -----	119
5.1.1 Parental Surveys -----	119
5.1.2 Lab Assessment -----	121
5.1.3 Participants -----	122
5.1.4 Method -----	123
5.1.5 Results -----	125
5.1.6 Discussion -----	137

Part III- Relations Between Language, Cognitive And Environmental Predictors, And Later Language Outcomes

Chapter 6: Connecting the Dots Through Structural Equation Modeling

6.1 Why Use SEM Instead of Multiple Regression Analysis? -----	142
6.2 Data Analysis With SEM -----	143
6.2.1 Participants -----	146
6.2.2 Results -----	146
6.2.3 Discussion -----	158

Part IV- Project Summary And Significance

Chapter 7: General Discussion and Conclusions

7.1	Summary of Measures of Receptive Language Abilities, Cognitive Skill, and Language Outcomes -----	164
7.2	Summary of SEM Analyses -----	168
7.3	Study Limitations (and Merits) -----	170
7.4	Future Directions -----	175
	References -----	178
	Appendix A: Guidelines for Completing the WG-MCDI -----	184
	Appendix B: Guidelines for Completing the WS-MCDI -----	186
	Appendix C: Mplus Codes for SEM Analyses -----	188

List of Tables

- Table 1. Subject demographic information.
- Table 2. Stimuli for experiments assessing infant sensitivities to high and low phonotactic probability patterns at 6 and 9 months.
- Table 3. Stimuli for experiment assessing sensitivities to mispronunciations at 19 months.
- Table 4. Stimuli for experiment assessing speed and efficacy of word processing at 26 months.
- Table 5. Mean novelty scores and standard deviations for Rose Task.
- Table 6. Correlations between cognitive outcome measures collected with the Rose Task at 12 months and the Mullen Visual Reception and Fine Motor subscales at 26 months.
- Table 7. Means, standard deviations, and t-test results comparing full-term and preterm participants' language outcome scores collected with the MacArthur Bates Communicative Development Inventories- Words and Gestures (WG-MCDI) at 14.5 months.
- Table 8. Means, standard deviations, and t-test results comparing full-term and preterm participants' language outcome scores collected with the MacArthur Bates Communicative Development Inventories- Words and Sentences (WS-MCDI) at 26 months.
- Table 9. Correlations between language outcome measures collected with the WG-MCDI at 14.5 months and the WS-MCDI at 26 months.
- Table 10. Correlations between language outcome measures collected with the WG-MCDI at 14.5 months and the Mullen Receptive Language Subscale (MRLS) at 26 months.
- Table 11. Correlations between language outcome measures collected with the WS-MCDI and the MRLS at 26 months.
- Table 12. Summary of latent constructs, indicator variables, and measurements used for structural equation modeling (SEM) analyses.

List of Figures

- Figure 1. Language learning model assessed with structural equation modeling (SEM).
- Figure 2. Mean difference scores and standard errors assessing group differences in sensitivities to lexical stress at 4 and 6 months.
- Figure 3. Mean difference scores and standard errors assessing group differences in sensitivities to phonotactic probability patterns at 6 and 9 months.
- Figure 4. Mean proportion looking to the target during Correct, Mispronunciation, and Novel trials at 19 months.
- Figure 5. Mean proportion looking to the target during Correct, 1-Feature Change, 2-Feature Change, 3-Feature Change, and Novel trials at 19 months.
- Figure 6. Example of trial structure for experiment assessing speed and efficacy of word processing at 26 months.
- Figure 7. Target accuracy following target noun onset for Vanilla, I-verb, U-verb, I-adj, and U-adj trials for experiment assessing speed and efficacy of word processing at 26 months.
- Figure 8. Target accuracy following target verb onset for Vanilla, I-verb, U-verb, I-adj, and U-adj trials for experiment assessing speed and efficacy of word processing at 26 months.
- Figure 9. Target accuracy following target adjective onset for Vanilla, I-verb, U-verb, I-adj, and U-adj trials for experiment assessing speed and efficacy of word processing at 26 months.
- Figure 10. Mean reaction time for target orientation following target noun onset for Vanilla, I-verb, U-verb, I-adj, and U-adj trials for experiment assessing speed and efficacy of word processing at 26 months.
- Figure 11. Mean reaction time for target orientation following verb onset for I-verb and U-verb trials for experiment assessing speed and efficacy of word processing at 26 months.
- Figure 12. Mean reaction time for target orientation following adjective onset for I-adj and U-adj trials for experiment assessing speed and efficacy of word processing at 26 months.
- Figure 13. Potential causal pathways for poor language outcomes.
- Figure 14. Example of a Rose Task trial.

- Figure 15. Rose task test phase stimulus pairings for face and abstract pattern trials.
- Figure 16. Mean proportion looking to abstract patterns and faces during the familiarization phase of the Rose Task.
- Figure 17. Mean proportion of time spent looking away during the familiarization phase of the Rose Task.
- Figure 18. Mean trial length required to attain 3- and 10-second Rose Task familiarization criteria for abstract patterns and faces, respectively.
- Figure 19. Test phase mean novelty scores for abstract pattern and face trials for the Rose Task.
- Figure 20. Mean number of words comprehended and produced by full-term and preterm participants at 14.5 months.
- Figure 21. Mean number of words produced by full-term and preterm children at 26 months.
- Figure 22. Mean use of words expressing Past, Future, and Absence, and mean use of Word Endings including plural –s, possessive ‘s, progressive –ing, and past tense –ed at 26 months.
- Figure 23. Means for Irregular Words and Overregularized Words used at 26 months.
- Figure 24. Mean number of more Grammatically Complex forms used by 26-month-olds.
- Figure 25. Mean Length of the Three Longest Utterances produced by 26-month-olds.
- Figure 26. Mean raw score on the Mullen Receptive Language Subscale (MRLS) for full-term and preterm toddlers surveyed at 26 months.
- Figure 27. Structural equation (SE) model of the three latent language factors and their corresponding indicator variables.
- Figure 28. SE model of F_1 on F_2 on F_3 with gestational age (GA) as a covariate.
- Figure 29. SE model of F_1 on F_2 on F_3 with GA and the Mullen Visual Reception Subscale (MVRS) score as covariates.
- Figure 30. Structural equation model of F_1 on F_2 on F_3 with GA, MVRS Score, and maternal education as covariates.
- Figure 31. Summary of key findings established through SEM analyses.

Chapter 1: Premature Birth and Language Development

1.1 Introduction

Preterm children constitute a population known to be at risk for language delays. Research has documented that very-low-birth-weight infants often display slower rates of growth in vocabulary and grammar than do full-term infants; moreover, preterm infants may be at heightened risk for specific language impairment (Foster-Cohen et al., 2007; Kern & Gayraud, 2007; Marston et al., 2007; Ortiz-Mantilla et al., 2008; Sansavini et al., 2006; Stolt et al., 2007; Vohr, Garcia-Coll, & Oh, 1988, 1989; Vohr et al., 2000). Such delays have been implicated in deficits in early school achievement (Pritchard et al., 2009; Wolke, et al., 2008). However, little is known about what underlies these language difficulties and whether there are indicators within preverbal stages of development that predict future complications. The aim of the longitudinal project reviewed in this dissertation addresses issues of linguistic delays in preterm children from a novel perspective by examining receptive abilities during infancy. Rather than focusing on standardized measures of language outcomes during childhood, the set of studies discussed will prospectively characterize the progression of skills required for word-learning in preterm infants relative to full-term controls.

1.1.1 Word-Learning in Typically Developing Children

Word-learning is fundamental to language acquisition. From birth, full-term infants have perceptual sensitivities and preferences that facilitate word-learning. At the most perceptually basic level, infants are born with preferences for speech over other types of

sounds (Vouloumanos & Werker, 2004) and for native-language speech over foreign speech (Moon et al., 1993). Furthermore, they are capable of recognizing mothers' voice over that of others' (DeCasper & Fifer, 1980). What infants are specifically tuning to during these early stages of language acquisition are the prosodic properties of speech. Sensitivities to rhythm, stress and intonation initiate the word-learning process by facilitating word segmentation and recognition in fluent speech. Critically, some of these infant sensitivities to native-language properties have been linked to language outcomes in the second year of life.

As early as 2 months of age English-learning infants can discriminate between native and non-native speech solely on the basis of intonational differences (Mehler et al., 1988). However, recognizing native-language speech is not enough to learn words; the task of word segmentation must first be mastered. As adults this problem becomes obvious when considering what it is like to listen to speech in an unfamiliar language: it is very difficult, if not impossible, to tell where one word ends and the next begins. In infancy, this problem is resolved in part by sensitivities to stress cues. At 4 months, full terms can distinguish between words with trochaic (strong/weak- e.g. GABA) and iambic (weak/strong- gaBA) stress patterns, showing a preference for that pattern which is dominant in their native language (Herold et al., 2008). By 7.5 months infants' preference for dominating stress patterns guides the early stages of word segmentation in fluent speech. For instance, infants learning English, which is dominated by trochaic stress, can segment trochaic but not iambic words from fluent; not until 10.5 months do infants succeed at segmenting both types of words (Jusczyk, Houston & Newsome, 1999).

In addition to prosody, neonates are sensitive to phonetic details that help them

develop their native-language sound system. This system of phonemes is essential to word-learning because it makes up the most basic components of words. Generally, before 6 months, infants can discriminate between native and non-native speech contrasts with equal ease and irrespective of native-language input. Over the next several months, perceptual sensitivity to non-native contrasts gradually declines while that to native contrasts increases (Kuhl et al., 1992; Kuhl et al., 2006; Werker & Tees, 1984). In fact, better native contrast discrimination at 7 months predicts accelerated vocabulary growth in the second year of life, whereas better non-native contrast discrimination is linked to poorer outcomes (Kuhl et al., 2005). By 12 months infants have developed a mature representation of their native-language phonemes (Werker & Tees, 1984). During early word-learning, infants' sensitivity to native phonetic details aids them in learning similar-sounding lexical items (Yoshida et al., 2009). During later word-learning, infants use phonetic detail to bolster their representation of words. At 19 months they can detect mispronunciations and show a graded sensitivity to them, suggesting that they have robust, adult-like representations of familiar words (White & Morgan, 2008). Like adults, these toddlers can ignore slight mispronunciations of familiar objects, understanding the intended meanings of such words rather than treating them as novel words. However, when the severity of the mispronunciation is sufficiently high, 19-month-olds are much less likely to attend to the intended target object, and instead focus their attention on an unfamiliar object. By 25 months the speed and accuracy with which an infant orients to a target during online speech processing predict vocabulary growth: those who are quicker and make fewer errors score significantly higher on the MacArthur-Bates Communicative Development Inventories (MCDI) (Fernald et al., 2006).

Infants also show an early sensitivity to the distributional probabilities of the sound sequencing patterns of their native language. By 9 months infants prefer to listen to words that follow phonotactic patterns that are legal over those that are illegal in their native language (Jusczyk et al., 1993). Additionally, at this age they prefer to listen to words with high- over words with low-probability phonotactic patterns (Jusczyk, Luce & Charles-Luce, 1994). Knowing what sound combinations are legal in one's native language can be of great use when trying to learn new words and can eventually help bootstrap learning to read. This is expected because phonological awareness (i.e., knowledge of native-language syllables and their onsets and rimes), has been implicated in enhanced reading acquisition (Stahl & Murray, 1994).

Early sensitivities to prosody, phonetic contrasts, phonotactic patterns, and mispronunciations are essential to speech recognition and segmentation as well as to word-learning. They facilitate language acquisition by helping the language learner hone into the characteristics that are specific and linguistically meaningful to their native language.

1.1.2 The Case for Linguistic Delay in Children Born Preterm

It remains unclear whether preterm infants are born with the same sensitivities to speech and language as their full-term counterparts. Lack of these sensitivities would be unsurprising given that these infants' earliest experiences with the auditory world are fundamentally different from those of typically developing full-term infants.

As treatment of preterm infants has become more sophisticated, it is no longer uncommon for infants born around 24 weeks gestational age to survive. In fact, the

survival rate for this group of infants has been estimated to be about 39% (<http://www.preemiesurvival.org/info/>). To put things in a developmental prospective, the cochlea is not fully developed and functional until about 26 weeks gestational age. Thus, many preterm infants are born either before or right around the time their hearing becomes functional. Under normal circumstances, once it starts hearing, a fetus predominantly experiences the low-frequency components of external sounds. Thus, the uterus acts like a filter, significantly attenuating high-pitched, arrhythmic sounds and protecting the fetus against auditory overstimulation.

In the case of speech, prosodic information is carried in its low-frequency components, providing the fetus with the first source of input by which it can start learning its native language. This low-pass filtered form of speech initiates and facilitates learning because it is a lot less complex than full-spectrum speech. In fact, prenatal exposure to low-pass filtered speech is enough for neonates to discriminate their native language from languages belonging to rhythmically different classes. However, unlike full-term infants, preterm infants are not exposed to low-pass filtered speech. Instead, they are bombarded with an array of full-spectrum sounds they would normally not experience during a critical period of development. In other words, they are deprived of the simplified and structured language input heard by typically developing full-term infants. To make matters worse, the sensory systems of preterm neonates are even less developed than those of their full-term counterparts. As a result, these systems are unlikely capable of processing and making sense of the input they are receiving. Together, these factors likely put preterm infants at risk for atypical developmental and linguistic outcomes.

In non-language domains, it has been suggested that premature and excessive exposure to broad spectra stimuli may disrupt and be detrimental to developmental processes (Turkewitz & Kenny, 1982, see also hypotheses of “Less is More”, Newport, 1990, and “Starting Small”, Elman, 1993). Such disruptions can reorganize the trajectory of development and subsequently impact learning by directing what the organism pays attention to in its environment. For example, research with rat pups shows that premature visual stimulation impacts the development of homing behavior (Kenny & Turkewitz, 1986). Under normal circumstances, rat pups are born with their eyes closed and early orientation to the home nest is strictly guided by olfactory cues; homing behavior gradually increases until the pups’ eyes open, at which point the behavior decreases. However, prematurely stimulating rat pups’ visual system by opening their eyes shortly after birth disrupts the trajectory of homing behavior: instead of declining, homing continues to increase much later than typically expected. According to Kenny and Turkewitz, providing premature visual stimulation affects the development of homing behavior by altering the cues the rat pups can attend to when finding their nest. Having multiple cues to guide homing can be detrimental, as it can create unnecessary competition between cues for attentional resources. Thus, the timing, type, and amount of stimuli received by an organism are critical in defining its developmental trajectories and subsequent outcomes. Yet, these are obviously all disrupted for preterm infants from the very beginning.

Preterm infants end up spending weeks and even months in neonatal intensive care units (NICU), exposed to a full range of sound frequencies and volumes as well as continuous and acute noises. Simultaneously, they are receiving premature input from all

other sensory systems, creating the perfect opportunity for sensory competition, confusion, and overstimulation. Thus, instead of being able to focus on the low-frequency characteristics of speech, preterm infants get a full spectrum of linguistic input. Such atypical stimulation may be overwhelming to an underdeveloped auditory system, thereby altering and/or slowing its course of development (Fouchée & Lickliter, 2002). These ideas give rise to the More is Less Hypothesis (MiLH), which stipulates that excessive and mistimed multisensory stimulation interferes with the organization of developmental processes.

The general expectation is that greater levels of prematurity result in more opportunities to disrupt developmental trajectories and outcomes. Work by Foster-Cohen and colleagues (2007) provides evidence for this idea in the domain of language. These researchers have shown that there is a linear relationship between gestational age and word and sentence comprehension and production at 2 years: toddlers born extremely premature, under 28 weeks, scored significantly lower on the MCDI than those born less premature, between 28-32 weeks. Both groups performed significantly worse than children who were born full-term, between 38-40 weeks. Moreover, preterm 3.5-year-olds born by 33 weeks gestational age not only had significantly shorter mean length utterances (MLU) than their age-matched full-term counterparts, but also produced more morphological errors and omissions in word-repetition tasks, and scored lower on phonological working memory measures (Sansavini et al., 2007). Sansavini and colleagues also found that preterm children scored lower on IQ tests than controls. Furthermore, Le Normand & Cohen (1999) showed that, relative to full-terms, preterm 3.5- and 5-year-olds' emergence of verb type and token usage is delayed and decreased.

Together, these findings suggest that time spent *in utero* potentially modulates language outcomes in early childhood.

Linguistic and cognitive deficits among preterm infants continue to be evident throughout the elementary school years. At 6 years children born very preterm are 3-5 times more likely to have problems with speech articulation and pre-reading skills compared to their age-matched full-term peers (Wolke & Meyer, 1999), while at 8 years they not only show decreased performance on semantic tasks, but also lack of activation in cortical areas typically associated with semantic processing (Peterson et al., 2002). Differences between preterm and full-term infants may even persist through adolescence: preterm and full-term 12-year-olds show differential activation in the neural pathways used to complete lexical semantic tasks (Schafer et al., 2009). Though no significant differences were found between these groups in response time (RT) and accuracy measures nor in activation of classic semantic processing areas (i.e., left superior and middle temporal gyri and inferior frontal gyrus), preterm subjects exhibited stronger connectivity between language-related and sensory-motor cortical areas as well as less activation of correlated areas within the frontal lobes. Despite showing no significant semantic differences at the behavioral level, preterm subjects did have significantly lower scores on receptive and expressive verbal tasks as well as on full-scale IQ measures relative to their full-term counterparts. Furthermore, subjects born prematurely were shown to have significantly different brain volumes. More importantly, these linguistic and cognitive deficits appear to be rooted in biological factors, such as sex, level of disability, and weight, rather than social factors, such as socioeconomic status (Martson et al., 2007; Wolke & Meyer, 1999).

Though it is undeniable that premature birth has a negative impact on language outcomes, it is unclear what is at the root of the shortcomings observed in word acquisition, production and comprehension. A major gap exists in the literature: most research has focused on language and language-related outcomes in preschool and school-aged children, once word-learning is well underway. Little is known about what is going on in infancy that might indicate future language delays. One exception is provided by Herold and colleagues (2008) and suggests that the effects of premature birth on vocabulary development could lie in infants' early sensitivities to native-language prosody. Evidence suggests that preterm infants born on average at 28 weeks gestational age show perceptual deficits to native-language lexical stress patterns in early infancy: German-acquiring infants born prematurely cannot distinguish between trochaic and iambic stress as late as 6 months corrected age (i.e., age after the expected delivery date) even though full-term infants show a preference for trochaic stress, the predominant stress pattern in German, by 4 months.

Given premature German-learning infants' early difficulties with word stress, this project seeks to identify other receptive language precursors to word-learning that may be impaired in premature infants. Of particular interest is characterizing the developmental patterns of these precursors, including sensitivities to word stress and phonotactic patterns in English-learning preterm infants, as well as abilities to learn similar-sounding words, detect word mispronunciations, and speech processing efficiency. Critically, unlike other studies, subjects participating in this project were tracked longitudinally on skills representative of both the earliest and later stages of language acquisition. Additionally, this project seeks to establish a link between these receptive language

abilities and word production and comprehension outcomes during the second and third years of life. Ultimately, the goal of this exploratory work is to establish a baseline for behavioral differences between preterm and full-term infants that could provide clinicians with insight on detecting those infants who are at risk of developing language delays and impairments in childhood. Early diagnosis as well as a better understanding of the root cause of specific language delays and impairments is essential for early intervention strategies and treatment. After all, it is much more difficult to intervene on a stable, fully developed system than on one that is still developing.

1.2 Project Details

1.2.1 Goals

The primary aim of this dissertation project is to investigate whether receptive language skills, particularly sound-structure sensitivities and word form representations, can be used to predict later language outcomes across full-term and preterm groups. Specifically, this project was designed to capture when and how these two groups of interest differ in their attention and sensitivities to native-language input, and subsequently assess whether these differences can account for the observed language outcomes. These issues are addressed in three parts.

The first part of this body of work reviews a series of longitudinal experiments investigating differences in the development of receptive language skills and cognitive skills in full-term and preterm infants in the first and second years of life. Chapter 2 focuses on sensitivities to native-language sound structure, while Chapter 3 centers around representations of native-language word forms. Chapter 4 discusses differences in

cognitive skill outcome measures and the potential mediating roles they play in language development.

The second part of this dissertation project provides an overview of a number of language outcome measures collected for both full-term and preterm participants during their second and third years of life. Chapter 5 outlines and discusses results from the three assessment tools used to evaluate both comprehension and production during the early and later stages of word learning. The reviewed outcomes are based on both parental surveys and lab-administered assessments. Chapter 5 also provides a summary of correlational relationships between the various outcome measures collected.

The third part of this dissertation work explores relations between the measures of receptive language abilities and cognitive skill collected in the first and second years of life and language outcome measures collected during the second and third years of life. These relations are assessed in Chapter 6 using structural equation modeling (SEM). Relations between receptive language skills and language outcomes are also explored using correlational analyses.

1.2.2 Hypotheses and Outline of Receptive Language Experiments and Collection of Language Outcome Data

The general hypothesis tested throughout this project is that preterm infants and toddlers who are at risk for language delays are likely to show different patterns of receptive language skills than those observed in typically developing full-term children. This is because preterm infants' initial experience with language is fundamentally different from that of their full-term counterparts. Unlike full-term newborns, preterm

infants are deprived of months of exposure to low-pass filtered speech. Instead, they are confronted with full-spectrum speech significantly earlier in development, and their perceptual systems have to figure out how to make sense of this input despite being physically underdeveloped,

During a typical pregnancy, the uterus and amniotic fluid act as low-pass filters, allowing only the low frequency components of sound to reach fetal ears. In the case of speech, these components are limited to its rhythmic and intonational qualities (i.e., prosody). Consequently, the uterine environment may help the developing fetus focus on certain qualities of native-language speech by restricting what the auditory system is exposed to. Thus, at birth, full-term infants are born with knowledge of their language's prosody, and this knowledge may provide a framework that is crucial for further acquisition of that language's sound system.

Given that preterm newborns lack prenatal familiarization with native-language prosody, they may be disadvantaged in further language learning. Though little research has been done to support this idea, one recent study (Herold et al., 2008) does report that 4-month-old age corrected German-learning preterm infants, but not age matched full-terms, are unable to recognize their native language's predominant lexical stress pattern over other stress patterns common to their language. Furthermore, this inability persists through the sixth month. These data suggest that preterm infants not only start off behind in their language processing skills, but continue to stay behind throughout the first year of life. This initial lag is likely to cause cascading negative effects for subsequent language acquisition, specifically for word-learning, which requires the learner to successfully segment words from fluent speech. Figuring out where word boundaries occur in speech

is contingent on the acquisition of a language's phonological structure, which in turn is facilitated by perceptual sensitivities that form the basis for early receptive language skills, including knowledge about lexical stress and phonotactic patterns.

To test the hypothesis that preterm infants have atypical patterns of receptive language skills, the present longitudinal project examined three aspects of receptive language abilities that have been shown to play major roles in word-learning in full-term children and are likely to show deficits in preterm infants. These three aspects included sensitivities to prosody, recognition of phonotactic patterns, and development of word-form representations.

Prosody, was assessed with one set of experiments. These evaluated whether preterm English-learning 4- and 6-month-olds are sensitive to different patterns of lexical stress in the same way as their full-term counterparts. Sensitivities to prosodic qualities, such as lexical stress, are important in helping the infant recognize the rhythmic patterns that identify her native language and its words. Prosody is essentially the first way in which the infant can distinguish her language from other languages and from other sounds. Given preterm infants' limited prenatal exposure to prosodic structure, prematurely born participants were expected to have difficulty discriminating between different patterns of native-language lexical stress.

The second aspect, recognition of native language phonotactic patterns at 6 and 9 months, was also assessed with one set of experiments. Phonotactics refers to restrictions on the combinations of phonemes permissible in a language; such constraints define the structure of syllables, consonant clusters and vowel sequences that occur in words. Being aware of what sound sequences are permissible aids the language learner to segment

words from fluent speech as well as to learn new words. Full-term, English-learning infants at 9 months, but not 6 months, can distinguish between words with high- and low-probability phonotactic patterns common to their native language (Jusczyk et al., 1994). Given preterm infants' tendency to show deficits in vocabulary growth, they were expected to show reduced phonotactic abilities at both 6 and 9 months.

The third aspect assessed by this work concerned the development of language-appropriate word-form representations. This portion of the project involved two experiments. The first experiment explored 19-month-olds' sensitivity to varying degrees of mispronunciations to familiar words. Age-matched full-terms show graded sensitivity to such mispronunciations (White & Morgan, 2008). At this later stage of word-learning toddlers can use their knowledge of familiar words to determine when a mispronunciation is linguistically meaningful (i.e., is it just an error or is it a new word?). Child participants who are at risk for language delays were expected to be more susceptible to confusion caused by mispronunciations. For example, preterm participants might be more likely to accept a mispronounced word as a novel word. The second experiment tested speed and efficiency of speech processing at 26 months and its relationship to vocabulary knowledge. Full-term toddlers show a strong correlation between speech processing abilities and vocabulary growth: those who orient to a target image more quickly and accurately when directed to do so are significantly more likely to show increased vocabulary growth as assessed by the MCDI (Fernald et al., 2006). Preterm infants participating in this project were expected to show little to no relationship between processing efficiency and vocabulary growth. Furthermore, they were expected to show slower and less efficient word-processing skills as well as reduced vocabulary scores

relative to full-term peers.

In addition to data on the receptive language skills listed above, this project collected outcome measures of comprehension and production for all participants. Language outcomes were evaluated using the MacArthur-Bate Communicative Development Inventories-Words and Gestures (WG-MCDI) at 14.5 months and the MacArthur-Bate Communicative Development Inventories-Words and Sentences (WS-MCDI) as well as the Mullen Receptive Language Subscale (MRLS) at 26 months. The WG-MCDI and WS-MCDI are parental reports, and each is divided into a number of sub-components measuring various aspects of child language including early communicative intent and complexity of language used. The WG-MCDI is primarily designed to measure language production and comprehension at the earlier stages of language acquisition, between 8 to 16 months, whereas, the WS-MCDI focuses on production at the later stages of language acquisition, between 16 to 30 months. The MRLS is a lab-administered standardized test that measures various skills related to language comprehension between infancy and early childhood. Such skills include command following, recognition of word-object pairings, understanding of words indicating spatial relationships, etc.

Besides providing novel information about the linguistic developmental trajectory of preterm children, this study investigated individual differences both within and across full-term and preterm groups. The intent was to discover early signs diagnostic of future language deficits. Determining what abilities best predict future outcomes was accomplished using correlational analysis and SEM. In essence, SEM is a statistical method for testing the validity of hypothesized relationships between theoretically

motivated latent constructs within a model. Each of these latent constructs is measured by a set of observed variables, or indicators, that are grouped conceptually and assume a common underlying cause. In practice, SEM is a combination of factor analysis and multiple regression that can handle multiple predictor and multiple outcome variables under a single model.

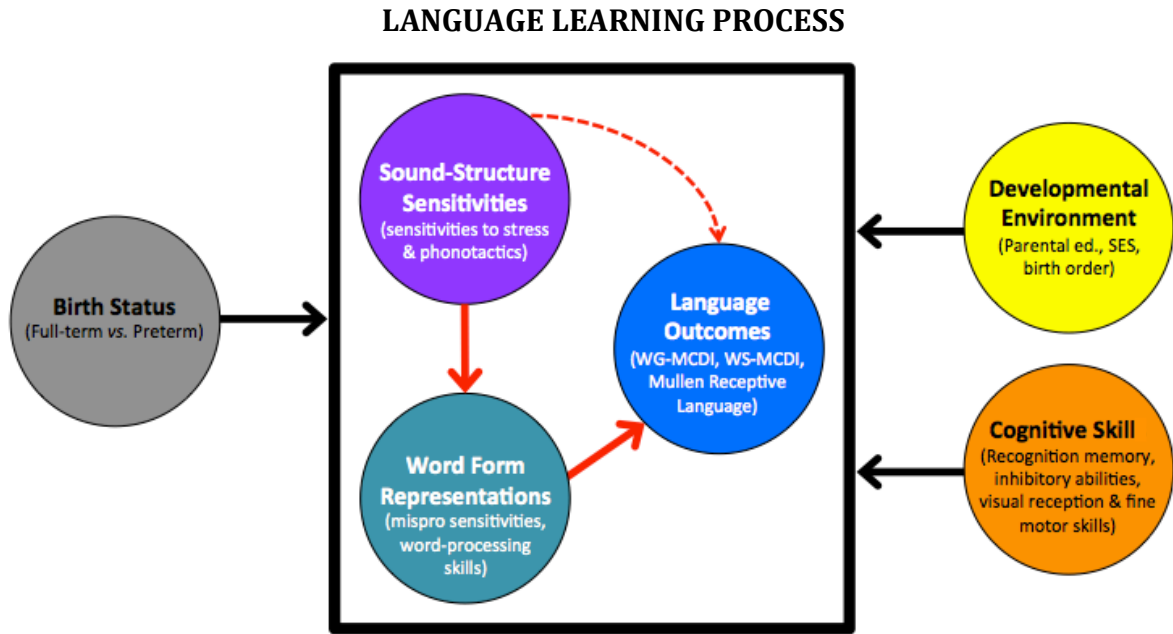
In this project SEM was used to test a possible model of language development. Language learning is a complex, multidimensional process, and to better understand it, it is critical to investigate how it unfolds from a number of different perspectives. This project addressed this issue by exploring how early sound-structure sensitivities, later word form representations, cognitive skills, birth status, and developmental environment relate to one another as well as to later language outcomes. These six factors represent the latent constructs comprising the language development model tested (see Figure 1). The indicators used to represent each of these constructs included: sensitivities to prosody and phonotactics for Sound-Structure Sensitivities; sensitivities to mispronunciations and speed and efficacy of word processing for Word Form Representations; visual recognition, visual reception, and fine motor skills for Cognitive Skill; full-term *versus* preterm birth for birth status; maternal education for developmental environment; and scores on the WG-MCDI, the WS-MCDI and the MRLS for Language Outcomes.

Besides playing an exploratory role characterizing language development in preterm infants, the experiments reviewed in this dissertation were aimed at investigating potential issues with receptive language skills that underlie difficulties in word-learning. If the testing methods used throughout this body of work successfully identify such issues then they could subsequently be used as diagnostic tools to assess children at risk for

poor outcomes. These methods would be ideal diagnostic tools for a number of reasons. First, they can be used to test children in early infancy as they only rely on looking behavior to assess response to auditory or visual stimuli. This is unlike other assessment tools presently used, which require responses that are much more complex (e.g., verbal responses), and, thus, cannot be used until the child is significantly older. Second, the methods used throughout this work are easily executed: the necessary equipment is few, relatively uncomplicated and inexpensive, and only requires the presence of single experimenter during testing. Third, testing sessions are short, lasting no more than 10 minutes. This is ideal for testing small children, who generally have short attention spans.

Additionally, the methods and study design used in the present project allowed for the construction of detailed individual profiles of participants' developmental trajectories. The advantages of this are two fold. First, having a collection of measures for each child can provide estimates of individual differences and their significance in shaping later outcomes. Second, the availability of many measures can better specify the nature of a particular language delay. Having a better understanding of what is causing a particular issue can, in turn, help clinicians develop more effective treatment interventions.

Figure 1. Language learning model examined in this study with SEM. Circles enclosed by the square represent latent constructs, with their corresponding indicators appearing in parentheses. In this model, Sound-Structure Sensitivities was tested as a predictor of Word Form Representations, which was subsequently tested as a predictor of Language Outcomes; the relationships between these constructs are marked by solid red arrows. The relationship between Sound-Structure Sensitivities and Language Outcomes was tested indirectly, as marked by the dashed red arrow. Circles appearing outside the square contain some of the indicators that were tested as model covariates.



1.2.3 Participants

The total test sample consisted of 30 preterm (19 females) and 59 full-term (26 females) control participants. Because it was difficult to recruit infants to start the project at 4 months, some participants were enrolled at 6 or 9 months. In the preterm group, three infants started participating at 6 months and three at 9 months. In the full-term group one infant enrolled at 6 months and three at 9 months. Initially, recruitment was limited to single birth infants. However, this placed an extra limitation on preterm recruitment, as

many multiple birth pregnancies result in preterm birth. Four sets of twins were subsequently recruited, two in the preterm group and two in the full-term group.

All infants were English-learning with little to no exposure to alternate languages. Preterm infants were born between January 2010 and June 2011, and were recruited from the Women and Infants Hospital Neonatal Follow-up Clinic in Providence, Rhode Island. Full-term infants were born between April 2010 and February 2011, and were recruited through the Metcalf Infant Lab Database at Brown University in Providence, Rhode Island. At birth, preterm participants were on average 9.8 weeks older and weighed 2221 g less than their full-term counterparts. Of the initial 30 infants, 9 were extremely preterm (EP, gestational age <28 weeks), 13 were very preterm (VP, gestational age between 28 to 32 weeks, and 8 were moderately preterm (MP, gestational age between 33 and 36 weeks); 2 were of normal birthweight (≥ 2500 g), 4 were low birthweight (< 2500 g), 14 were very low birthweight (VLBW, ≤ 1500 g), 2 were extremely low birthweight (ELBW, ≤ 1000 g), and 7 were micropremature (≤ 750 g). Preterm participants were always tested at corrected ages, meaning that their age at test was calculated based on their expected rather than their actual birth dates (e.g., an 11-month-old infant born 2 months premature is 9 months corrected age). The decision to test at corrected ages, rather than at chronological ages, was made because, from a maturational perspective, a preemie who is age-matched chronologically to a full-term infant is not as neurologically or physically developed as that full-term infant. In fact, when evaluating whether or not preterm infants have achieved developmental milestones, physicians continue to correct for age until age two years.

Previous work has shown that maternal education, a proxy for social economic status (SES), highly correlates with measures of language outcomes (Hart & Risley, 1995). To control for this, we collected information on parental education using an 8-point classifier scale (i.e., 1= < 7th grade, 2= between 7th to 9th grade, 3= between 10th to 11th, 4= GED, 5= up to 12th grade, 6= Partial college, 7= College and 8= Graduate School). Means for maternal education were 6.7 and 7.2 for the preterm (n= 24) and full-term (n= 50) groups, respectively. Means for paternal education were 6.5 and 6.6 for the preterm (n= 23) and full-term (n=49) groups, respectively. To assess overall neonatal health, we collected information on incidence of intraventricular hemorrhaging (IVH) and periventricular leukomalacia (PVL), as well as days spent in the neonatal intensive care unit (NICU) after birth. Demographic information is further specified on Table 1. Given the longitudinal nature of the project, all subjects were not able to participate in all studies at every age group. Thus, specific subject counts and relevant demographic information are also provided for each study.

Table 1. Summary of demographic information for full-term and preterm groups across all studies. GA= gestational age, BW= birth weight, IVH= intraventricular hemorrhaging, PVL= periventricular leukomalacia, NICU= neonatal intensive care unit, gr.= grade, and bi.= bilateral.

	Sex	GA (weeks)	BW (g)	Birth Order	Maternal Ed.	Paternal Ed.	Incidence of IVH	Incidence of PVL	Days in NICU
Full-term n= 59	26 females	Mean: 39.6 SD: ±1.15 Range: 37-42	Mean: 3520 SD: ±484.9 Range: 2296-4423	Mean: 2.0 SD: ±1.3 Range: 1-6	Mean: 7.2 SD: ±0.83 Range: 5-8	Mean: 6.6 SD: ±0.99 Range: 3-8	0	0	Mean: 0.033 SD: ±0.26 Range: 0-2
Preterm n= 30	19 females	Mean: 29.8 SD: ±3.73 Range: 23-36	Mean: 1299 SD: ±571.3 Range: 620-2750	Mean: 1.4 SD: ±0.61 Range: 1-3	Mean: 6.7 SD: ±1.6 Range: 3-8	Mean: 6.5 SD: ±1.5 Range: 3-8	6 (3 gr. I, 1 gr. II, 1 bi. gr. I & II, 1 bi. gr. I & IV)	1	Mean: 61.2 SD: ±43.0 Range: 2-127

Given the longitudinal nature of this project, each subject had the potential to yield numerous data points. A total of 12 possible data points were collected per subject: four measures of Sound-Structure Sensitivities (i.e., sensitivities to lexical stress at 4 and 6 months and sensitivities to phonotactics at 6 and 9 months), two measures of Word Form Representations (i.e., sensitivities to mispronunciations at 19 months and speed and accuracy of speech processing at 26 months), three Cognitive Skill measures (visual recognition memory at 12 months, and visual reception and fine motor skills at 26 months), and three language outcome measures (WG-MCDI at 14.5 months, WS-MCDI at 26 months and MRLS at 26 months). All 12 measures were successfully collected for 18 full-term and 7 preterm participants. Within the full-term group, 22 infants had all six receptive language measures, 13 had five measures, 13 had four measures, 7 had three measures, and 4 had less than three measures; 34 infants had all 3 cognitive measures, 5 had two measures, and 8 had one measure; 38 infants had all 3 language outcome measures, 4 had two measures, and 5 had one measure. Within the preterm group, 13 infants had all six receptive language measures, 2 had five measures, 5 had four measures, 4 had three measures, and 6 had less than three measures; 16 infants had all three cognitive measures, 1 had two measures, and 5 had one measure; 14 infants had all three language outcome measures, 4 had two measures, and 5 had one measure. In order to be included in SEM analyses, a subject had to have a minimum of one sound-structure sensitivity measure, one word form representation measure, one cognitive measure, and one language outcome measure. Based on these inclusion criteria, 43 full-term and 19 preterm participants were included in SEM analyses.

PART I

MEASURING PREDICTORS OF LANGUAGE OUTCOMES:

RECEPTIVE LANGUAGE SKILLS AND COGNITIVE ABILITIES

IN THE FIRST AND SECOND YEARS OF LIFE

Chapter 2: Sensitivities to Native-Language Sound Structures

2.1 Sensitivities to Lexical Stress

As already reviewed, knowledge of native-language prosody is one of the most fundamental stepping-stones for recognizing and learning one's native language. Thus, impaired or atypical processing of prosodic cues is likely to have cascading negative effects on subsequent word-segmentation from fluent speech and vocabulary growth. One of the earliest cues infants are sensitive to is lexical stress. It is well established that when presented with lists of words with trochaic and iambic stress patterns, full-term, typically developing, English-learning 9-month-olds prefer to listen to words with trochaic stress than to those with iambic stress (Jusczyk et al. 1993). However, recent research with full-term, German-learning 4- and 6-month-olds shows that preference for the predominant native-language stress pattern can be elicited as early as 4 months if infants are presented with one rather than multiple lexical units (Herold et al., 2008). The present set of experiments replicated Herold et al.'s work with two goals in mind. The first was to potentially elicit preference for trochaic stress before 9 months in full-term, English-learning infants. The second was to examine potential differences in the patterns of sensitivity preterm and full-term infants would show to lexical stress at 4 and 6 months.

2.1.1 *Participants*

Participants were tested longitudinally at 4 and 6 months corrected age. At 4 months, 23 preterm (14 females, mean corrected age= 126.6 days, range= 111-146 days) and 49 full-term (23 females, mean age= 118.9 days, range= 101-138 days) infants were

tested. An additional eight full-term infants were tested but discarded due to fussiness. Chronologically, preterm infants were 80.4 days older than their preterm counterparts (mean chronological age= 199.3 days, range= 175-239 days). At 6 months, 21 preterm (13 females, mean corrected age= 186.2 days, range= 174-203 days) and 49 full-term (23 females, mean age= 184.9 days, range= 169-197 days) infants were tested. An additional three infants, two preterm and one full-term, were tested but discarded due to fussiness and technical difficulties, respectively. Chronologically, preterm infants were 66.8 days older than their preterm counterparts (mean chronological age= 251.7 days, range= 212-302 days). Out of the 23 preterm infants tested at 4 months, 17 were retested at 6 months; 4 new infants were recruited and tested at 6 months. Out of the 49 full-term infants tested at 4 months, 42 were retested at 6 months; 7 new infants were recruited and tested at 6 months.

2.1.2 Method

Infants were tested in a sound-attenuated laboratory room using the video Head-Turn Preference Procedure (v-HPP), which is a modified version of the Head-Turn Preference Procedure (HPP). This method was selected because it minimizes the need for head turning, which at younger ages is a limited skill, as head control is not yet been fully mastered. In the traditional HPP set-up, lights are placed to the left and right of the participant; at the beginning of each trial, one of these lights turns on and auditory test stimuli can only start playing when the infant orients toward it. v-HPP replaces these lateral lights with a set of adjacent monitors that instead face the infant. Like the lights, these serve to get the infant's attention at the beginning of each trial. v-HPP facilitates

testing because in order to initiate playing of the stimuli all the infant has to do is move her eyes (rather than her head) to the appropriate screen.

During testing, the participant sat on her caregiver's lap, approximately 90 cm away from the monitors. The caregiver listened to instrumental music over noise-cancelling headphones in order to mask the stimuli and discourage interactions that might sidetrack the participant during testing. The experimenter ran the test session from a control room, monitoring and coding online infant looking behavior over a closed-circuit video system.

Before each trial, a light centered between the monitors was flashed to capture the infant's attention. Once the infant oriented toward the midline, the light was extinguished and a blinking orange ball appeared on one of the monitors. When the infant looked at the appropriate monitor, a static checkerboard replaced the ball and stimulus words began to play. Trial length was infant-controlled, with words playing every half-second for a specified maximum duration. At 4 months maximum trial duration was set at 20 seconds; at 6 months it was set at 30 seconds. Minimum trial duration was one second at both ages. Trials not meeting this criterion were repeated. A trial was automatically terminated if the infant looked away from the monitor for two or more consecutive seconds, at which point a new trial was called. For each trial, looking time to the monitor was recorded as a proxy for attention to the auditory stimuli. Stimuli were played at conversational level (approximately 70 dB).

2.1.2.1 Stimuli

Throughout the test session participants heard one nonce, disyllabic CVCV word presented half of the time with trochaic stress and half of the time with iambic stress. To prevent any priming effects, different target words were used across age groups. At 4 months, the trochaic and iambic exemplars were “**G**Aba” and “ga**BA**”, respectively; at 6 months, they were “**G**Obo” and “go**BO**”, respectively. A female native English speaker recorded multiple tokens of each target word in isolation. At test, infants heard 15 unique tokens of each target word, “**G**Aba” and “ga**BA**” at 4 months, and “**G**Obo” and “go**BO**” at 6 months.

Target words were presented over five blocks of trials, with each block consisting of one trochaic trial and one iambic trial for a total of ten trials. Trial type presentation was counterbalanced across sides. At 4 months, infants were given two 30-second familiarization trials prior to testing. Unlike test trials, these trials were not infant-controlled and consisted of repetitions of the trochaic target word. Familiarization trials were not used at 6 months.

2.1.2.2 Measures of Interest

For each block of trials, looking time to iambic stimuli was subtracted from looking time to trochaic stimuli to determine infant preference for stress type. A positive difference score indicated that the infant preferred to listen to trochaic stimuli, whereas a negative score indicated preference for iambic stimuli. Difference scores were then averaged for blocks two through four, and compared across full-term and preterm groups to determine overall preference. Difference scores for the first block were not averaged

into the mean because the two trials in this block were used to get participants acquainted with the test set-up. The final block was excluded from analysis because interest in the stimuli had decreased by the time participants got to the last two trials. This approach to data analysis has been used to keep stimulus effects from being washed out by unfamiliarity with behavioral expectations at the beginning of the experiment and boredom at the end (see Herold et al., 2008 for an example of the former). Additionally, habituation data have previously show that, during a test session, infants will reverse their looking preference patterns over time.

2.1.3 Results

Full-term and preterm infants showed opposite patterns of sensitivity to lexical stress at both 4 and 6 months: full-term infants tended to prefer to listen to trochaic word tokens, while preterm infants tended to prefer to listen to iambic word tokens (see Figure 2). At 4 months, the mean difference scores in looking time were +0.382 (SD= ± 3.52) and -0.595 (SD= ± 2.84) for full-term and preterm infants, respectively. At 6 months, the mean difference scores were +0.441 (SD= ± 2.54) and -0.977 (SD= ± 2.77) for full-term and preterm infants, respectively. However, a 2 (mean difference scores) by 2 (birth status), repeated measures ANOVA revealed that the main effect of stress preference and the interaction between birth status and stress preference did not reach statistical significance, $F(1, 81) = .027, p = .869$, and $F(1, 81) = .019, p = .891$, respectively^{1,2}.

¹ Despite the large sample size differences between full-term and preterm groups, ANOVAs were used to compare performance between these groups in the numerous experiments discussed throughout this manuscript. Additional analyses showed that variances across groups were not statistically different, thus justifying ANOVA use.

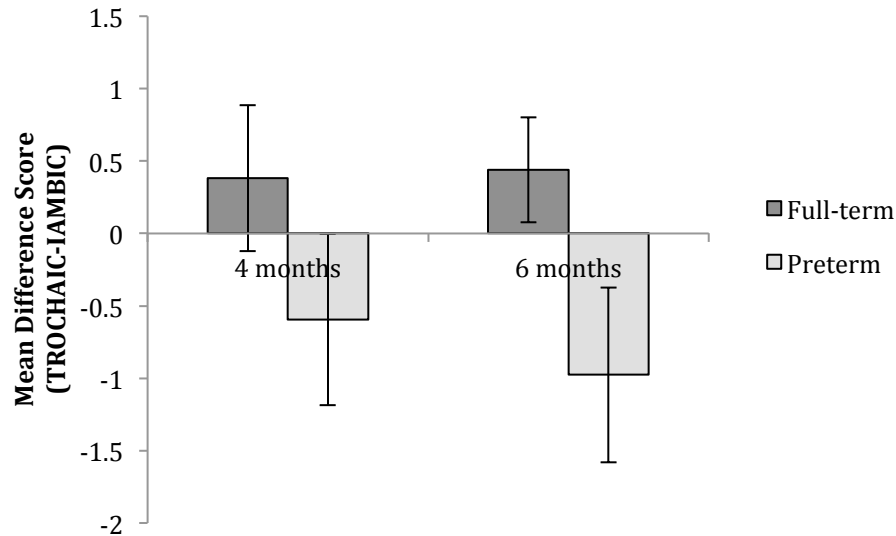
Planned comparisons between group means within each age group were conducted using independent samples t-tests³. Results showed that at 4 months, mean difference scores in looking time were not significantly different across full-term and preterm participants, $t(70) = 1.16, p = .249$; mean difference scores at 6 months were significantly different across groups, $t(68) = 2.08, p = .041$. Additional analyses using one-sample t-tests were conducted to explore whether mean difference scores at each time point for each group were significantly different from zero. For full-term participants, mean difference scores were not significantly different from zero at either 4 or 6 months, $t(48) = .758, p = .452$, and $t(48) = 1.21, p = .231$, respectively. For preterm participants, mean difference scores were also not significantly different from zero at either 4 or 6 months, $t(22) = -1.01, p = .326$, and $t(20) = -1.62, p = .121$, respectively.

Taken together, these analyses suggest that while neither full-term nor preterm infants showed a statistically significant preference for either trochaic or iambic stress at 4 or 6 months, the patterns of preference emerging across these two groups were distinctly different. While full-term infants tended to prefer words with trochaic stress, preterm infants tended to prefer words with iambic stress.

² To maximize use of collected data points, missing mean difference scores for a given participant were imputed using t-scores computed for that participant's mean difference scores for the 4-, 6- and 9-month test sessions. The mean of these t-scores was used to estimate the missing mean difference score, X , using the following equation: $X = (t_{\text{mean}}(SD_{\text{group}}/\sqrt{n})) + M_{\text{group}}$, where t_{mean} is the subject's mean t-score, SD_{group} is the standard deviation for the group's mean difference score, n is the number of subjects in the group, and M is the group's mean difference score. At 4 months, seven and four data points were imputed for full-term and preterm participants, respectively. At 6 months, seven and six data points were imputed for full-term and preterm participants, respectively.

³ Imputed data points were not included in t-test analyses.

Figure 2. Mean difference scores and standard errors measuring stress preferences for full-term (dark grey bars) and preterm (light grey bars) infants at 4 and 6 months.



2.1.4 Discussion

The purpose of these experiments was to explore whether preterm infants show the same pattern of sensitivities to native-language lexical stress as their full-term counterparts. Though this study did not yield statistically significant preferences for trochaic or iambic stress for either full-term or preterm infants, it did suggest the emergence of different strategies for early language processing between these two groups of participants. Like Jusczyk et al.'s 9-month-olds (1993) and Herold et al.'s 6-month-olds (2008), full-term infants in this study showed a preference trend for word tokens with trochaic stress, which is the dominant stress pattern in the language these infants were being exposed to, English. Given that this trend was evident as early as 4 months, the present study corroborates the notion that the development of sensitivity to native-

language stress emerges early in language acquisition. Furthermore, if a perceptual preference for the native language's dominant stress pattern is a proxy for typical language development, then preterm infants' preference trend for iambic stress in this study implies that this group is not sensitive to native-language stress cues in the same way full-term infants are.

The present results are unlike Herold et al.'s (2008) findings, which showed that preterm infants were delayed compared to their full-term counterparts as they could not distinguish between trochaic and iambic stress patterns as late as 6 months corrected age. Instead, preterm infants in this study evidenced some ability to discriminate between trochaic and iambic stress, despite, their pattern of sensitivity being opposite to that of their full-term peers. That said, differences between these two studies should be interpreted with a grain of salt, as the preterm sample tested in the present study was significantly more heterogeneous than that tested by Herold and colleagues. All infants participating in the latter had gestational ages between 26 and 29 weeks and weighed less than 1,500 g at birth. In contrast, preemies participating in the present experiments had more variable gestational ages and birthweights, ranging from 24 to 36 weeks and 620 to 2,750 g, respectively.

Furthermore, the magnitude of the effects observed for full-term participants in the present study were not nearly as large as those seen by Herold and colleagues. In the latter study, full-term infants as young as 4 months showed a significant preference for words with trochaic stress. In contrast, neither 4- or 6-month-old full-term participants in the current study showed a statistically significant preference for trochaic stress- they only showed a trend towards a preference for trochees. What might explain this

discrepancy in findings? Herold et al.'s infants were German-learning, while infants in this study were English-learning. Perhaps the behavioral differences observed across these studies can be attributed to cross-linguistic differences in the extent to which lexical stress is weighed as a cue to word segmentation. That is, something in the German input might make lexical stress more perceptually salient to the infant learner being exposed to this language. Of course, this is just speculation, and additional work is needed to test this hypothesis.

In short, this study points to early differences in perceptual sensitivity to lexical stress between full-term and preterm infants. If preterm infants are paying attention to their native language differently, then what they learn from linguistic input is expected to be different from what their full-term counterparts learn. This could subsequently have negative cascading effects on typical word-learning strategies. For example, an early preference for the less common lexical stress pattern could bias word segmentation when listening to fluent speech: if lexical stress cues word boundaries, then focus on iambic stress could create a mismatch between perceived and actual word boundaries. Thus, preterm children would learn improper word forms.

2.2 Sensitivities to Phonotactics

In addition to lexical stress, full-term infants are sensitive to native-language phonotactics (i.e., permissible patterns of sound combinations). Seminal experiments show that, when presented with nonce words with English phonotactics, full-term, English-learning 9-month-olds prefer to listen to words with high sound sequence probabilities than to words with low sound sequence probabilities. Critically, infants do

not show a preference for either type of word at 6 months, suggesting that experience with the native-language modulates learning of patterns of sound structures (Jusczyk et al., 1994). That said, Jusczyk et al.'s experimental paradigm was atypically complex for testing 6-month-olds: infants were tested with HPP on a total of 240 unique (120 high and 120 low probability) words over 16 preference trials. Testing full-term infants on a simpler task could reveal that they are sensitive to phonotactic probabilities as early as 6 months. As already noted, previous work shows that, when presented with a simplified stimulus set, infants as young as 4 months can distinguish between trochaic and iambic stress patterns (Herold et al., 2008), despite seminal not attributing this skill to infants until 9 months of age. Thus, it remains unclear how early typically developing infants are sensitive to native-language phonotactics. The first goal of this study was to investigate whether sensitivities to patterns of high and low phonotactic probabilities could be elicited before 9 months of age. This was accomplished by simplifying the task used by Jusczyk and colleagues (1994): the number of unique words and preference trials heard by participants was significantly reduced. Establishing a benchmark for when sensitivities to phonotactics develop in full-term infants is important for assessing whether preterm infants are following typical developmental trajectories for receptive language abilities.

The second goal of this study was to explore potential differences in how full-term and preterm infants process high and low patterns of phonotactic probabilities. Like lexical stress, phonotactic information can provide cues to word boundaries, which can facilitate word segmentation in fluent speech. For example, some sound-sequence combinations occur across word boundaries and rarely within words, and other sequences

only occur word initially (e.g., “*str-*” in English), while others are limited to medial or final positions (e.g., “*-rf*” in English). Atypical patterns of sensitivity to native-language sound combinations could mediate poor word-learning via effects on word segmentation abilities. Previous work with very preterm (≤ 32 weeks gestational age) monolingual Spanish- and Catalan-learning 8-month-olds provides evidence suggesting these infants have difficulty segmenting familiar words from fluent speech (Bosch, 2011). Specifically, when familiarized with passages containing multiple repetitions of two target words very preterm infants did not show a listening preference for those words when tested on lists of words containing familiar and novel words. This suggests that very preterm infants could not segment and learn the target words they were familiarized with. Importantly, measures collected during the segmentation task were positively correlated with measures of expressive vocabulary at 12 and 18 months. Specifically, greater looking time during the last familiarization trial was linked to greater word production at 12 months, and greater looking time to novel words during test trials was associated with greater word production at both 12 and 18 months. The present study hoped to expand on these findings by investigating sensitivity to phonotactic probability patterns across full-term and preterm groups and subsequently exploring how these relate to later vocabulary measures. If knowledge about native-language phonotactics bolsters word-learning by facilitating word segmentation, then correlations between measures of sensitivity to phonotactics and later language outcomes are expected.

2.2.1 *Participants*

Participants were tested longitudinally at 6 and 9 months corrected age. At 6 months, 21 preterm (12 females, mean corrected age= 186.2 days, range= 174-203 days) and 48 full-term (22 females, mean age= 184.9 days, range= 169-197 days) infants were tested. An additional three infants, two preterm and one full-term, were tested and discarded due to fussiness (n= 1 preterm), experimenter error (n= 1 preterm), or technical difficulties (n= 1 full-term). Chronologically, preterm infants were 66.8 days older than their preterm counterparts (mean chronological age= 251.7 days, range= 212-302 days). At 9 months, 25 preterm (15 females, mean corrected age= 278.3 days, range= 267-292 days) and 49 full-term (22 females, mean age= 276.4 days, range= 258-290 days) infants were tested. Chronologically, preterm infants were 74.4 days older than their preterm counterparts (mean chronological age= 350.8 days, range= 306-405 days). Out of the 21 preterm infants tested at 6 months, 19 were retested at 9 months; 6 new infants were recruited and tested at 9 months. Out of the 48 full-term infants tested at 6 months, 41 were retested at 9 months; 8 new infants were recruited and tested at 9 months.

2.2.2 *Method*

Infants were tested using v-HPP. For set-up details, refer to the *Method* section under *Sensitivities to Lexical Stress*. No familiarization trials were used prior to testing.

2.2.2.1 *Stimuli*

Participants heard 36 unique nonce monosyllabic CVC words, 18 with high and 18 with low phonotactic probabilities conferring to American English. At each visit,

infants were presented with one of two unique word-lists. List presentation was counterbalanced across visits such that all infants heard both List 1 and List 2. Target words were presented over six blocks, with each block consisting of one high and one low phonotactic probability trial for a total of 12 trials. Each trial consisted of six unique words that were repeated throughout the duration of the trial. Trial length was infant-controlled, with one target word presented every half-second for a maximum of 20 seconds or until the infant looked away from the monitor for two or more seconds. Minimum trial length was set at two seconds. Trial type presentation was counterbalanced across sides.

Once target words were generated and selected using the Phonotactic Probability Calculator (Vitevitch & Luce, 2004). This program calculates the sums of segment-to-segment and biphone co-occurrence probabilities to determine word phonotactic probabilities. For List 1, the means for the sums of segment-to-segment and biphone probabilities were 0.1535 and 0.0045 and 0.0878 and 0.0026 for high and low probability words, respectively; for List 2, these values were 0.1571 and 0.0044 and 0.0884 and 0.0025, respectively (see Tables 2a and 2b for selected target words and their corresponding statistics). Occurrence of vowel type was controlled across lists of high and low probability words to ensure that any detected sensitivities to phonotactics were not the product of vowels specifically occurring in each type of list. Six vowels were selected as nuclei, including æ, i, e, ^, ai, and R. Each of these vowels occurred in three high and three low probability words in both List 1 and List 2. Where possible, consonant occurrence within List 1 and List 2 was also equated across high and low probability words. This was significantly more challenging to do than controlling for vowel types, as

the probability of consonant occurrence inherently contributes to determining a language's phonotactic structure. Once selected, target words were recorded by a female native-English speaker.

Table 2. (a) List 1 and (b) List 2 of nonce words used to test infants at 6 and 9 months. Darker shades of purple and green highlight high probability stimuli and lighter shades of purple and green highlight low probability stimuli. Words are listed according to their vocalic nucleus (i.e., æ, i, e, ^, aɪ, and R). Sums for segment-to-segment and biphone co-occurrence probabilities are provided for each word. Means and standard deviations for each of these properties are also provided for each list.

(a) List 1

HIGH				LOW		
Vowel	Word	Σ Seg-to-Seg	Σ Biphone	Word	Σ Seg-to-Seg	Σ Biphone
æ	sæð	0.1849	0.0050	ðæg	0.1002	0.0032
	pæz	0.1839	0.0094	ʃæd	0.1271	0.0038
	tʃæɪ	0.1620	0.0107	tʃædʒ	0.0991	0.0029
i	gin	0.1539	0.0031	nih	0.0588	0.0021
	jɪn	0.1358	0.0032	θɪp	0.0757	0.0024
	pɪdʒ	0.1270	0.0028	tɪʒ	0.0780	0.0014
e	seʒ	0.1333	0.0019	jes	0.1159	0.0031
	keθ	0.1293	0.0022	ðetʃ	0.0401	0.0004
	hen	0.1647	0.0040	dʒes	0.1218	0.0034
^	s^ʃ	0.1493	0.0063	j^s	0.1259	0.0042
	f^t	0.1518	0.0034	l^θ	0.0807	0.0019
	p^m	0.1730	0.0075	tʃ^m	0.0975	0.0056
aɪ	bam	0.1816	0.0071	naɪz	0.0782	0.0024
	saitʃ	0.1447	0.0033	naɪp	0.0952	0.0029
	kaɪz	0.1471	0.0015	ʃaɪk	0.0975	0.0033
R	sRd	0.1651	0.0048	zRb	0.0533	0.0014
	ʃRn	0.1305	0.0022	nRs	0.0562	0.0005
	fRl	0.1450	0.0027	lRf	0.0785	0.0010
	Mean	0.1535	0.0045	Mean	0.0878	0.0026

(b) List 2

HIGH				LOW		
Vowel	Word	Σ Seg-to-Seg	Σ Biphone	Word	Σ Seg-to-Seg	Σ Biphone
æ	sædʒ	0.1926	0.0054	næð	0.1063	0.0023
	ðær	0.1607	0.0075	næʒ	0.1049	0.0021
	pæʒ	0.1655	0.0089	jæf	0.1070	0.0016
i	kit	0.1905	0.0029	zis	0.1132	0.0021
	hin	0.1673	0.0043	jid	0.0777	0.0024
	sif	0.1419	0.0030	nif	0.0753	0.0026
e	pem	0.1630	0.0042	hep	0.1057	0.0027
	kez	0.1420	0.0031	weθ	0.0569	0.0016
	sed	0.1696	0.0041	meʃ	0.0941	0.0033
^	w^l	0.1332	0.0052	v^b	0.0876	0.0043
	j^n	0.1432	0.0060	r^dʒ	0.1001	0.0031
	p^θ	0.1310	0.0025	t^ʃ	0.0914	0.0018
aɪ	saitʃ	0.1447	0.0033	ðark	0.0907	0.0031
	kaiθ	0.1344	0.0007	θaɪp	0.0782	0.0016
	baɪn	0.1816	0.0071	laɪtʃ	0.0764	0.0037
R	fRn	0.1674	0.0037	zRʒ	0.0290	0.0003
	mRl	0.1556	0.0021	θRk	0.0850	0.0036
	vRn	0.1432	0.0044	tʃRs	0.1124	0.0034
Mean		0.1571	0.0044	Mean	0.0884	0.0025

2.2.2.2 Measures of Interest

Preference for phonotactic probability type was calculated for each block of trials by subtracting looking time to the low stimuli trial from looking time to high stimuli trial. A positive difference score indicated a listening preference for words with high phonotactic probabilities, whereas a negative difference score indicated preference for words with low phonotactic probabilities. Difference scores were then averaged for blocks two through five, and compared across full-term and preterm groups to determine overall preference. Difference scores for the first block were not averaged into the mean

because the trials in this block were used to train infants on the experimental set-up. The final block was excluded because looking time during the last two trials had decreased significantly, indicating that participants were bored with the experiment. Thus, including the sixth block in mean difference score calculations would attenuate potential differences in sensitivity to phonotactics across full-term and preterm groups. This approach to data analysis was also undertaken when assessing infants' sensitivities to lexical stress (see *Measures of Interest* section for "Sensitivities to Lexical Stress" for a more detailed justification).

2.2.3 Results

Full-term and preterm infants showed opposite patterns of sensitivity to phonotactics at both 6 and 9 months: full-term infants tended to prefer to listen to low probability words, while preterm infants tended to prefer to listen to high probability words (see Figure 3). At 6 months, mean difference scores in looking time were -0.278 (SD= ± 3.04) and +0.324 (SD= ± 3.18) for full-term and preterm infants, respectively. At 9 months, mean difference scores were -0.825 (SD= ± 2.60) and +0.841 (SD= ± 2.63) for full-term and preterm infants, respectively. A 2 (mean difference scores) by 2 (birth status), repeated measures ANOVA yielded no significant main effect of phonotactic preference, $F(1, 81) = .472$, $p = .494$; critically, the interaction between mean difference scores and birth status was marginally significant, $F(1, 81) = 2.98$, $p = .088^4$.

⁴ Again, missing data points for a given subject were imputed using t-scores calculated for that participant's mean difference scores at the 4-, 6- and 9-month test sessions. The mean of these t-scores was computed and used, as previously described, to estimate the missing mean difference scores. At 6 months, eight and six data points were imputed for

Planned comparisons between group means within each age group were conducted using independent samples t-tests⁵. Results showed that at 6 months, mean difference scores in looking time were not significantly different across full-term and preterm participants, $t(67) = 1.57, p = .458$; mean difference scores at 9 months were significantly different across birth status groups, $t(72) = .019, p = .011$. Additional analyses using one-sample t-tests were conducted to explore whether mean difference scores at each time point for each group were significantly different from zero. At 6 months, mean difference scores for full-term and preterm infants were not significantly different from zero, $t(47) = -.635, p = .528$, and $t(20) = .466, p = .646$, respectively. At 9 months, the mean difference score was significantly different from zero for full-term infants, $t(48) = -2.22, p = .031$; for preterm infants this difference did not reach statistical significance, $t(24) = 1.60, p = .123$.

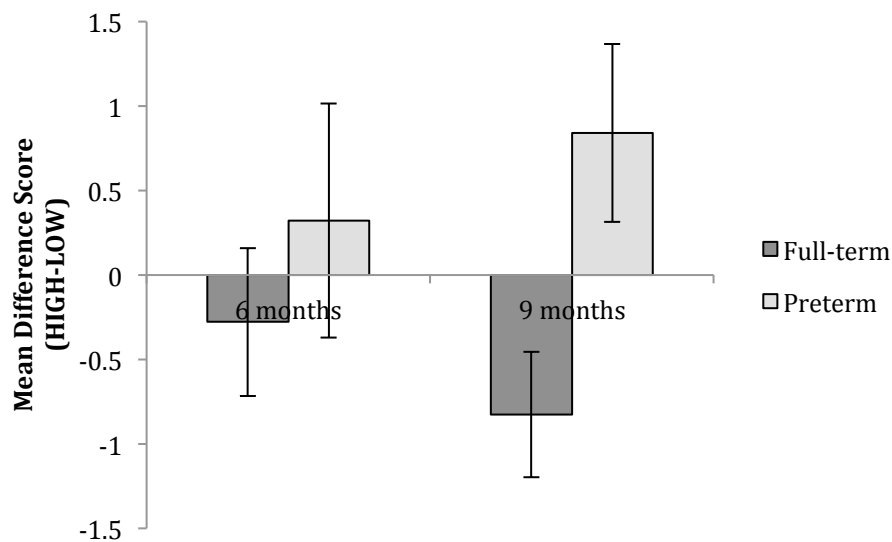
Taken together, these analyses suggest a couple of things. First, at 6 months, neither full-term nor preterm participants show a statistically significant preference for either high or low phonotactic probability words. That said, distinct and opposite trends in preference seem to be emerging within each group: full-term infants tended to prefer listening to low probability words, while preterm infants tended to prefer listening to high probability words. Second, by 9 months, full-term infants consistently preferred listening to low probability words. Though preterm infants continued to show the opposite preference pattern, their preference for high probability words was more variable than that of full-term infants' for low probability words (see Figure 3 error bars). Interestingly,

full-term and preterm participants, respectively. At 9 months, four and two data points were imputed for full-term and preterm participants, respectively.

⁵ Imputed data points were not included in t-test analyses.

variability within the preterm group at 9 months was unlikely attributable to weeks of gestational age, Spearman's $r = -.291, p = .158$. That is, less preterm infants were not more likely to behave like full-term infants by preferring to listen to low probability words.

Figure 3. Mean difference scores and standard errors measuring phonotactic preferences for full-term (dark grey bars) and preterm infants (light grey bars) at 6 and 9 months.



2.2.4 Discussion

The goal of these experiments was to investigate whether preterm and full-term infants show different patterns of sensitivity to native-language phonotactics. Like the previous study assessing early sensitivities to lexical stress, this study showed that full-term and preterm infants have opposite listening preferences for high and low phonotactic probability words at both 6 and 9 months. While full-term participants preferred to listen to nonce words with less probable sound structures, preterm participants preferred to

listen to words with more probable sound structures. Thus, from the perspective of novel-familiar responses, full-term infants displayed a novelty preference by listening longer to words with less familiar sound-pattern combinations. On the other hand, preterm infants showed a familiarity preference by listening longer to words with more common sound-pattern combinations. These findings underscore the notion that preterm infants do not have the same perceptual sensitivities to their native language as their full-term counterparts.

Note that the present results for full-term infants are incongruous with Jusczyk and colleagues' (1994) findings, which showed that, when given a similar task, full-term, English-learning 9-month-olds preferred to listen to words with high phonotactic probabilities (i.e., familiarity preference). This discrepancy likely stems from differences in the difficulty of the stimuli with which subjects were presented. In the original Jusczyk et al. study, infants were tested on a total of 240 unique nonce words (120 with high and 120 with low phonotactic probabilities) over 16 trials (i.e., 15 unique words per trial). This experimental paradigm was unusually complex, especially for testing younger infants; unsurprisingly, 6-month-old participants could not distinguish between high and low phonotactic probability patterns. In an effort to elicit a phonotactic preference before 9 months and minimize subject attrition rates, the present test set-up was simplified: infants were instead presented with only 36 unique nonce words (18 with high and 18 with low phonotactic probabilities) over 12 trials (i.e., 6 unique words per trial). This set-up allowed for repeated presentations of each unique word, thus decreasing task demands and permitting full-term infants to focus their attention on those sound patterns they were less familiar with. Support for this idea comes from seminal work by Hunter and Ames

(1988), which showed that increased exposure to familiar stimuli results in an attentional shift from the more familiar to the less familiar stimuli. Assuming that full-term participants came into the lab with knowledge of their native-language's phonotactic structures and probabilities, then it is not surprising that they would find words with low probabilities more interesting to listen to than those with high probabilities.

Exhibiting a novelty preference has important implications for general learning strategies, especially when considering the significance of the behavioral differences observed across full-term and preterm participants in the current study. Specifically, researchers believe that the attentional shift from more to less familiar stimuli arises when the more familiar stimuli are fully encoded, or learned, subsequently allowing attentional processes to focus on learning the less familiar stimuli (Houston-Price & Nakai, 2004). On this account, preterm participants' preference for words with more familiar phonotactic patterns could indicate inadequate encoding of their native-language's sound structures. This is surprising, given that, on average, preterm participants were chronologically almost two and a half months older than their full-term counterparts. Thus, greater *ex-utero* experience with their native language does not afford preterm infants successful acquisition of that language's phonotactic properties.

Additional support for this notion comes from mismatch-negativity EEG data showing that preterm infants who are chronologically 12 months old (and 9 months corrected age) continue to discriminate a non-native phonetic contrast even though their 12-month-old full-term counterparts cannot (Peña, Werker, & Dahan-Lambertz, 2012). In fact, these preterm infants showed the same pattern of neural activity as full-term 9-month-old

controls, suggesting that neural maturation rather than duration of language exposure modulates the ability to discriminate phonetic contrasts.

The results to the present experiment further support the idea that preterm birth could indirectly impact language development by interfering with attentional skills necessary for successful learning. Like the data on sensitivities to lexical stress, these data show that preterm infants attend to phonotactic probability patterns differently than their full-term counterparts. Specifically, the latter prefer to listen to words with low (i.e., less familiar) probabilities, while the former prefer to listen to words with high (i.e., more familiar) probabilities. Preemies' atypical attentional pattern to phonotactics could impact word-learning by affecting how words are segmented from fluent speech as well as the ease with which novel words are learned, stored and accessed. An early bias for more common sound-sequence combinations could delay word-learning through continued focus on words with familiar phonotactic patterns, thus detracting attention from learning and storing new and less familiar-sounding lexical units.

2.3 General Discussion

Results from the experiments assessing sensitivities to lexical stress and phonotactic probabilities showed that, while preterm infants do seem to have some degree of discriminability between trochaic and iambic stress and between high and low phonotactic probabilities, these infants are not behaving like their full-term counterparts. Both tasks showed that, when compared to full-term controls, preterm infants have opposite patterns of preference for stress and phonotactic patterns. While these data cannot specify why and how these differences can lead to language delays, they

underscore atypicalities in how preterm infants perceive and process spoken language. A potential outcome resulting from this atypicality is uncharacteristic representations of native-language structure early in the process of language learning. Importantly, any prospective language delays observed in this preterm population are unlikely to result from a lack of perceptual sensitivities to native-language structures, but rather from differences in how these infants are listening to and learning about their language.

Chapter 3: Representations of Native-Language Word Forms

3.1 Sensitivities to Mispronunciations

Previous work in our lab has shown that, by 19 months of age, full-term, typically developing, English-learning toddlers have a graded sensitivity to word-initial (White & Morgan, 2008) and word-final (Ren and Morgan, 2012) mispronunciations. That is, when presented with a pair of familiar and unfamiliar objects, toddlers tend to look less at the target object (and more at the novel object) as the degree of target word mispronunciation becomes more severe. Thus, words that are only slightly mispronounced are still accepted as target words rather than labels for the unfamiliar objects. On the other hand, target words that are severely mispronounced are potentially considered to be new words for the unknown referent. These findings suggest that native-language word form representations at 19 months are adult-like: like adults, toddlers at this age are sensitive to phonological discrepancies. Furthermore, these results show that semantic processing at this age is already context-dependent and flexible: given an object with a known label and one with an unknown label, toddlers can reason that the mispronounced word could be the name of the unfamiliar object. Having such sensitivities to the phonological specifications of native-language words can act as bootstraps for subsequent word-learning.

The goal of the current experiment was to replicate White and Morgan (2008) with preterm toddlers in order to explore and compare their sensitivities to word-initial mispronunciations to those of their full-term counterparts. Given that preterm children are known to have reduced vocabulary growth, their patterns of sensitivity to mispronunciations were expected to differ from those reported for age-matched, full-term

toddlers. Furthermore, preterm participants were anticipated to be less flexible in their interpretations of mispronounced target words.

3.1.1 Participants

Twenty preterm (13 female, mean corrected age= 583.2 days, range= 570-597 days) and 40 full-term toddlers (20 female, mean age= 580.2 days, range= 571-592 days) were tested at 19 months. Chronologically, preterm infants were 68.5 days older than their full-term counterparts (mean chronological age= 648.8 days, range= 602-695 days). Two additional full-term participants were tested but excluded from analyses due to technical complications and brain development issues.

3.1.2 Method

As previously stated, the present study was a replication of White and Morgan (2008). Participants were tested in a sound-attenuated room using the Intermodal Preferential Looking Paradigm (IPLP). The parent sat with the child on his/her lap, and heard instrumental music over noise-cancelling headphones to discourage him/her from listening to the stimuli and influencing his/her child's behavior. The child was seated approximately 90 cm away from two monitors with a light centered between them. The experimenter observed the child's looking behavior over a closed-circuit video system. The testing session was recorded on a digital camcorder at 30 frames per second for later off-line coding. Speech stimuli were played at a conversational level (approximately 70 dB).

Trials were comprised of two phases: salience and test. Each trial began with the center light flashing until the child fixated at the midline. The light was then extinguished and the experimenter initiated the salience phase. During this phase, two objects were simultaneously presented to the child in silence for 4 seconds. The goal of salience was to establish the child's baseline looking preferences to each object. Once the screens went off, and after a minimum pause of 1 second, the experimenter called for the center light to start flashing again. Once the child's attention was recaptured, the test phase was initiated. According to White and Morgan (2008), the purpose of re-centering attention between phases was to prevent participants from fixating on one screen for the duration of the trial. At test, the same two objects reappeared on the monitors and were displayed for 9 seconds. Unlike in the salience phase, the object images were now accompanied by two audio stimuli. The first stimulus, "*Where's the (target word)?*", was synchronized with the onset of object presentation; the second audio stimulus, "*Find the (target word)!*", was presented 3 seconds after the offset of the target word, and was used to sustain the child's attention to the remaining portion of the trial. Presentation of target and distracter objects was counterbalanced across monitors.

Following the testing session, the parent completed a questionnaire on his/her child's comprehension and production of the words pertaining to the stimulus objects. Half of the objects were expected to be familiar and the other half were expected to be unfamiliar.

3.1.2.1 Stimuli

Visual and auditory stimuli were the same as those used by White and Morgan (2008) in one of the four counterbalancing conditions they used for Experiment 1. Visual stimuli were cartoons of target objects displayed against white backgrounds. Auditory stimuli were recorded by a female native English speaker.

Children were tested on 18 trials, each comprised of a familiar-unfamiliar object pairs. The familiar object stimuli had labels that, according to parental report norms, are understood by at least 50% of infants by age 14 months. Thus, 19-month-olds participating in the present study were expected to know the words corresponding to these objects. In contrast, the unfamiliar object stimuli consisted of artifacts that 19-month-olds are unlikely exposed to: with the exception of *pickle*, none of the labels corresponding to these stimulus objects appear on normed parental reports such as the MCDI. Table 3 provides a complete list of the stimulus pairings shown to participants.

Of the 18 trials, half involved correct labeling of the target object and half involved mispronunciations of the target object. Mispronunciations were limited to onset consonant changes that ranged in severity: 1-feature changes involved alterations to place of articulation, 2-feature changes to place and voice, and 3-feature changes to place, voice and manner. Each child heard five Correct trials, three 1-Feature Change trials, three 2-Feature Change trials, and three 3-Feature Change trials; the target in these four trial types was always the familiar object. Four Novel trials were also presented to discourage subjects from developing a strategy in which they always looked to the familiar object. The target for these trials was the unfamiliar object, and its label was always correctly pronounced. According to White and Morgan (2008), Novel trials

provide a means for assessing looking behavior in the context of an unfamiliar object-word pairing. By virtue of mutual exclusivity, participants were expected to look to the correct object during these trials, despite their lack of familiarity with both the object and its corresponding label.

Given the longitudinal nature of the present study, the order of presentation for stimulus pairings was kept the constant for all participants. However, trial type presentation was randomized.

Table 3. Stimulus pairings displayed to participants organized according to trial type: 1- (place), 2- (place + voice) and 3-feature (place + voice + manner) onset consonant changes. Familiar objects are listed on the second column and unfamiliar objects on the third column. Pronunciation of the target word for each trial type is provided on the fourth column.

DISPLAY PAIRINGS			
Trial Type	Familiar Image	Unfamiliar Image	Pronunciation of Target Word
1-Feature Change	shoe	padlock	<i>foo</i>
	ball	pump	<i>gall</i>
	bear	hourglass	<i>gear</i>
2-Feature Change	cookie	shuttlecock	<i>dookie</i>
	bird	pickle	<i>kird</i>
	hand	can opener	<i>zand</i>
3-Feature Change	book	paint roller	<i>sook</i>
	car	barrel	<i>yar</i>
	keys	garlic	<i>zeys</i>
Correct	bottle	accordion	<i>bottle</i>
	cup	artichoke	<i>cup</i>
	foot	bottle opener	<i>foot</i>
	balloon	door knocker	<i>balloon</i>
	bunny	French horn	<i>bunny</i>
Novel	banana	lantern	<i>lantern</i>
	chair	trophy	<i>trophy</i>
	flower	bullhorn	<i>bullhorn</i>
	telephone	beehive	<i>hive</i>

3.1.2.2 Measures of Interest

Looking behavior during each trial was coded off-line for both salience and test phases; frame-by-frame (1 frame= 33 ms) analyses were conducted using the Metcalf Infant Lab Coder program (in-house software). For salience, looking time to the target and distracter objects was coded for the full 4-second duration of the phase. For test, looking time to the target and distracter was only coded for the 3 seconds following target word onset. The proportion of time spent looking to the target object was then computed for each phase by dividing the time spent looking at the target by the total time spent looking at the target and distracter for that phase. A difference score was calculated for each trial, and these were averaged according to trial type for each participant. The mean difference score for each trial type was then calculated using the following formula:

$$[\text{Mean Proportion Looking (Target)}_{\text{TEST}}] - [\text{Mean Proportion Looking (Target)}_{\text{SALIENCE}}].$$

This measure was used to assess change in looking to the target in response to target word onset. Furthermore, comparing looking time to the target across the test and salience phases corrected for potential preferences for a particular object in a stimulus pair. That is, each stimulus pair served as its own control.

Additional measures of interest were calculated for the test phase. These included: total number of fixations⁶ to the target and to the distracter, length of the longest fixation to the target and to the distracter, direction of the first fixation and total number of gaze switches between the target and distracter. Again, these calculations were limited to the 3 seconds following target word onset. The purpose of these computations was to assess the level of uncertainty participants encountered as a function of the degree of

⁶ A fixation was defined as ≥ 10 consecutive frames of looking to a given object.

mispronunciation. For instance, children were expected to have a greater number of gaze switches between the target and distracter as their uncertainty for target word interpretation increased. Under this stipulation, the number of switches were anticipated to be greater for mispronounced target words than for correctly pronounced target words.

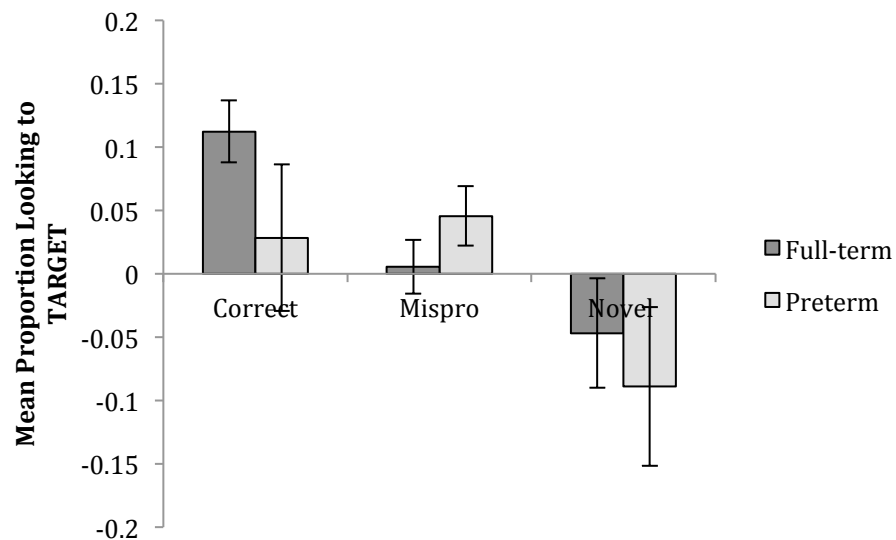
Trials were excluded from looking time analyses if the child: (a) was unfamiliar with both objects in a pairing, (b) only looked at the target or distracter during the salience phase (c) looked at neither the target or distracter during salience and/or test phases, and (d) looked at the target and distracter for less than 25% of the time during the 3-second interval of interest during the test phase.

3.1.3 Results

The data show that, while full-term participants spent significantly less time looking at the target object when the target word was mispronounced, preterm participants spent about the same amount of time looking at the target object regardless of whether the word was pronounced correctly or not. Interestingly, neither group of participants showed evidence for mutual exclusivity: during Novel trials, toddlers spent a greater proportion of time looking at the familiar objects instead of at the target unfamiliar ones (see Figure 4). For full-term toddlers, the mean proportion looking times to the target object were 0.112 (SD= ± 0.154), 0.00548 (SD= ± 0.134) and -0.0468 (SD= ± 0.273) for Correct, Mispronunciation, and Novel trials respectively. For preterm infants, the mean proportion looking times to the target were 0.0285 (SD= ± 0.252), 0.0457 (SD= ± 0.105) and -0.0892 (SD= ± 0.273) for Correct, Mispronunciation, and novel trials respectively. A 3 (trial type) by 2 (birth status) repeated measures ANOVA confirmed a

main effect of trial type, $F(1.72, 57) = 5.88, p = .006$; the interaction between trial type and birth status was not significant, $F(1.72, 57) = 1.52, p = .226$. To explore the basis for the main effect, a 3 (trial type) by 1 repeated measures ANOVA was run for each group of participants. These showed that the main effect of trial type was significant for full-term, $F(1.45, 38) = 5.90, p = 0.01$, but not preterm, $F(1.86, 18) = 2.53, p = .097$, participants. Bonferroni corrected pairwise comparisons confirmed that the difference in mean proportion looking measures between correct and mispronounced trials was statistically significant for full-term participants, $p = .003$.

Figure 4. Mean proportion looking to the target during Correct, Mispronunciation (i.e., combined means of 1-, 2-, and 3-Feature Change), and Novel trials for full-term (dark grey bars) and preterm participants (light grey bars). Error bars represent standard error.

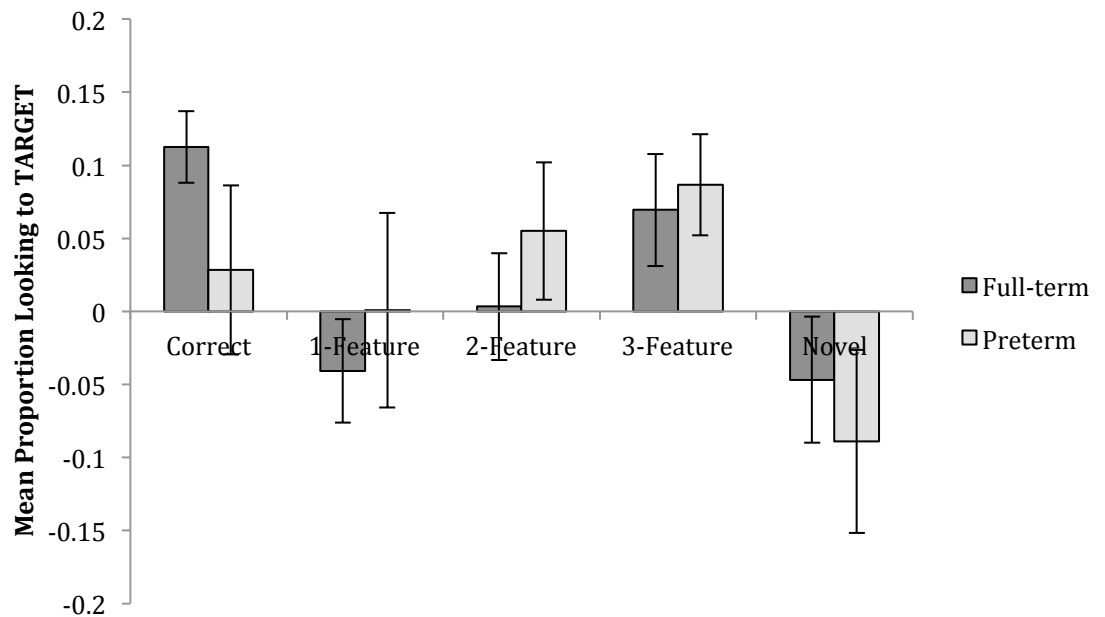


An additional set of analyses was conducted to explore whether participants across groups showed a graded sensitivity to type of mispronunciation. Results showed that, for both full-term and preterm participants, the mean proportion looking to the target increased as the severity of the mispronunciation increased (see Figure 5). This pattern of looking behavior was unexpected, as White and Morgan (2008) showed the opposite effect: in their study an increase in the degree of mispronunciation resulted in decreased looking to the target.

For full-term toddlers, the means for proportion looking to the target during 1-Feature, 2-Feature, and 3-Feature Change trials were -0.0407 (SD= ± 0.224), 0.00328 (SD= ± 0.228), and 0.0694 (SD= ± 0.243), respectively. For preterm infants, the means were 0.000621 (SD= ± 0.298), 0.0549 (SD= ± 0.210), and 0.0865 (SD= ± 0.151) for 1-Feature, 2-Feature, and 3-Feature Change trials respectively. A 5 (trial type) by 2 (birth status) repeated measures ANOVA reaffirmed a main effect of trial type, $F(3.95, 55) = 3.65$, $p = .007$; the interaction between trial type and birth status was again not significant, $F(3.95, 55) = 1.01$, $p = .405$, suggesting that the pattern of looking behavior was the same for full-term and preterm participants. Two follow-up 5 (trial type) by 1 repeated measures ANOVAs were performed to further explore the relationship between the different types of trials within each group of participants. These revealed that differences in mean proportion looking to the target across trial types were significant for full-term, $F(3.51, 36) = 3.45$, $p = .014$, but not preterm, $F(4, 16) = 1.72$, $p = .154$, children. Bonferroni corrected pairwise comparisons within the full-term group indicated that the mean for Correct trials was significantly different from the means for 1-Feature Change, $p = .003$, and Novel, $p = .039$, trials. Means were not significantly different across 1-Feature, 2-

Feature, and 3-Feature Change trials: $p= 1.00$ for 1- vs. 2-Feature Change trials, $p= .476$ for 1- vs. 3-Feature Change trials, and $p= 1.00$ for 2- vs. 3-Feature Change trials.

Figure 5. Mean proportion looking to the target during Correct, 1-Feature, 2-Feature and 3-Feature Change, and Novel trials for full-term (dark grey bars) and preterm (light grey bars) participants. Error bars represent standard error.



3.1.4 Discussion

The aim of this experiment was to investigate whether preterm individuals are sensitive to mispronunciations in the same way full-term infants are. The data showed that, unlike full-term toddlers, preterm toddlers were not sensitive to mispronunciations: while the former spent significantly more time looking at the target during Correct trials than Mispronunciation trials, the latter spent about an equal amount of time looking at the target regardless of whether it was pronounced correctly or not (see Figure 4). This

finding can be interpreted in a couple of different ways. First, equal looking to the target during Correct and Mispronunciation trials could indicate that preterm toddlers are perceptually insensitive to differences between 1-, 2-, or 3-feature changes to a word-initial segment. This seems unlikely given that mean looking to the target for 1- and 2-Feature Change trials was different from that for 3-Feature Change trials (see Figure 5). That said, these differences did not reach statistical significance, so this possibility cannot be entirely ruled out on the basis of this study's findings. At the same time, it is unsurprising that differences in means across types of mispronunciations were not statistically reliable given the extensive variability in looking behavior observed, especially within the preterm group (see error bars for Figures 4 and 5). At any rate, the chances that preterm infants cannot perceptually distinguish between 1-, 2-, and 3-feature mispronunciations seem unlikely.

A second possibility for explaining the lack of differences in target looking across Correct and Mispronunciation trials is that preterm participants simply accept changes to word-initial segments more readily than full-term controls. This could indicate that preterm children's representations of known words is not as detailed as that of their full-term counterparts. Of course, this is problematic for word-learning, particularly in situations when similar sounding words with distinct referents have to be learned. Such a task would become infinitely more challenging as confusion would likely ensue when determining object-word correspondences. Thus, less detailed representations of known words could delay the rate at which novel words are learned.

Though full-term participants in the present study replicated White and Morgan's (2008) general finding that typically developing toddlers look less to a target when it is

mispronounced, these toddlers did not demonstrate the same pattern of graded sensitivity to mispronunciations as that reported by these researchers. Specifically, White and Morgan showed that, as mispronunciations become more severe, 19-month-olds look increasingly less at a target. Unexpectedly, looking data from the present study showed the opposite pattern: looking to the target actually increased as mispronunciations became more severe. Though only significant in the full-term group, this pattern of looking behavior was also observed in the preterm group (see Figure 5). If the pattern of looking behavior to mispronunciations described by White and Morgan is considered typical, then even full-term participants in the present study behaved aberrantly. What can explain this unexpected behavior among full-term controls? Closer inspection of the full-term participant pool revealed that, by their final visit to the lab at 26 months, 6 of the 40 toddlers who took part in this experiment were either receiving early intervention services for speech ($n=4$) or some condition that could impact language development ($n=2$). An additional 8 children were reported to have a relative with a diagnosed language ($n=3$), learning or developmental ($n=5$) issue. While problems with language development were expected among preterm participants, such a high incidence was unexpected among full-term participants. Additional analyses of this data set need to be performed while excluding full-term toddlers who were at risk for language delays and impairments (i.e., receiving early intervention services by 26 months or had relatives with language, learning or developmental disorders). This would reveal whether these toddlers are driving the unexpected patterns of looking behavior observed within the full-term group. Why might so many individuals in this study's full-term sample be at risk for language delays and impairments? One reason is parents' motivation for committing to participate

in a 22-month long study. Perhaps knowing that someone in your family has a language or developmental disorder may have provided an additional incentive for parents to bring their full-term children to the lab for multiple visits. In future work it would be worthwhile to ask parents about their motivations for bringing their infants and children to the lab.

Another point of departure between these results and those of White and Morgan (2008) lies in looking behavior during Novel trials. Interestingly, neither full-term nor preterm participants in this study showed evidence of mutual exclusivity for these trials: both groups of toddlers spent more time looking at the familiar object than at the target object when the target's label was unfamiliar to them. This could represent a general bias or inability to disengage attention from that which is familiar to the participants. This is an interesting point to consider, given that with increasing age, children tend to show a preference for novelty. This is potentially beneficial to word-learning, as it creates opportunities for the child to learn labels for novel things. For example, a mother is more likely to name a novel object for her child if she notices her child is attending to it. Perhaps, the familiarity preference observed for Novel trials in this study is an indirect measure of poorer language outcomes among full-term and preterm participants in this study. Future work should focus on investigating how performance on this task relates to word-learning rates. The role of looking behavior in this experiment on language outcomes is explored in *Part III* of this paper, *Relations Between Language, Cognitive And Environmental Predictors, And Later Language Outcomes*.

On a more general note, some of the result discrepancies between this study and White and Morgan's (2008) might be attributed to the stimuli used to test participants.

The latter study had four counterbalancing conditions to ensure that each familiar-unfamiliar stimulus pair was assigned to each of the four pronunciation conditions, Correct, 1-, 2-, and 3-Feature Change. However, since the present study was part of a larger longitudinal study, all participants had to be tested in the same condition. Though the condition used was identical to one of the four used by White and Morgan, it may not have been the best condition to use. That is, relative to the other three conditions, the one selected may have not reflected the overall looking pattern described by the researchers as strongly. Since White and Morgan did *not* report *statistically significant* differences in their results across conditions, the condition used in this experiment was randomly selected. Additionally, unlike White and Morgan's 19-month-olds, the toddlers participating in the present study had been tested in the lab multiple times with methods similar or identical to those used in this experiment. Perhaps re-test effects could partially account for the unexpected patterns of looking behavior observed in this study.

3.2 Speed and Efficacy of Word Processing

The speed and efficiency with which words are processed in fluent speech also play a critical role in facilitating word-learning. Fernald, Perfors and Marchman (2006) showed that full-term 25-month-old toddlers who process language more quickly and with greater accuracy show accelerated rates of expressive vocabulary growth between the ages of 12 to 25 months. In this study, children's online language comprehension was assessed using the Looking-While-Listening procedure, in which children were presented with pairs of familiar images representing a target and a distracter, and then directed to look at the target. These researchers showed that reaction time (RT) and looking accuracy

to the target image were highly correlated with three concurrent production measures collected with the WS-MCDI: number of words produced, grammatical complexity, and mean length of the three longest utterances produced by the child. Importantly, the predictive role of online comprehension measures of language outcomes held true for children with atypical patterns of language growth. Specifically, Fernald and Marchman (2012) tested a sample of full-term late-talking toddlers and showed that word recognition skills at 18 months were predictive of vocabulary growth between the ages of 18 to 30 months: children with longer RT's and less accurate word processing skills on the Looking-While-Listening task tended to have stunted vocabulary growth.

The goal of the present experiment is to assess preterm toddlers' online language comprehension skills relative to those of their full-term counterparts at 26 months. Behavioral comparisons between these two groups of toddlers could reveal whether preterm infants show patterns of looking behavior that are not only different from those of typically developing full-term children, but also different from full-term late-talkers. Furthermore, in light of the findings described above, knowing how quickly and accurately preterm toddlers respond to familiar words when listening to fluent speech could provide insight on their vocabulary growth rates and why they are at risk for language delays and impairments.

3.2.1 Participants

Nineteen preterm (mean corrected age= 789.7 days, range= 776-804 days) and 38 full-term children (mean age= 790.0 days, range= 776-811 days) were tested at 26 months. Chronologically, preterm infants were 67.7 days older than their full-term counterparts (mean chronological age= 857.7 days, range= 811-898 days).

3.2.2 Method

The present experiment was modeled after work currently being conducted by Anne Fernald and her team of researchers with preterm children at 24 months chronological age. In the present study toddlers were tested in a sound-attenuated room using IPLP. The parent sat with the child on his/her lap and wore noise-canceling headphones to discourage him/her from listening to the stimuli and influencing his/her child's behavior. The child was seated approximately 90 cm away from two monitors with a light centered between them. The experimenter observed the child's looking behavior over a closed-circuit video system and recorded on a digital camcorder at 30 frames per second for later off-line coding. Speech stimuli were played at a conversational level (approximately 70 dB).

Each trial began with the center light flashing until the child fixated at the midline. The light was then extinguished and two objects were presented simultaneously. After 2 seconds of silent presentation, the audio stimuli came on and directed the child to look at the target object. Once the audio stimuli went off, the objects remained on the screens for an additional second before the trial ended. Side presentation of target objects was counterbalanced.

3.2.2.1 Stimuli

Auditory and visual stimuli were borrowed from Anne Fernald, director of the Language Learning Lab at Stanford University. She is using these stimuli on an ongoing project investigating word processing abilities in preterm children at 24 months chronological age. Auditory stimuli were recorded by a female native English speaker, and visual stimuli were target objects photographed against white backgrounds.

Toddlers were tested on 40 trials, each comprised object-pairs with labels that are highly familiar to 2-year-olds. During each trial, the participant was asked to look at one of the two objects, with the target object occurring at the end of a carrier phrase (e.g., “*Where’s the (TARGET)?*”). Children were presented with five types of trials: eight Vanilla trials (Vanilla), eight Informative Verb trials (I-verb), eight Uninformative Verb trials (U-verb), eight Informative Adjective trials (I-adj) and eight Uninformative Adjective trials (U-adj).

From a speech processing prospective, Vanilla trials constituted the easiest trials: the participant simply had to wait until the target noun was enunciated (at the end of the carrier phrase) to determine what object to look at. The purpose of these trials was to help establish a baseline for orientation behavior to a target in the absence of context cues that could help identify the target before it was said. The informative trials, I-verb and I-adj, contained a verb or an adjective, respectively, that was related to the target noun, thus potentially disambiguating the identity of the target before it was enunciated. For example, in an I-verb trial, the child was presented with images of a glass of juice and of a cookie and was asked to, “***Drink** the juice*”; in this context, hearing the word “***drink***” should be enough to elicit looking towards the glass of juice before “***juice***” is said.

Similarly, in an I-adj trial, the child was shown images of a blue shoe and a red ball and asked “Which is the red ball?”; in this context, hearing the word “red” should be sufficient to elicit looking towards the ball before “**ball**” is said. Thus, in the informative trials, the participant did not have to wait until the end of the carrier phrase to determine what object she should orient to. The purpose of these types of trials was to determine if a participant could efficiently use words related to the target noun to guide her looking behavior.

Alternatively, uninformative trials were like Vanilla trials in that the participant had to wait until the end of the carrier phrase to determine the identity of the target noun, and orient their looking accordingly. For instance, during an U-verb trial, the child was presented with images of a glass of juice and a cookie and told “Show me the juice”. Likewise, in an U-adj trial, the child was shown a red shoe and a red ball and asked, “Which is the red ball?”. Neither of these cases provides the child with the context cues necessary to determine the target noun before it is enunciated.

Throughout the course of the experiment, participants were presented with six unique object pairs. For Vanilla trials, these were *baby – doggie* and *kitty – birdie*. Verb trial pairs were *car – book* and *cookie – juice*. Finally, Adjective trial pairs were *cup – sock* and *ball – shoe*. With the exception of Adjective trial pairs, object pairings were presented four times, with each object serving as the target twice. Pairings for the I-adj and U-adj trials only occurred once to ensure that each color adjective was paired with each object. Refer to Table 4 for all possible audio-visual pairings within each trial type. Side of target object presentation was counterbalanced across trials. To reduce boredom

resulting from repetitious visual object pairings, two unique pictures were used for babies, doggies, kitties, birdies, car, books, cookies, and juice.

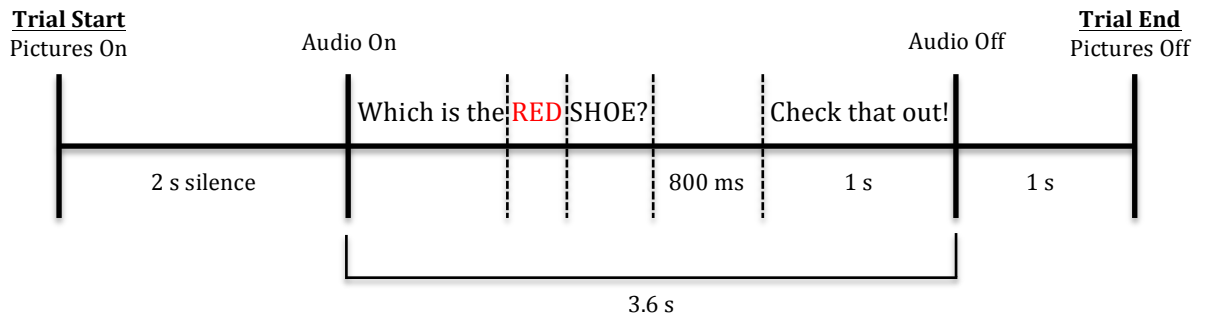
In addition to the 40 test trials, participants were given four attention-getting trials to keep them engaged throughout the duration of the testing procedure. These trials were interspersed among the test trials and consisted of pairs of child-friendly cartoon characters (e.g., Elmo – Cookie Monster) or animals (e.g., koala – monkey) accompanied by an attention-getting audio stimulus (e.g., “Do you like the pictures?”).

Trials lasted 6.6 seconds a piece, with the target and distracter images remaining on screen for the entire duration of the trial. Each trial was comprised of 2 seconds of silent object presentation followed by 3.6 seconds of auditory stimuli, and ended with an additional second of silent object presentation. The auditory portion of the trial consisted of two statements: the carrier phrase containing the target noun and any related (or unrelated) verbs or adjectives, and an attention-maintaining phrase to keep the participant on task (e.g., “*Can you it?*”, “*Check that out!*”). These were separated by 800 ms of silence. Refer to Figure 6 for a sample of trial structure.

Table 4. List of auditory stimuli and corresponding visual displays for each trial type. Target nouns and potentially informative verbs and adjectives appear in bold and are underlined. With the exception of auditory stimuli for I-adj and U-adj trials, each auditory stimulus was presented twice throughout the testing session.

Trial Type	Auditory Stimulus	Display Pairings	
Vanilla	<i>Where's the <u>baby</u>?</i> <i>Where's the <u>doggie</u>?</i>	baby	dog
	<i>Where's the <u>kitty</u>?</i> <i>Where's the <u>birdie</u>?</i>	kitten	bird
I-verb	<i><u>Drive</u> the <u>car</u>.</i> <i><u>Read</u> the <u>book</u>.</i>	car	book
	<i><u>Eat</u> the <u>cookie</u>.</i> <i><u>Drink</u> the <u>juice</u>.</i>	cookie	juice
U-verb	<i><u>Look</u> at the <u>car</u>.</i> <i><u>Take</u> the <u>book</u>.</i>	car	book
	<i><u>Show</u> me the <u>cookie</u>.</i> <i><u>Show</u> me the <u>juice</u>.</i>	cookie	juice
I-adj	<i>Which one's the <u>blue cup</u>?</i> <i>Which one's the <u>red cup</u>?</i> <i>Which one's the <u>blue sock</u>?</i> <i>Which one's the <u>red sock</u>?</i>	blue cup red cup red cup blue cup	red sock blue sock blue sock red sock
	<i>Which is the <u>blue ball</u>?</i> <i>Which is the <u>red ball</u>?</i> <i>Which is the <u>blue shoe</u>?</i> <i>Which is the <u>red shoe</u>?</i>	blue ball red ball red ball blue ball	red shoe blue shoe blue shoe red shoe
U-adj	<i>Which one's the <u>blue cup</u>?</i> <i>Which one's the <u>red cup</u>?</i> <i>Which one's the <u>blue sock</u>?</i> <i>Which one's the <u>red sock</u>?</i>	blue cup red cup blue cup red cup	blue sock red sock blue sock red sock
	<i>Which is the <u>blue ball</u>?</i> <i>Which is the <u>red ball</u>?</i> <i>Which is the <u>blue shoe</u>?</i> <i>Which is the <u>red shoe</u>?</i>	blue ball red ball blue ball red ball	blue shoe red shoe blue shoe red shoe

Figure 6. Structure of a sample Adjective trial. Each trial began with 2 seconds of silent presentation of target-distractor images. The participant was then directed to look at the target image. Verbal instructions were followed by an additional second of silent target-distractor image presentation after which point the screens went off and the trial ended.



3.2.2.2 Measures of Interest

Looking behavior during each trial was coded off-line, frame-by-frame (1 frame= 33 ms), using the Metcalf Infant Lab Coder program (in-house software). Two measures of interest were subsequently calculated from coded data: accuracy and speed of target word recognition.

Accuracy to the target object was defined as the proportion of time spent looking at the target object for the 2 seconds following 200 ms after the onset of a word that could potentially disambiguate target object identity (i.e., target nouns, verbs and/or adjectives). For Vanilla trials, accuracy was only calculated for the 2 seconds following target noun onset. For Verb trials, accuracy was computed for the 2 seconds following the phrase-initial verb onset and the 2 seconds following target noun onset. Similarly, for Adjective trials, accuracy was calculated for the 2 seconds following adjective onset and the 2 seconds following target noun onset. Accuracy was calculated for each trial using the following formula: $[\text{Looking}_{\text{TARGET}}] / [(\text{Looking}_{\text{TARGET}}) + (\text{Looking}_{\text{DISTRATER}}) +$

(Looking_{AWAY})]⁷. These proportions were then used to determine the mean proportion looking to the target for each type of trial. Trials in which the child spent less than 25% of the time on both the target and distracter during the 2-second window of interest were discarded from analyses. If toddlers successfully used verbs and adjectives to guide gaze orientation, then mean proportion looking to the target was expected to be greater in the 2 seconds after (1) verb onset in the I-verb trials than the U-verb trials, and (2) adjective onset in the I-adj trials than the U-verb trials. The mean proportion looking time to the target after target noun onset was expected to be similar across trial types.

Speed of target word recognition, or reaction time (RT), was defined as the amount of time it took participants to orient to the target object 200 ms following the onset of a word that could potentially disambiguate target object identity. For Vanilla, U-verb and U-adj trials, the disambiguating word was the target noun itself. Thus, RT for these trials was measured following the onset of the noun. For I-verb and I-adj trials, two words could help participants orient to the target object: a verb and the target noun, or an adjective and the target noun, respectively. Thus, two RT measures were collected for these trials: one following the onset of the verb or adjective and one following the onset of the target noun. Trials were excluded from analyses if the child (1) oriented to the target object before the 200 ms period following target noun onset for Vanilla, U-verb, and U-adj, or verb or adjective onset for I-verb and I-adj trials, or if (2) took more than 2 seconds to orient to the target object after target noun onset. Looking to the target object

⁷ Time spent looking away from the screens was included in these calculations because subjects had no reason to be looking at either the target or distracter when uninformative verbs or adjectives were first enunciated. Limiting calculations to only include looks to the target and the distracter would thus inflate mean proportion looking to the target, especially following verb and adjective onset in U-verb and U-adj trials.

in these situations is likely guided by chance, and unlikely reflects the participant's behavioral response to the auditory stimulus. Trials were also excluded from RT analyses if, during windows of interest, the child failed to look at both the target and distracter at least 25% of the time. While greater proportion looking to the target suggests better accuracy in speech processing, shorter RT's indicate more efficient word recognition abilities. If preterm toddlers in this study are indeed at risk for language delays and impairments, then they are expected to have longer RT's and spend less time looking at the target than their full-term counterparts.

3.2.3 Results

3.2.3.1 Accuracy Data

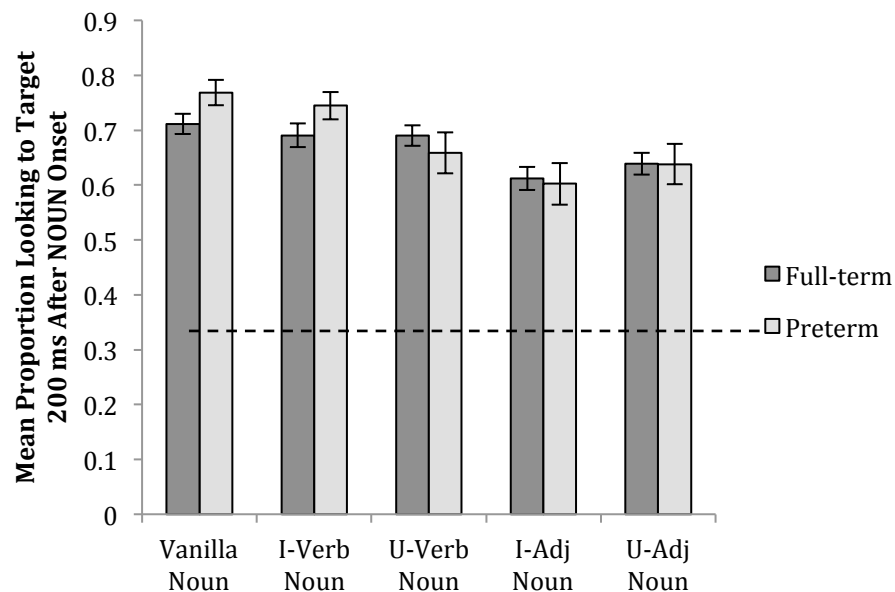
Mean proportion looking to the target was evaluated in three parts. Looking to the target was considered following (1) target noun onset for all five trial types, (2) verb onset for I-verb and U-verb trials, and (3) adjective onset for I-adj and U-adj trials.

Noun Onset

Within each trial type, full-term and preterm participants spent about the same proportion of time looking at the target following target noun onset (see Figure 7). Means for proportion of time spent on the target for full-term and preterm participants, respectively, were: 0.712 (SD= ± 0.114) and 0.769 (SD= ± 0.101) for Vanilla trials, 0.691 (SD= ± 0.134) and 0.745 (SD= ± 0.110) for I-verb trials, 0.690 (SD= ± 0.115) and 0.659 (SD= ± 0.164) for U-verb trials, 0.612 (SD= ± 0.129) and 0.602 (SD= ± 0.166) for I-adj trials, and 0.639 (SD= ± 0.122) and 0.638 (SD= ± 0.160) for U-adj. A 5 (noun type) by 2

(birth status) repeated measures ANOVA revealed a significant main effect of noun type, $F(4, 52) = 12.48, p < .001$; as expected, given the similar means across groups within trial types, the interaction between noun type and birth status was not significant, $F(4, 52) = 1.63, p = .167$. Bonferroni corrected pairwise comparisons indicated that the noun type main effect was driven by the following differences in mean proportion looking to the target: Vanilla *versus* U-verb, $p = .027$, I-adj, $p < .001$, and U-adj, $p < .001$; I-verb *versus* I-adj, $p < .001$, and U-adj, $p = .014$; and U-verb *versus* I-adj, $p = .052$.

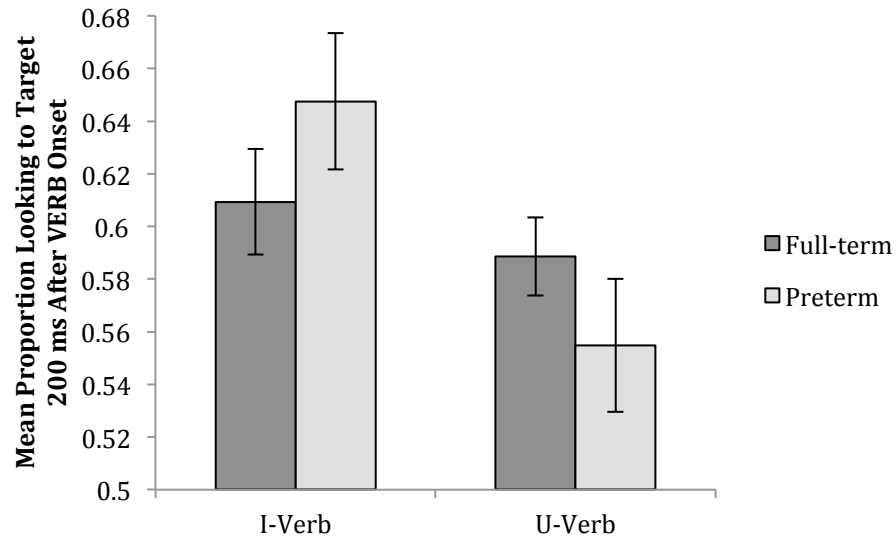
Figure 7. Target accuracy, or mean proportion looking to the target object, 200 ms following target noun onset for Vanilla, I-verb, U-verb, I-adj, and U-adj trials. Measures for full-term and preterm participants are shown in dark and light grey bars, respectively. Proportion looking to the target was significantly greater than 0.33, the expected proportion looking to the target if participants were behaving at chance (marked by the dashed line). Error bars represent standard error.



Verb Onset

Data for verb trials show that both full-term and preterm participants spent a greater proportion of time looking to the target after verb onset in I-verb than in U-verb trials (see Figure 8). For full-term toddlers, the means for the proportion of looking time to the target were 0.609 (SD= ± 0.124) and 0.589 (SD= ± 0.0916) for I-verb and U-verb trials, respectively. For preterm participants, these values were 0.648 (SD= ± 0.113) and 0.555 (SD= ± 0.110) for I-verb and U-verb trials, respectively. A 2 (verb type) by 2 (birth status) repeated measures ANOVA confirmed a main effect of verb type, $F(1, 55) = 8.58$, $p = .005$; the interaction between verb type and birth status was marginally significant, $F(1, 55) = 3.43$, $p = .069$. Follow-up independent-samples t-tests indicated that group differences in mean proportion looking within trial types were not statistically significant for either I-verb trials, $t(55) = -1.13$, $p = .264$, or U-verb trials, $t(55) = 1.22$, $p = .227$. Additional paired-samples t-tests showed that greater looking to the target after verb onset in I-verb trials was significant for preterm, $t(18) = 3.37$, $p = .003$, but not full-term, $t(37) = .881$, $p = .384$, participants.

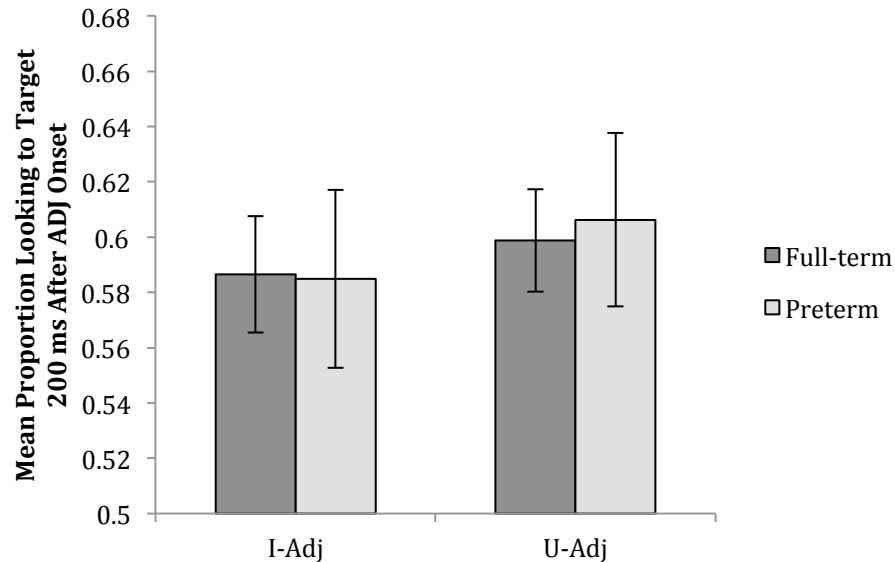
Figure 8. Target accuracy, or mean proportion looking to the target object, 200 ms following verb onset for I-verb and U-verb trials. Measures for full-term and preterm participants are shown in dark and light grey bars, respectively. Proportion looking to the target is again above chance for both groups of subjects. Error bars represent standard error.



Adjective Onset

Results for I-adj and U-adj indicate that mean proportion looking to the target following adjective onset was about the same across subject groups and adjective trial types (see Figure 9). I-adj means for full-term and preterm participants were 0.587 (SD= ± 0.130) and 0.585 (SD= ± 0.140), respectively. Means corresponding to U-adj trials were 0.599 (SD= ± 0.115) and 0.606 (SD= ± 0.137) for full-term and preterm subjects, respectively. A 2 (adjective type) by 2 (birth status) repeated measures ANOVA confirmed that neither the main effect of adjective type nor the interaction between adjective type and birth status were significant, $F(1, 55) = .548, p = .462$, and $F(1, 55) = .040, p = .842$, respectively.

Figure 9. Target accuracy, or mean proportion looking to the target object, 200 ms following adjective onset for I-adj and U-adj trials. Measures for full-term and preterm participants are shown in dark and light grey bars, respectively. Proportion looking to the target is again above chance for both groups of participants. Error bars represent standard error.



3.2.3.2 Speed of Target Word Recognition Data

Like accuracy, speed of target word recognition, or RT, was examined following target noun onset for all trial types, verb onset for I-verb and U-verb, and adjective onset for I-adj and U-adj trials. The purpose of the last two analyses was to explore whether hearing informative verbs and adjectives was helping participants orient to the target sooner than verbs and adjectives providing no information to disambiguate target identity.

Noun Onset

RT data following target noun onset show that, as a group, preterm participants were faster at orienting to the target object than full-term controls during I-verb, U-verb,

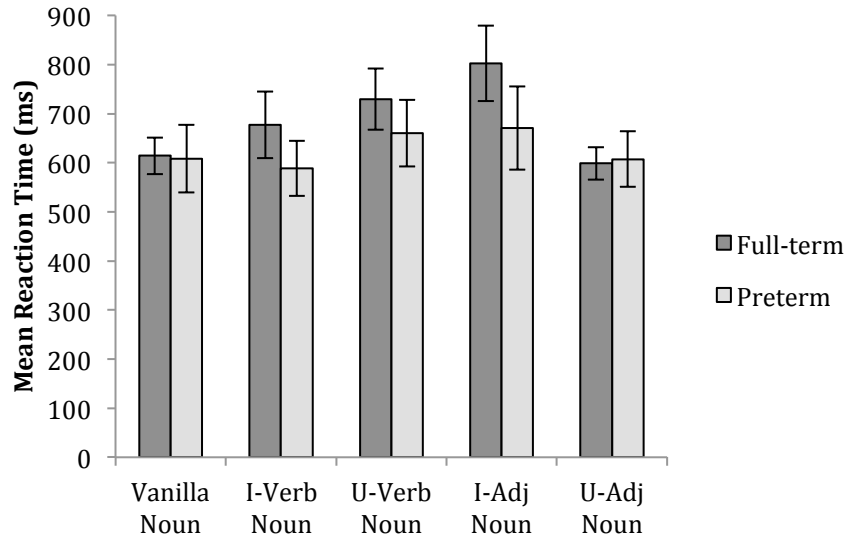
and I-adj trials; mean reaction times for Vanilla and U-adj trials were comparable across subject groups (see Figure 10). Vanilla mean RT's were 614.2 ms (SD= ± 215.7 ms) and 608.7 ms (SD= ± 293.4 ms) for full-term and preterm toddlers, respectively. I-verb mean RT's were 676.9 ms (SD= ± 393.9 ms) and 588.3 ms (SD= ± 208.8 ms) for full-term and preterm participants, respectively. U-verb mean RT's were 729.9 ms (SD= ± 380.7 ms) and 660.6 ms (SD= ± 281.2 ms) for full-term and preterm subjects, respectively. I-adj mean RT's were 802.3 ms (SD= ± 459.8 ms) and 670.5 ms (SD= ± 338.3 ms) for full-term and preterm children, respectively. Finally, U-adj mean RT's were 598.8 ms (SD= ± 201.7 ms) and 607.4 ms (SD= ± 232.4 ms) for full-term and preterm subjects, respectively.

Despite the noted differences in speed of target word recognition across trial types and subject groups, a 5 (trial type) by 2 (birth status) repeated measures ANOVA showed that neither the main effect of trial type, $F(3.45, 51) = 2.03, p = .103$, nor the interaction between trial type and birth status, $F(3.45, 51) = .708, p = .567$, were statistically significant⁸. This is unsurprising given the high levels of variability observed both within and between groups. To ensure that data imputation did not overshadow potentially significant RT differences within trial types, group means were compared using independent-samples t-test. These also indicated that group differences for each trial type

⁸ To maximize data use in statistical analyses, missing data were imputed using the t-score method previously described. For a given subject with missing data, t-scores were first computed for each RT measure; the mean of these was then used to estimate the missing data point. For Vanilla trials, five data points were imputed for full-term participants. For I-verb trials, four data points were calculated for each group of participants. For U-verb trials, one data point was computed for each group of subjects. For I-adj trials, two data points were calculated for each group of children. Finally, for U-adj trials, one data point was imputed for each group of participants.

were not significant, $t(49) = .075$, $p = .940$ for Vanilla trials, $t(46) = .793$, $p = .432$ for I-verb trials, $t(52) = .671$, $p = .505$ for U-verb trials, $t(50) = 1.03$, $p = .309$ for I-adj trials, and $t(52) = -.139$, $p = .890$ for U-adj trials.

Figure 10. Mean reaction time for target orientation following target noun onset for Vanilla, I-verb, U-verb, I-adj, and U-adj trials. Data for full-term and preterm subjects are shown in dark and light grey, respectively. Error bars represent standard error.



Verb Onset

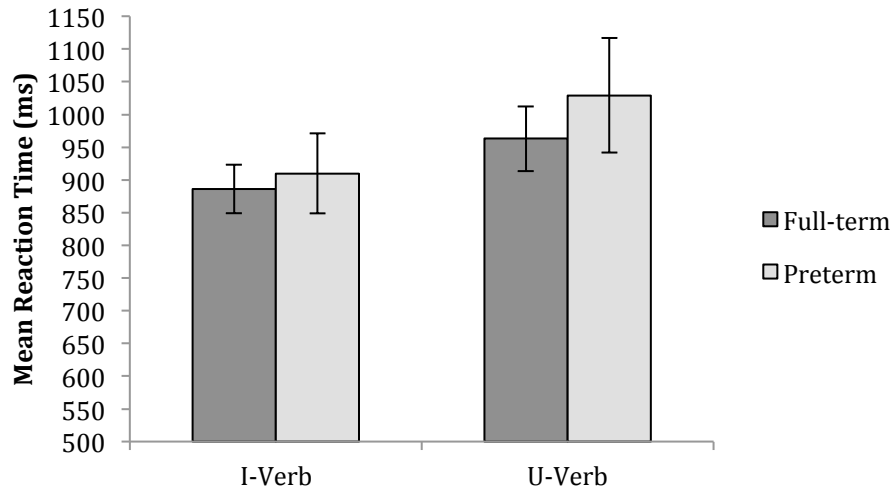
Though both full-term and preterm participants oriented to the target object faster following the onset of an informative than an uninformative verb, preterm children were slower than their full-term children during both I-verb and U-verb trials (see Figure 11). For I-verb trials, mean reaction times following verb onset were 886.4 ms (SD= ± 227.0 ms) and 909.9 ms (SD= ± 260.1 ms) for full-term and preterm participants, respectively.

For U-verb trials, mean RT's following verb onset were 963.0 ms ($SD = \pm 305.9$ ms) and 1028.9 ms ($SD = \pm 371.4$ ms) for full-term and preterm subjects, respectively. A 2 (verb type) by 2 (birth status) repeated measures ANOVA indicated that the main effect of verb type was significant, $F(1, 54) = 5.36, p = .024$, while the interaction between verb type and birth status was not, $F(1, 54) = .236, p = .629^9$. Follow-up paired-samples t-tests suggested that the main effect was driven by a marginally significant difference in RT measures between I-verb and U-verb trials for the full-term group, $t(36) = -1.92, p = .063$; cross-trial differences were not significant in the preterm group, $t(17) = -1.46, p = .162$. Independent-samples t-tests further confirmed that within trial group differences in RT were not significant, $t(53) = -.343, p = .733$ for I-verb, and $t(54) = -.703, p = .485$ for U-verb trials¹⁰.

⁹ ANOVA results include one imputed data point in the I-verb condition for the full-term group. This RT was estimated using the previously described t-score method. First, t-scores were calculated for RT measures in U-verb, I-adj, and U-adj trials. The mean of these three scores was then averaged and used to compute the missing RT value.

¹⁰ T-test analyses did not include the imputed data point.

Figure 11. Mean reaction time for target orientation following verb onset for I-verb and U-verb trials. Means for full-term and preterm participants are shown in dark and light grey bars, respectively. Error bars represent standard error.



Adjective Onset

RT data for I-adj and U-adj trials also indicated that both full-term and preterm participants oriented to the target object faster after hearing an informative than an uninformative adjective. However, preterm participants were this time faster than their full-term counterparts at recognizing the target for both I-verb and U-verb trials (see Figure 12). For I-adj trials, mean RT's following adjective onset were 772.5 ms (SD= ± 357.1 ms) and 755.8 ms (SD= ± 226.7 ms) for full-term and preterm participants, respectively. For U-adj trials, mean RT's following adjective onset were 887.6 ms (SD= ± 237.7 ms) and 804.4 ms (SD= ± 339.4 ms) for full-term and preterm subjects, respectively. Despite the observed trends, a 2 (adjective type) by 2 (birth status) revealed that neither the main effect of adjective type, $F(1, 54) = 2.07$, $p = .156$, nor the interaction

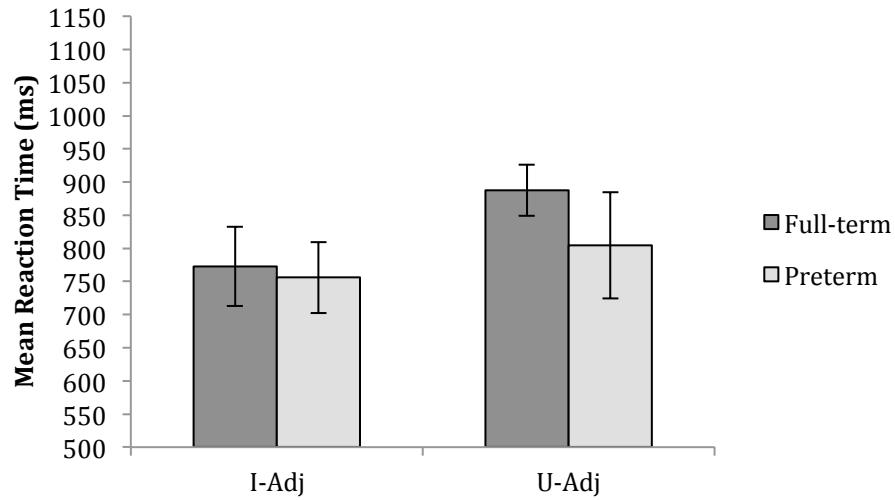
between adjective type and birth status, $F(1, 54) = .336, p = .565$, were statistically significant¹¹.

Despite the lack of significant effects found by omnibus testing, follow-up independent-samples t-tests were run to further examine group differences in RT within each type of adjective trial. These results further confirmed that RT measures were not significantly different between full-term and preterm groups for either I-adj, $t(52) = .181, p = .857$, or U-adj trials, $t(54) = 1.06, p = .293$. Further testing was conducted using paired-samples t-tests to explore within group differences across trial type. These analyses showed that while RT differences across I-adj and U-adj trials were not significant for preterm participants, $t(17) = -.565, p = .580$, they approached statistical significance for the full-term participants, $t(35) = -1.65, p = .107$ ¹².

¹¹ ANOVA results included two imputed data points in the I-adj condition for the full-term group. These RT estimates were computed using the same method previously described for the I-verb condition.

¹² T-test analyses excluded imputed data points.

Figure 12. Mean reaction time for target orientation following adjective onset for I-adj and U-adj trials. Means for full-term and preterm participants are shown in dark and light grey bars, respectively. Error bars represent standard error.



3.2.4 Discussion

The primary goal of the present study was to explore potential differences in word processing abilities across full-term and preterm 26-month-olds. This was approached in two ways using an online speech-processing task. First, this study compared accuracy and speed of target word recognition were across groups. Second, it assessed whether participants could successfully use target-related words combined with context cues to guide their target looking behavior. This was achieved by presenting children with two types of trials, informative (I-verb or I-adjective) trials, and uninformative (U-verb or U-adjective) trials, and comparing looking times between these trials. Comparisons were expected to show that target orientation would be faster for informative than uninformative trials. For example, when presented with a book and a car in a verb trial,

and told to “Read the book” (I-verb), children were predicted to look to the book sooner than when told to “Look at the book” (U-verb). Similarly, for adjective trials, children were expected to look to the ball faster when “Which is the red ball?” was paired with a red ball and a blue shoe rather than with a red ball and a red shoe.

Given preterm children’s known risk for language delays, full-term participants were hypothesized to outperform preterm participants on all measures, showing greater accuracy, shorter RT’s, and more effective use of informative verbs and adjectives to guide target looking. Though the resulting data did corroborate some of these expectations, it also yielded some unexpected patterns of looking behavior. These findings are further discussed below. Note that, although the stimuli used for this experiment were borrowed from one of Anne Fernald’s studies, comparisons cannot be drawn at this time between these results and her results because she continues to collect and analyze data.

3.2.4.1 Accuracy

Noun Onset

When presented with a target noun, preterm participants were just as accurate at target word recognition as their full-term counterparts: both groups of toddlers spent a greater proportion of time looking at the target following target noun onset, regardless of trial type. Significant differences in mean proportion of time spent looking at the target were found across trial types. Specifically, compared to Vanilla trials, subjects looked significantly less to the target after target noun onset during U-verb, I-adj, and U-adj trials. This finding was expected for I-adj trials: hearing the adjective directed looking to

the target before the target noun was heard, thus, by the time the target was enunciated, the child's interest in the target object had likely started to wane. However, decreased target looking during U-verb and U-adj trials relative to Vanilla trials was unexpected given that all these trial types shared a common structure; like Vanilla trials U-verb and U-adj trials did not contain any information that would help the child identify the target prior to target word onset.

What might be driving these differences in looking behavior across Vanilla, U-verb, I-adj, and U-adj trials? One possibility, specifically for I-adj and U-adj trials, is choice of target-distracter object pairings. Unlike I-verb and U-verb trials, I-adj and U-adj trials shared the same object pairings, *cup – sock* and *ball – shoe*. Thus, children saw these objects a lot more often than *baby – doggie* and *kitty – birdie* during Vanilla trials, *car – book* during I-verb trials, and *cookie – juice* during U-verb trials. Image repetition likely caused decreased attention to the target during I-adj and U-adj trials.

As for U-verb trials, the source of decreased looking to the target compared to Vanilla trials is less obvious. Again, one possibility lies in the nature of the target-distracter object pairings. Both pairings in U-verb trials consist of inanimate objects, whereas those in Vanilla trials consist of animals. Perhaps, looking time differences between U-verb and Vanilla trials simply reflect relative differences in interest to inanimacy *versus* animacy, whereby the latter is more interesting than the former.

Verb Onset

Comparison of mean proportion looking to the target following verb onset across I-verb and U-verb trials indicated that preterm participants were more apt to use verb

information to orient to the target than their full-term counterparts. Though within trial type comparisons indicated that preterm children spent a greater proportion of time looking at the target during I-verb trials and full-term children spent a greater proportion of time looking at the target during U-verb trials, these within trial group differences were not statistically significant. However if these differences do in fact reflect a real effect, then they could be indicative of attentional differences between full-term and preterm toddlers. However, the analyses made thus far of the present data set are not sufficient to make any claims regarding this issue.

Regardless, the critical analysis to consider for this study is the relative difference in mean proportion looking to the target between I-verb and U-verb trials. These data showed while both full-term and preterm children looked at the target longer during I-verb than U-verb trials, this difference was only statistically significant for the preterm group. If greater proportion looking to the target following verb onset in I-verb than U-verb trials is truly a reflection of greater accuracy, then these results suggest that the preterm group is processing informative language cues more efficiently than the full-term group.

When considering these conclusions, there is an important caveat to note regarding the window of analysis chosen to calculate proportion looking to the target. Within each verb trial, the window of analysis for target looking following verb onset significantly overlapped the window of analysis for target looking following target noun onset. This overlap may subsequently be masking priming effects of the verb on the target noun because it includes a significant proportion of time when the participant is no longer uncertain about the identity of the target. This does not seem far-fetched given that

mean proportion looking within each type of verb trial was not significant across full-term and preterm participants; yet, we found significant increased looking to the target during I-verb trials for preterm children. Perhaps decreasing the window of analysis following verb onset would reveal the same priming effect in our full-term group.

Adjective Onset

Evaluation and comparison of looking behavior during I-adj and U-adj trials suggested that neither full-term nor preterm participants made use of informative adjectives to guide their looking to the target. Following adjective onset, both groups of children had similar mean proportion looking to the target across I-adj and U-adj trials. However, the lack of differences in target looking across trial types could, again, be an artifact of the selected window of analysis. Like in verb trials, the window used to calculate mean proportion looking to the target following adjective onset significantly overlapped the window of analysis of target looking following target noun onset. In fact, the overlap was a lot greater for these trials than for verb trials, as the adjectives were adjacent to the target nouns. Thus, the potential to overshadow the effects of subjects' adjective processing was even greater for I-adj than for I-verb trials. Again, limiting the window of analysis following adjective onset (so that it ends prior target noun onset) would likely reveal significant differences in looking behavior across I-adj and U-adj trials.

Another interesting point to make regarding accuracy analyses is that the means for proportion looking to the target following the onset of uninformative verbs and adjectives were a lot higher than would be expected by chance. Again this is likely

attributable to the overlap between the windows of analyses for target looking following the onset of verbs or adjectives and the onset of target nouns. One of our future goals is to reduce the window of analyses for target looking following verb and adjective onsets so that they end prior to the onset of the target noun. This change would enable us to use a more sensitive measure to determine whether 26-month-olds can use informative verbs and adjectives to direct their looking behavior.

3.2.4.2 Speed of Target Word Recognition

Noun Onset

Like mean accuracy measures following target noun onset, mean RT following target noun onset did not reveal statistically significant differences in the speed of target word recognition between full-term and preterm children. That said, preterm children tended to have shorter mean RT's than their full-term counterparts: these participants took less time to orient to the target on three of the five trial types (i.e., I-verb, U-verb, and I-adj trials). Thus, the proposed hypothesis of reduced language processing efficiency in preterm children is unsupported by the present data: these children are processing speech just as efficiently, if not more so, than their full-term counterparts. Additionally, unlike accuracy data, these data did not show significant RT differences across trial types. This finding was as expected.

Verb Onset

While mean accuracy measures following verb onset suggested that preterm subjects used informative verbs more efficiently than full-term subjects, mean RT values

following verb onset indicated the opposite outcome. Full-term children were significantly faster at recognizing the target following informative than uninformative verbs. Though preterm children also had shorter mean RT's for I-verb than U-verb trials, these differences were not statistically significant. These results suggest that preterm participants were not able to use verb information as successfully as their full-term counterparts. These conclusions are incongruent with those drawn from mean accuracy data. What can explain this discrepancy across measures? Perhaps longer looking to the target following the onset of a word of interest is an inadequate conceptualization of accuracy during word processing. Longer looking to the target could simply indicate that a participant has a harder time disengaging her attention from a particular visual stimulus once she has fixated it. If this were the case, then such a measure would have nothing to do with language processing skill. A better measure of word recognition accuracy would be the object the participant orients to first, target or distracter, following the onset of a word of interest.

RT measures, unlike mean proportion looking to the target, are a more direct way of examining whether verb cues are being utilized to guide target looking. After all, these measures are relatively instantaneous; rather than summarizing looking over a long period of time, during which many cognitive processes can occur, RT's capture a participant's immediate reaction to a stimulus. If mean proportion looking to the target is an inadequate measure of language processing, then previously discussed accuracy data for verb trials do not justify concluding that preterm infants recognize target words more accurately than their full-term counterparts. Given that RT measures are a better measure of language processing skill, then full-term participants' shorter latencies to the target

following the onset of informative verbs suggest that they use verb information to guide their looking behavior during an online word-processing task more efficiently than preterm children.

Adjective Onset

Like mean RT data following noun onset, data following adjective onset suggest that there are no significant group differences in the speed of target word recognition in either I-adj or U-adj trials. Furthermore, though both full-term and preterm participants tended to have shorter RT's for informative than uninformative adjective trials, these cross-trial differences were not statistically significant for either group of children. These findings suggest a couple of things. First, neither full-term nor preterm children seem to be exploiting adjective information to orient to the target more quickly. Why are children using verb but not adjective information to guide their looking behavior? One explanation could lie in the type of adjectives used in this study: colors (i.e., red and blue). Though an informal poll taken among participants' parents suggested that their children had some understanding of these color terms, it could be that participants had not fully conceptualized the meaning of these adjectives. After all, knowledge about colors at this age is the product of explicit pedagogy. For example, a parent might sit with her child, hold up a red crayon and say, "*This is a red crayon*". At this point the child has limited information regarding what "*red*" can refer to: is it limited to that particular crayon or can it be used to describe other objects? If adjectives are harder to conceptualize and process than verbs, then it is not surprising that this study did not yield significant RT differences across I-adj and U-adj trials for either group of participants. The other critical

point to draw from adjective trials is that preterm participants were just as efficient as their full-term counterparts at using adjective information during an online word-processing task.

3.2.4.3 General Remarks

Results from this experiment showed preterm participants had similar patterns of target looking behavior as their full-term counterparts. Though accuracy data was hard to interpret, they did show that preterm toddlers were just as accurate as full-term toddlers following target noun onset, and like full-term toddlers they showed greater accuracy during I-verb than U-verb trials and equal accuracy during I-adj and U-adj trials. Similarly, RT data showed that, following target noun onset, preterm participants oriented to target objects just as quickly as full-term participants, and, like full-term participants, they took less time to orient to the target during informative than uninformative verb and adjective trials. Where differences did arise across these groups was in the magnitude of RT differences between informative and uninformative verb trials. Specifically, full-term participants were faster than preterm participants. Similarities in patterns of behaviors in conjunction with differences in speed of target object recognition across these groups could suggest that poor language outcomes associated with preterm birth are not language-specific but rather result from deficits in speed of processing.

Chapter 4: Cognitive Skills as Modulators of Language Development

Preterm children across the board, regardless of gestational age, exhibit a high incidence of cognitive delays, impairments and behavioral problems (Baron & Rey-Casserly, 2010; Kapellou et al., 2006; Peterson et al., 2000; Rushe et al., 2001). The nature and severity of these deficits span a wide range. While many of the more severe disabilities (i.e., moderate to severe mental retardation, neurosensory hearing loss and blindness, epilepsy and cerebral palsy) result from focal brain injuries identified early in infancy, many low-severity disabilities (i.e., learning disabilities, ADHD, borderline mental retardation, specific neuropsychological deficits such as visual motor integration and behavioral disorders) arise without any obvious brain insult and may not become apparent until children are of school age (Aylward, 2002).

The severity of developmental issues associated with prematurity is inversely proportional to gestational age and birthweight: every additional week spent *in utero* or gram at birth significantly reduces the chances for cognitive impairments (Baron & Rey-Casserly, 2010).

By age 30 months, at least half of all surviving infants born extremely preterm (≤ 25 weeks) are estimated to have neurodevelopmental impairments, and by age 6 years 40% of these children exhibit cognitive deficits (as referenced by Kapellou et al., 2006).

Critically, at least 33% of those born very or moderately preterm also end up developing neurocognitive or behavioral issues. Additionally, reduced birthweight, which co-varies with gestational age, also has a negative impact on development. Approximately 12% to 32% of very low birthweight (VLBW, $\leq 1,500$ g) children are prone to

neuropsychological and behavioral problems and are eventually diagnosed with major neurodevelopmental handicaps.

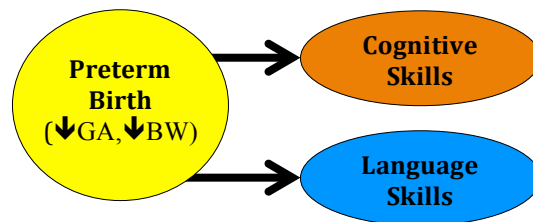
Though poor outcomes are often attributed to neonatal health complications, evidence suggests that even preterm children with low-risk infancies experience considerable cognitive deficits and educational difficulties. For example, IQ scores for VLBW children have been reported to be nearly 15 points below those of their age-matched full-term counterparts. Additionally, more than 50% of preterm children with low-risk infancies require special education resources, and by age 8 years, about 20% have repeated a grade. To complicate matters, despite an increased survival rate among VLBW children (i.e., 85%) their development is poorly understood. These numbers pose a major public health concern as the nature of the neurocognitive deficits afflicting preterm children are varied even when outcomes are similar (Peterson et al., 2000). This makes it especially difficult to develop and implement intervention strategies that can effectively (and universally) help preterm children.

While it is clear from the statistics reviewed above that prematurity puts children at risk for poor cognitive outcomes, the studies reporting these numbers do not specify what aspects of cognition are affected by prematurity. Thus, it remains unknown whether the cause of these cognitive issues is operating at a local or global level. This distinction is critical when considering causal explanations of why language is susceptible to delays and deficits in preterm children. At least three plausible explanations can account for the poor cognitive and linguistic outcomes observed among preterm children. First, prematurity could negatively impact the developmental trajectories of cognition and language independently of one another (see Figure 13a). Second, prematurity could cause

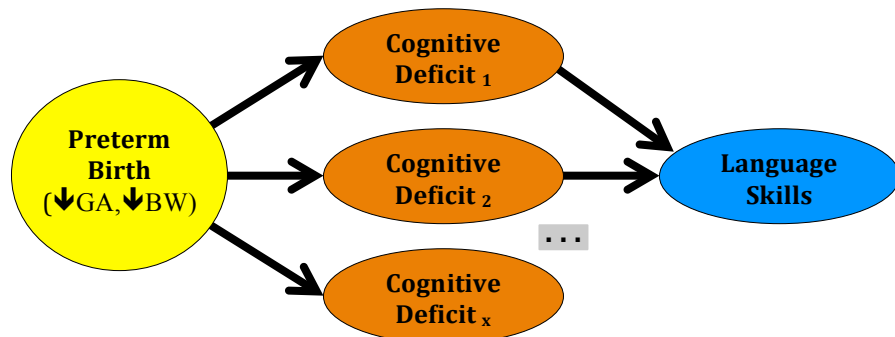
deficits in specific cognitive skills that mediate language development; in this case, prematurity affects language indirectly (see Figure 13b). Third, prematurity could impact certain aspects of language directly and others indirectly via effects on specific cognitive skills (see Figure 13c). Given that cognition and language co-exist in the brain and are unlikely to function independently of one another, the latter two explanations are more plausible.

Figure 13. Potential causal pathways for poor language outcomes.

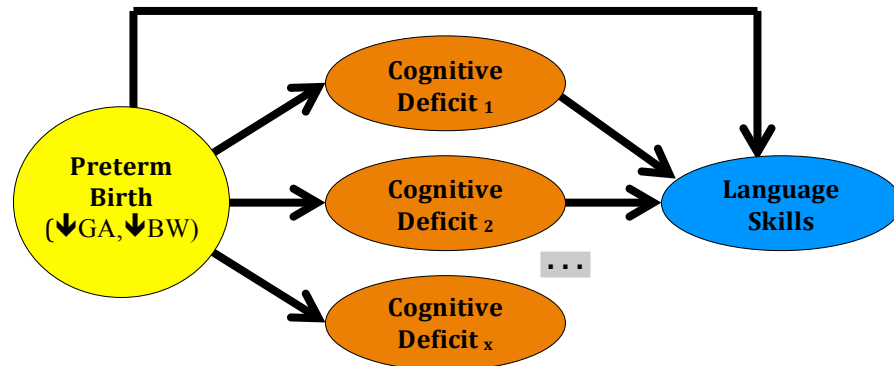
(a) Prematurity independently affects cognitive and language outcomes.



(b) Prematurity indirectly affects language by altering specific aspects of cognition. Under this model, not all cognitive deficits interfere with language. Additionally, the degree to which each deficit impacts language could vary. For the sake of simplicity, language skills are summarized in one oval. It is possible that specific cognitive deficits only impact a limited subset of linguistic abilities.



(c) Prematurity impacts language outcomes both directly and indirectly. Under this model, preterm birth could result in specific cognitive deficits that subsequently affect language skills. At the same time, preterm birth could directly affect specific language abilities.



In fact, Rose, Feldman and Jankowski (2009) provide evidence for cognition as a mediating factor in language development. These researchers attempted to disentangle the relationship between prematurity, cognition and language by tracking abilities in specific cognitive domains and relating these to both concurrent and later language outcomes. Specifically, these researchers wanted to see if memory, processing speed, attention and representational competence at 12 months corrected age were predictive of scores on the WG-MCDI at 12 months and the Peabody Picture Vocabulary Test (PPVT) at 36 months corrected age. These four cognitive domains are of particular interest because they relate to information processing, a skill that is important for successful speech processing and language acquisition.

Rose et al. found that measures of memory and representational competence, but not of processing speed and attention, were correlated with language outcomes at both 12 and 36 months corrected age. Surprisingly, not all measures of memory and

representational competence were related to comprehension and production measures. In the case of memory, measures of immediate visual recognition, delayed recognition, and elicited recall significantly correlated with language outcomes at both 12 and 36 months. On the other hand, short-term memory capacity was not associated with language at either age. In the case of representational competence, measures of cross-modal transfer, object permanence, and symbolic play significantly correlated to language outcomes at 12 and 36 months. However, measures of anticipation did not relate to language at either age. Subsequent regression analyses revealed that information processing skills at 12 months accounted for 31% and 17% of the variance seen in scores of language comprehension and production at 36 months, respectively. Recognition memory, elicited recall and cross-modal transfer each contributed uniquely to comprehension; cross-modal transfer, alone, had unique predictive power on production.

Based on these findings, Rose et al. concluded that language development is guided by domain-general information processing abilities. Prematurity is likely to negatively impact language outcomes indirectly by triggering specific cognitive deficits that make it more difficult to learn language. Specifically, this study's findings suggest that representational competence and memory are likely targeted by prematurity. Rose et al. explain that these two aspects of cognition are critical to language learning: representational competence is needed to successfully draw connections between words and their corresponding objects and actions, and memory is needed to consolidate these abstract representations.

Rose and colleagues did not find any relationship between attention, processing speed, and language. These results are unexpected given that attention and processing

speed are assumed to be important for normal language development. For instance, attention is necessary during parent-child interactions that could result in word-learning situations. Specifically, when a parent points to and names a novel object, an infant or child must attend to what the parent is pointing to in order to successfully learn the new word. At a more basic level, a parent is more likely to engage with an infant or child that appears to be attentive. Speed of processing is equally important in word-learning situations. For example, faster processing can facilitate word segmentation in fluent speech by giving quick access to familiar words. This subsequently allows the novice language learner to focus on and learn unfamiliar words in the input. Though attention and processing speed likely play critical roles in language learning, Rose et al. probably did not find correlations between these two measures and language outcomes because these information-processing skills were assessed indirectly. Specifically, the researchers drew conclusions about attention and processing speed from looking-time data collected from tasks designed to evaluate memory and representational competence. Regardless, Rose and colleagues' findings provide support for the idea that prematurity could, in part, impact language outcomes via effects on specific cognitive skills.

The following portion of the present longitudinal project aimed at comparing cognitive skill in full-term and preterm participants. Three specific cognitive measures were taken, immediate visual recognition memory, visual reception, and fine motor skill, with two goals in mind. The first goal was to assess possible interactions between these specific cognitive skills and the receptive language abilities measured between 4-26 months. The second goal was to explore whether these specific cognitive measures

explain some of the variability in language outcomes typically observed across full-term and preterm children.

Immediate visual recognition memory was assessed at 12 months using methods employed by Rose and colleagues, and visual reception and fine motor skills were assessed at 26 months using the Mullen Scales of Early Learning. The mediating roles of these cognitive skills on measures of receptive language skills between 4-26 months and language outcomes at 14.5 and 26 months were evaluated using structural equation modeling. These findings are discussed in Chapter 6, which is in Part III Relations Between Language, Cognitive And Environmental Predictors, And Later Language Outcomes.

4.1 Rose Task: Assessing Immediate Visual Recognition Memory at 12 Months

In this experiment full-term and preterm children's immediate visual recognition memory was measured using the Rose Task, a visual paired comparison (VPC) procedure that compares memory for abstract patterns and faces. This task was chosen for a three reasons. First, it is well validated as it has been used to test both full-term and preterm infants at various ages. Second, Rose et al. (2009) found strong correlations between immediate visual recognition abilities at 12 months and language outcomes at 36 months. Third, measures of visual recognition memory in infancy, as evaluated by VPC, have been shown to predict later cognitive abilities. Specifically, outcomes have been shown to predict childhood IQ. Thus, the Rose Task was used because it is a good proxy for general cognitive ability (especially given that IQ cannot be measured in infancy and

toddlerhood), and because it can provide insight on the relationship between participants' cognitive ability and language outcomes.

The Rose Task is comprised of two phases: familiarization and test. During familiarization, the participant is shown a pair of identical images, either two faces or two abstract patterns, until she reaches a predetermined criterion, which, at 12 months, is 10 seconds for faces and 3 seconds for abstract patterns. At test, the participant is shown the familiarized image paired with a novel image of the same category. Looking time to each image is measured during the 10-second duration of this phase. If the participant successfully recognizes the familiarized image, she should show differential looking to the old and new image. Under normal circumstances, typically developing infants show a novelty preference, meaning that they spend more time looking at the novel image than at the familiarized one.

The purpose of the present study was to quantify potential differences in visual recognition memory abilities between full-term and preterm groups. Given that preterm participants have shown different patterns of attention than their full-term counterparts on many of the receptive language experiments reviewed thus far, the expectation for this experiment was that they would again behave differently than full-term participants. Critically, regardless of birth status, participants at risk for language delays were expected to show atypical patterns of recognition memory. For instance, if recognition memory does play a predictive role in language outcomes, then children at risk for delays might display a familiarity (rather than the expected novelty) preference.

4.1.1 Participants

Participants included 21 preterm (12 females, mean corrected age= 369.5 days, range= 360-382 days) and 42 full-term (17 females, mean age= 368.4 days, range= 355-386 days) 12-month-old infants. One additional full-term child was tested but excluded from data analyses due to fussiness. Chronologically, preterm infants were 70.8 days older than their full-term counterparts (mean chronological age= 439.2 days, range= 397-479 days).

4.1.2 Method

Testing was completed in a sound-attenuated room. The participant sat on her caregiver's lap, positioned approximately 90 cm away from a pair of adjacent monitors with a light centered between them. The caregiver wore noise-cancelling headphones and listened to instrumental music to discourage interactions that might impact the participant's behavior. The experimenter observed the child's looking behavior over a closed circuit video system and recorded it on a digital camcorder at 30 frames per second. Videos for each participant were later coded off-line using in-house software to determine how much time was spent looking at each image during each trial.

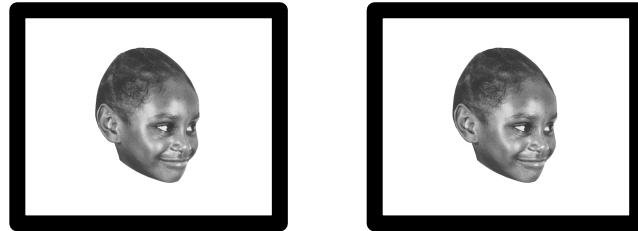
Trials consisted of two phases: familiarization and test (see Figure 14). The latter was divided into two blocks to counterbalance side of novel image presentation. Each trial began with the center light flashing. Once the infant oriented toward it, the light was extinguished and identical faces (or abstract patterns) appeared on the monitors for familiarization. This phase continued until the infant fixated the images for a total of 10 seconds in the case of faces or 3 seconds in the case of abstract patterns (see Figure 14a).

These familiarization times were used because they are thought to be the minimum exposure time needed for each type of stimuli for 12-month-olds to succeed during an immediate visual recognition task (see Rose et al., 2001 for citations). Image pairs were displayed for no more than 30 seconds, regardless of whether the participant reached the looking criterion.

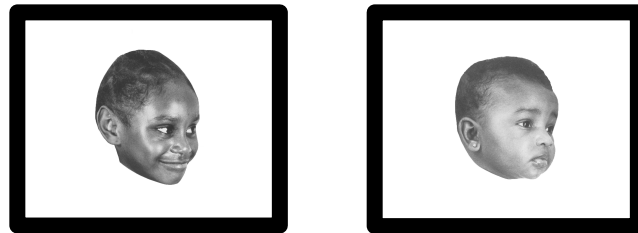
Once familiarization was complete, the test phase was initiated with the center light flashing until the infant's attention was redirected to the midline. At this point a new pair of images consisting of the familiarized image and a novel image appeared on the monitors (see Figure 14b). Familiarized faces were always paired with novel faces and familiarized abstract patterns were always paired with novel abstract patterns. After 5 seconds, the images disappeared, and the center light flashed again until the participant oriented towards the midline. The familiarized and novel images then reappeared on opposite monitors (see Figure 14c). This was done to control for possible side preferences. After another 5 seconds, the images disappeared and the trial ended.

Figure 14. Three part composition of a Rose Task trial: (a) familiarization, (b) test block 1 and (c) test block 2.

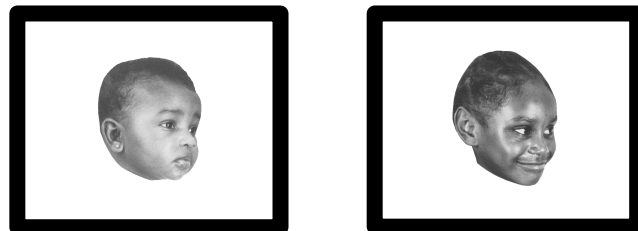
(a) Familiarization consisted of two identical images displayed on the monitors for 10 seconds for faces and 3 seconds for abstract patterns.



(b) Test block 1 consisted of the familiarized image paired with a novel image. The pair was displayed on the monitors for 5 seconds.



(c) Test block 2 reversed the order of presentation of the familiarized and novel images. The stimulus pair again remained on the monitors for 5 seconds.



4.1.2.1 Stimuli

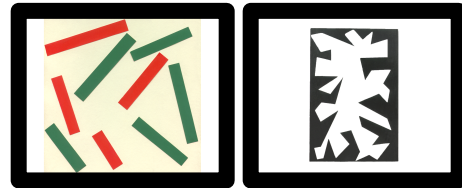
Abstract pattern and face pairs were the same as those used by Rose, Feldman and Jankowski (2001, 2009). Infants were presented with a total of nine trials; five were face trials and four were abstract pattern trials. Faces were black-and-white photographs and abstract patterns were multi-colored and black-and-white hand drawn designs. In the original study, Rose et al. wanted to minimize “procedural variation that could contribute to differences across infants” (p.138, 2001), which they accomplished by fixing the order of trial presentation and by keeping constant the stimuli used as familiar and novel in each pairing. The present study replicated this set-up. For specific stimulus pairings and order of trial presentation refer to Figure 15.

Figure 15. Test phase stimulus pairings for faces (left) and abstract patterns (right). For the sake of simplicity the familiarized image is presented on the right side and the novel image on the left side; side presentation for familiar and novel images was counterbalanced during the study. Pairings are given in the order in which they were presented to infants, with abstract pattern trials interspersed between face trials.

TRIAL 1



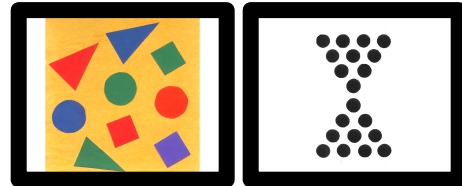
TRIAL 2



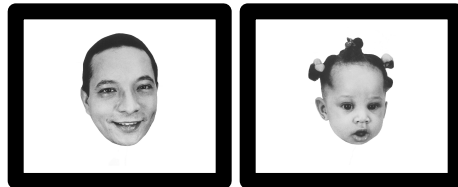
TRIAL 3



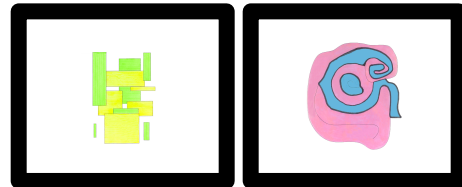
TRIAL 4



TRIAL 5



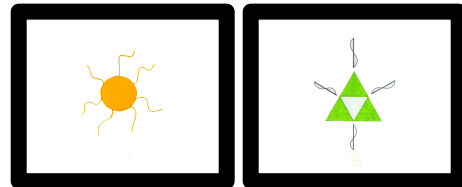
TRIAL 6



TRIAL 7



TRAIL 8



TRIAL 9



4.1.2.2 Measures of Interest

Looking behavior was coded off-line frame-by-frame (1 frame= 33 ms) using the Metcalf Infant Lab Coder program. Three measures were collected for each trial's familiarization phase. These included the average duration of all looks (*mean look*), the average duration of looks away from the target images (*mean pause*), and the total amount of time the stimuli had to be displayed in order for familiarization criteria to be reached (*exposure time*).

One additional measure was calculated for each trial's test phase: the difference in looking time between the familiarized and novel images (*novelty score*)¹³. Novelty scores were computed by subtracting looking time to the familiarized image from looking time to the novel image. A positive score indicated preference for the novel image, whereas a negative one indicated preference for the familiarized image. Novelty scores were first determined for each test block, and then calculated for the sum of the two blocks (i.e., composite novelty score). Blocks with less than 25% looking to the presented images were excluded from block analyses. However, these blocks were included in composite analyses as long as they yielded at least 25% looking to the stimuli when added to the other block within a trial. Thus, composite novelty scores retained more of the data collected during the testing session.

¹³ Rose and colleagues (2001) calculated novelty scores as the percentage of total looking time spent on the novel image. We instead decided to use difference scores so as to maintain consistency with the measures calculated for the other experiments comprising this longitudinal study.

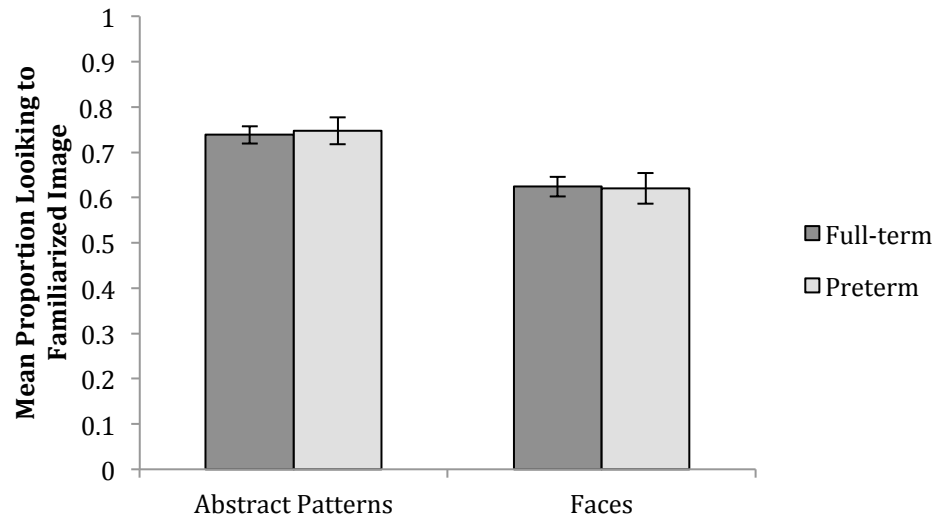
4.1.3 Results

4.1.3.1 Familiarization Phase

Mean Look

The mean proportion looking times within each stimulus group were very similar across full-term and preterm participants; both groups of infants spent a higher proportion of time looking at abstract patterns than at faces (see Figure 16). Mean proportion looking to abstract patterns was 0.738 (SD= $\pm$.125) and 0.747 (SD= $\pm$.136) for full-term and preterm infants, respectively. Mean proportion looking to faces was 0.624 (SD= $\pm$.140) and 0.620 (SD= $\pm$.156) for full-term and preterm participants. A 2 (stimulus type) by 2 (birth status) repeated measures ANOVA validated a highly significant main effect of stimulus type, $F(1, 61)= 28.3, p< .001$; the interaction between stimulus type and birth status was not significant, $F(1, 61)= .072, p= .789$. Follow-up paired-samples t-tests confirmed that both full-term and preterm infants had significantly higher mean proportion looking time for abstract patterns than for faces, $t(41)= 4.14, p< .001$, and $t(20)= 3.92, p= .001$, respectively.

Figure 16. Mean proportion looking times to abstract patterns and faces during familiarization for full-term (dark grey bars) and preterm (light grey bars) participants. Error bars represent standard error.

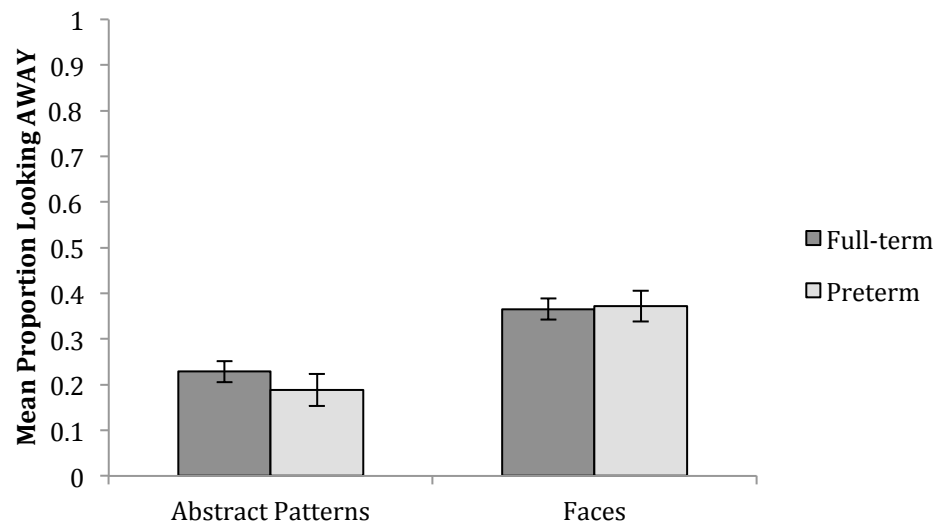


Mean Pause

The mean proportion of time spent looking away from the screens (i.e., mean pause) during familiarization was similar across full-term and preterm participants; infants spent a greater proportion of time looking away during face presentations than during abstract pattern presentations (see Figure 17). Mean pause for abstract patterns was 0.228 (SD= $\pm$.148) and 0.188 (SD= $\pm$.161) for full-term and preterm participants, respectively. Mean pause for faces was 0.365 (SD= $\pm$.151) and 0.372 (SD= $\pm$.155) for full-term and preterm infants, respectively. A 2 (stimulus type) by 2 (birth status) repeated measures ANOVA confirmed a highly significant main effect of stimulus type, $F(1, 61)= 44.1, p< .001$; the interaction between stimulus type and birth status was not significant, $F(1, 61)= .937, p= .337$. Follow-up paired-samples t-tests verified that both

full-term and preterm infants spent a greater proportion of time looking away during face familiarization trials than during abstract pattern familiarization trials, $t(41) = -4.69$, $p < .001$, and $t(20) = -5.23$, $p < .001$, respectively.

Figure 17. Mean proportion of time spent looking away (i.e., not at the screens) during familiarization with abstract patterns and faces for full-term (dark grey bars) and preterm (light grey bars) participants. Error bars represent standard error.

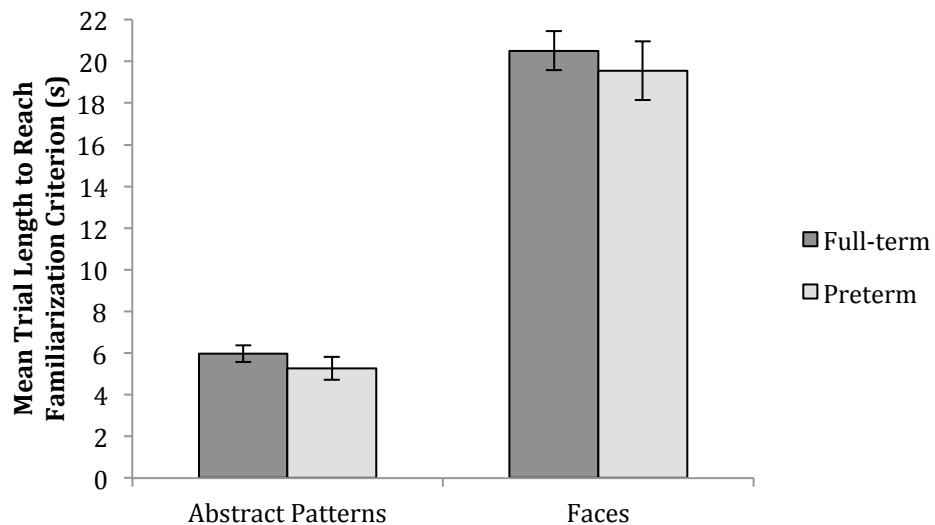


Exposure Time

The mean trial duration (i.e., exposure time) required for infants to reach the 3- and 10-second familiarization criteria for abstract patterns and faces was again similar across full-term and preterm participants (see Figure 18). On average, it took full-term infants almost 3 more seconds to attain the 3-second familiarization period for abstract patterns (mean exposure time = 5.97 s, $SD = \pm 2.55$ s) and almost 11 more seconds to reach the 10-second familiarization period for faces (mean exposure time = 20.5 s, $SD = \pm 6.02$ s).

Preterm infants took almost 2.5 more seconds to reach the abstract pattern criterion (mean exposure time= 5.27 s, SD= ± 2.51 s) and almost 10 more seconds to attain the face criterion (mean exposure time= 19.6 s, SD= ± 6.45 s). Independent-samples t-tests confirmed that mean exposure time was not significantly different across groups for either abstract patterns, $t(61)= 1.04$, $p= .305$, or faces, $t(61)= .580$, $p= .564$.

Figure 18. Mean trial length in seconds required to attain 3- and 10-second familiarization criteria for abstract patterns and faces, respectively. Data for full-term and preterm participants are depicted in dark grey and light grey bars, respectively. Error bars represent standard error.



4.1.3.2 Test Phase

Novelty Scores

Results indicated that both full-term and preterm participants had a preference for novel abstract patterns and faces during both test blocks. However, preterm participants had higher novelty scores for abstract patterns than for faces, while full-term infants had

higher novelty scores for faces than for abstract patterns. Mean novelty scores and standard deviations for each type of stimulus are summarized for each group of infants on Table 5; data are provided for block 1, block 2, and their composite scores. Figures 19a, 19b, and 19c provide graphical representations of these data.

A 2 (stimulus type) by 2 (block number) by 2 (birth status) repeated measures ANOVA revealed a significant main effect of block, $F(1, 61) = 8.98, p = .004$, and a significant interaction between stimulus type and birth status, $F(1, 61) = 8.52, p = .005$. The block main effect confirms that novelty scores for patterns and faces were different across blocks. Planned comparisons using paired-samples t-tests revealed that this main effect was driven by a significant drop in full-term infants' mean novelty scores for patterns across blocks, $t(41) = -3.39, p = .002$. Though full-term participants' mean novelty scores for faces and preterm participants' mean novelty scores for patterns and faces also decreased across blocks, these differences did not reach statistical significance, $t(41) = -.976, p = .335$, $t(20) = -1.40, p = .176$, and $t(20) = -1.11, p = .281$, respectively.

Follow-up 2 (stimulus type) by 2 (birth status) repeated measures ANOVAs were conducted for each block to examine the basis for the interaction between stimulus type and birth status. Results for block 1 did not yield a significant main effect of stimulus type, $F(1, 61) = .053, p = .819$, or of the stimulus type by birth status interaction, $F(1, 61) = 2.25, p = .139$. Again, block 2 analysis did not yield a significant main effect of stimulus type, $F(1, 61) = .565, p = .455$; however, this time, the interaction between stimulus type and birth status was significant, $F(1, 61) = 5.31, p = .025$. This confirms that full-term and preterm participants showed different patterns of novelty scores for abstract patterns and faces. Follow-up independent-samples t-tests revealed that this interaction was driven by

a significant difference in abstract pattern novelty scores, $t(61)=2.17, p=.034$: preterm infants had higher scores than their full-term counterparts. Additional paired-samples t -tests for block 2 indicated that full-term participants' novelty scores for faces were significantly higher than those for patterns, $t(41)=2.72, p=0.010$. Though preterm participants' mean novelty scores were not significantly different across stimulus types, $t(20)=-.903, p=.377$, paired-samples correlations showed that these scores were negatively correlated with one another, $r=-.466, p=.033$. Thus, preterm infants who had higher novelty scores for patterns tended to have lower novelty scores for faces.

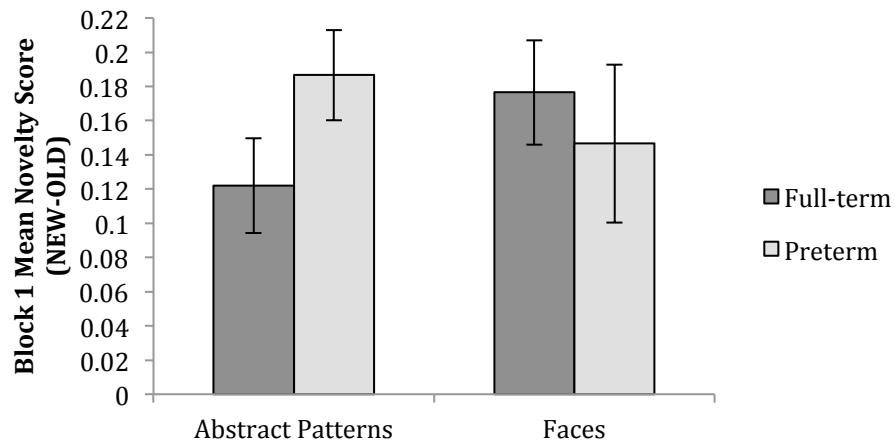
A 2 (stimulus type) by 2 (birth status) ANOVA was also run on the mean composite novelty scores. Composite scores were examined because they were used in SEM analyses, and they also allowed greater data retention. Results yielded no significant main effect of stimulus type, $F(1, 61)=.331, p=.567$. The interaction between stimulus type and birth status was again significant, $F(1, 61)=8.36, p=.005$. Follow-up independent-samples t -tests showed that at the root of this interaction was a significant difference in abstract pattern scores, $t(61)=2.37, p=.021$, and a marginally significant difference in face scores, $t(61)=-1.86, p=.068$, across participant groups. That is, preterm children had larger novelty scores for patterns than full-term children, and full-term children had larger novelty scores for faces than preterm children. Additional paired-samples t -tests showed that composite novelty scores across types of stimuli were significantly different for full-term, $t(41)=2.91, p=.006$, but not preterm, $t(20)=-1.52, p=.144$, children. Though preterm novelty scores for abstract patterns and faces were not different at the statistical level, there is a clear tendency for higher pattern novelty scores than face novelty scores. This is evident in both testing blocks.

Table 5. Summary of block 1 (blue columns), block 2 (pink columns) and composite (purple columns) mean novelty scores and standard deviations for abstract patterns and faces for full-term and preterm participants.

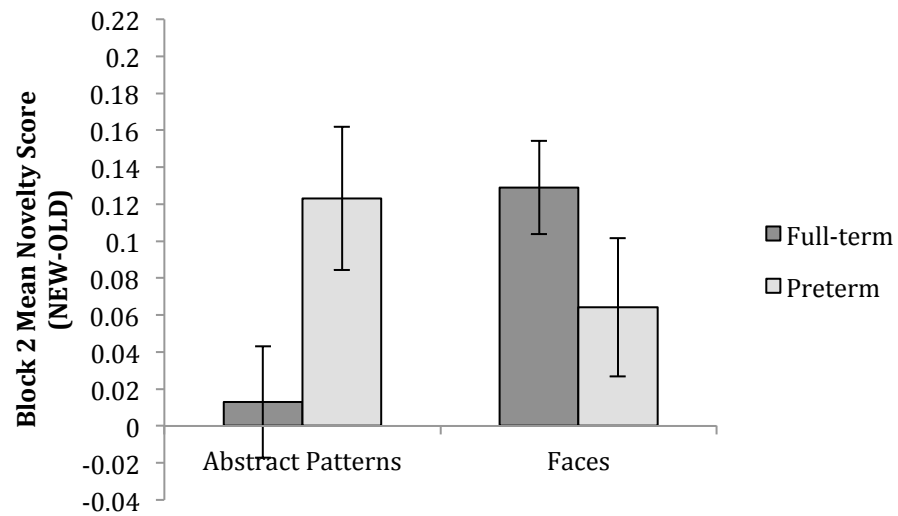
	Block 1 Mean Novelty Scores		Block 2 Mean Novelty Scores		Composite Mean Novelty Scores	
	Patterns	Faces	Patterns	Faces	Patterns	Faces
Full-Term	0.122 SD= $\pm$.180	0.176 SD= $\pm$.198	0.0129 SD= $\pm$.195	0.129 SD= $\pm$.164	0.0758 SD= $\pm$.124	0.146 SD= $\pm$.0891
Preterm	0.187 SD= $\pm$.121	0.147 SD= $\pm$.212	0.123 SD= $\pm$.178	0.0642 SD= $\pm$.172	0.150 SD= $\pm$.100	0.103 SD= $\pm$.0793

Figure 19. Mean novelty scores for abstract patterns and faces at test for full-term (dark grey bars) and preterm (light grey bars) participants for (a) block 1, (b) block 2 and (c) the composite of blocks 1 and 2. Error bars represent standard error.

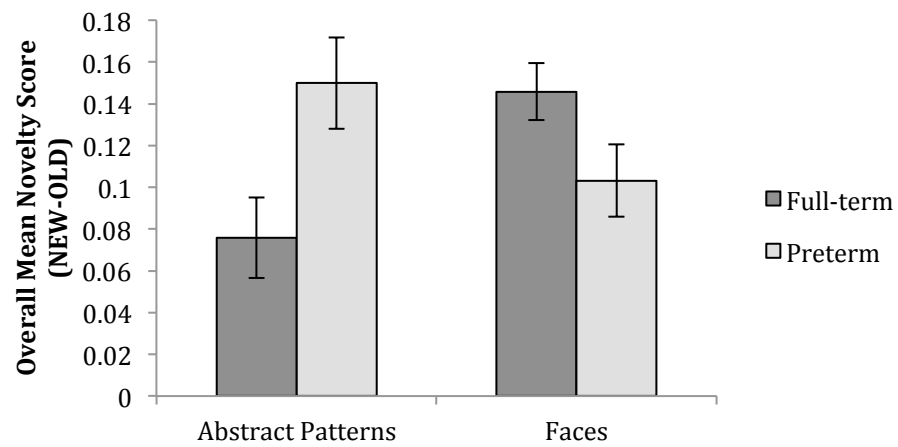
(a)



(b)



(c)



4.1.4 Discussion

This study assessed full-term and preterm children's visual recognition memory using the Rose Task, a VPC procedure that compares memory for abstract patterns and faces.

It has been well documented that, when presented with a VPC task, a typically developing infant will look longer at the novel than at the familiarized object during test. This is interpreted to mean that the child remembers the familiarized object, as she can discriminate it from the new object (Rose et al., 2001). The present findings show that, like their full-term counterparts, preterm children exhibited the expected novelty preference: they looked longer to the novel patterns and faces than to the familiarized ones. This is consistent with what Rose and colleagues (2001) found with their age corrected 12-month-olds. Critically, the observed preference for novel patterns and faces suggested that recognition memory is intact among the preterm sample tested in this study.

Like Rose et al. (2001), the present study also found a significant interaction between novelty scores for stimulus type and birth status, thus showing a difference in performance for abstract patterns and faces across full-term and preterm children. That said, the pattern of scores detected was opposite of that reported by Rose and colleagues: while their full-term participants had higher abstract pattern novelty scores than preterm participants, the preterm participants tested in this study had higher abstract pattern novelty scores than full-term participants. More importantly, the present data show a dissociation between novelty scores across types of stimuli. Specifically, full-term infants had significantly higher novelty scores for faces than for patterns, while preterm infants tended to have higher novelty scores for patterns than for faces. Importantly, these differences cannot be attributed to looking behavior during familiarization as mean proportion looking times to and away from familiarized patterns and faces were the same within stimulus type across full-term and preterm participants.

The dissociation in novelty scores presently observed is a novel finding, and could signal potential differences in how full-term and preterm children pay attention to face *versus* non-face stimuli. Specifically, mismatches in attentional patterns across subject groups could be tapping into differences in local *versus* global *versus* configural visual processing. Local processing refers to the analysis of the components making up an object. Global and configural processing are similar in that they refer to the analysis of the object as a whole; what distinguishes them is whether the spatial arrangement of the components is important for the identity of the object as a whole. Specifically, in global processing the relative position of the parts is not critical to identification, whereas in configural processing these interrelations are essential for object identity. For example, when considering faces, seeing eyes, a mouth, and a nose is sufficient to realize that one is looking at a face (i.e., global processing). However, these features are not sufficient to distinguish between two similar faces; in this case, one would have to rely on spatial relationships between the eyes, mouth, and nose to differentiate between different faces (i.e., configural processing). Thus, intact configural processing is necessary for successful facial recognition.

Critically, Santos and colleagues (2009) have shown that healthy 6- to 9-year-olds born very preterm exhibit atypical configural processing compared to their full-term counterparts. When presented with a shape-matching task that can be interpreted at the local or configural level (e.g., a 6x8 rectangular array composed of equidistantly spaced triangles), full-term children prefer to match at the configural level (i.e., they select a 6x8 rectangular array composed of equidistantly spaced circles as the match), while preterm children prefer to match at the local level (i.e., they pick a 3x4 rectangular array

composed of equidistantly spaced triangles as the match). In other words, preterm individuals show a bias for local processing. Given that face recognition requires configural processing and preterm individuals are more likely to have a predisposition for local processing, then it is not surprising that preterm sample tested in the present study had higher novelty scores for non-face stimuli.

Preterm participants' relatively smaller face novelty scores can also be interpreted as a reduced interest in faces compared to their full-term counterparts. This is atypical, as even newborns, who have limited experience with faces, have a visual preference for face and face-like stimuli over non-face stimuli (Johnson et al., 1991). A disinclination to attend to faces could subsequently have major implications for how preterm children learn about the world around them. This is particularly relevant to language learning, as an extensive body of work indicates that infants' attention to faces, especially to the mouth, makes a significant contribution to later language outcomes (e.g., Young et al., 2009). Furthermore, decreased attention to faces could significantly impact social development. If preterm children are paying less attention to faces than their full-term counterparts, then they are less likely to efficiently learn from and use information made available by other people.

Of course, these explanations are just speculation. The novelty score dissociation observed across types of stimuli in full-term and preterm participants may have simply been the product of the stimuli used. All the face stimuli used during this experiment were black-and-white whereas the abstract pattern stimuli were a combination of colorful and black-and-white designs. Perhaps preterm participants simply found the more colorful and variable abstract patterns more interesting than the monochromatic faces.

Regardless of the explanation there is an undeniable difference between full-term and preterm participants' attention to face *versus* abstract pattern stimuli, and subsequent experimental work is needed to investigate the validity of the proposed explanations.

In summary, this work shows that preterm participants tested in this experiment do not have issues with recognition memory. Like their full-term counterparts, these children had positive novelty scores indicating that they could distinguish between familiarized and novel images. Though preference for the novel stimuli is considered a typical developmental pattern, incongruities were observed in how preterm children paid attention to different types of stimuli. Specifically, these participants exhibited a looking bias for abstract patterns, while full-term participants preferred faces. These differences in attentional patterns across participant groups are yet another indication of how preterm children are paying attention to the world differently than their full-term counterparts. What remains to be seen is whether participants' novelty scores on this VPC task are related to their language outcomes, as suggested by Rose and colleagues' (2009) work. It is possible that novelty scores close to or below zero (i.e., cannot distinguish old from new or prefer the old image) are indicative of poorer language outcomes.

4.2 Mullen Scales of Early Learning: Assessing Visual Reception and Fine Motor Skills at 26 Months

In addition to visual recognition memory assessments at 12 months, subjects' performance on the cognitive subscales of the Mullen Scales of Early Learning AGS Edition (MSEL) was also collected at 26 months. This standardized test was selected over the more commonly used Bayley Scales of Infant Development because the latter was

used to evaluate most of this study's preterm participants during follow-up doctor's appointments on at least two occasions. Thus, the MSEL was used to minimize potential priming effects that could boost performance within the preterm group.

The MSEL is divided into five sections evaluating different aspects of a child's development, including gross motor, visual reception, fine motor, receptive language, and expressive language abilities. Cognitive testing in the present samples was limited to the visual reception (MVRS) and the fine motor (MFMS) subscales. The MVRS primarily evaluated participants' abilities to match 2- and 3-D shapes (e.g., putting a triangle puzzle piece into its corresponding place on a form board) as well as more complex objects (e.g., asking the child to find a specific flower among five different flowers). Other tasks involved sorting objects according to their shape (e.g., putting forks in one box and blocks in another box) and remembering the identity of a previously shown object (e.g., finding the clown the child had been previously familiarized with). The MFMS assessed children's abilities to manipulate objects in ways requiring fine motor skills. For example, children were asked to turn pages one at a time, stack blocks, draw vertical and horizontal lines, and string beads on a shoelace.

While the predictive value of early visual reception skills for later cognitive outcomes may seem obvious, fine motor abilities in infancy/toddlerhood are not typically thought of as meaningful predictors of later cognition. However, research is increasingly showing that there is a strong link between these skills and cognition. For instance, Grissmer and his team of researchers (2010) showed that measures of fine motor skills collected around kindergarten entrance were highly correlated with math and reading abilities in the fifth grade. These researchers proposed that the link between motor and

cognitive skills stems from bidirectional neural connections between the cerebellum/basal ganglia and the prefrontal cortex, which historically, were thought to be exclusively involved in motor and cognitive activities, respectively. Furthermore, Grissmer and colleagues review evidence suggesting that motor and cognitive development occur concurrently and facilitate general learning processes. Specifically, they argue that (a) many types of cognitive skills rely on “specialized control and modulation functions located in the cerebellum and basal ganglia that develop during motor acquisition” (p. 1013), and that (b) part of the neural infrastructure that guides and controls learning during motor development also controls learning during cognitive development. Additionally, developmental gains in motor skills afford the child new ways of interacting with and learning about the world around her, which can subsequently promote changes in cognitive development.

In addition to evaluating visual reception and fine motor skills in preterm and full-term subjects, the relationship between these skills and novelty scores on the Rose Task at 12 months was explored. This was accomplished by running correlational analyses between scores on the MVRS and the MFMS and block 1, block 2, and composite novelty scores on the Rose Task.

4.2.1 Participants

Subjects included 18 preterm (12 female) and 39 (18 female) full-term participants. MVRS data could not be collected for two preterm participants because of refusal to cooperate during testing (n=1) and an inability to return to the lab to complete testing (n= 1). Since completion of MVRS and MFMS typically involved two visits to the

lab, the average testing age was not equivalent across subscales; with few exceptions, visits were scheduled no more than two weeks apart from one another. For MVRS, the mean testing age was 795.5 days (range= 777-860 days) and 790.9 days (range= 776-819 days) for preterm and full-term participants, respectively; chronologically, preterm participants were 69.6 days older than their full-term peers (mean chronological age= 860.5 days, range= 811-923 days). For MFMS, the mean testing age was 794.4 days (range= 777-860 days) and 789.8 days (range= 776-819 days) for preterm and full-term subjects, respectively; chronologically, preterm participants were 70.5 days older than their full-term peers (mean chronological age= 860.3 days, range= 811-923 days).

4.2.2 Method

Testing was conducted in a small room with minimal visual stimulation to minimize subjects' distractibility. The participant sat at a table across from the experimenter, either next to or on her caregiver's lap. The MVRS and MFMS have 33 and 30 test items, respectively, with items becoming increasingly complex as testing progresses. Since both of these subscales are designed to assess participants from infancy to early childhood, their start and end points are dependent on the subject's age and level of maturity: older children start with more complex items, and testing continues until participants score zeros on three consecutive test items. Participants in this study began testing on item 16 for the MVRS and item 13 for the MFMS. The maximum number of points that can be scored on the MVRS and the MFMS are 50 and 49, respectively. Note that the MSEL uses different scoring scales across its five subscales, thus mean

performance scores on the MVRS and the MFMS cannot be compared to one another. The scores reported in the following section are raw scores.

4.2.3 Results

4.2.3.1 MVRS and MFMS Outcomes

Mean scores on MVRS and the MFMS were 32.4 points (SD= ± 4.08 points) and 26.1 points (SD= ± 2.10 points) for preterm subjects, and 33.4 points (SD= ± 5.16 points) and 26.1 points (SD= ± 2.00 points) for full-term subjects, respectively. Independent-samples t-tests confirmed that cross-group differences for each of these mean scores were not significant, $t(53) = .671, p = .505$ for MVRS and $t(55) = -.007, p = .994$ for MFMS.

4.2.3.2 Correlations Between Novelty Scores at 12 Months and Scores on the MVRS and MFMS at 26 Months

Bivariate correlations between novelty scores for abstract patterns and faces at 12 months and scores on the MVRS at 26 months indicated that there were no significant relationships between these two types of cognitive measures. Similarly, there were no significant bivariate correlations between novelty scores for abstract patterns and faces at 12 months and scores on the MFMS at 26 months. Table 6 summarizes all correlations tested between novelty scores and scores on the MVRS and the MFMS.

Table 6. Summary of correlations between cognitive outcome measures collected with the Rose Task at 12 months and the MVRS and MFMS collected at 26 months. All correlations are Spearman correlations.

			12 MONTHS					
			Abstract Pattern Novelty Scores			Face Novelty Scores		
			Block 1	Block 2	Composite	Block 1	Block 2	Composite
26 MONTHS	MVRS Score	<i>r</i> <i>N</i>	-.130 50	.006 50	-.067 50	.051 50	-.067 50	-.063 50
	MFMS Score	<i>r</i> <i>N</i>	-.101 51	-.019 51	-.009 51	-.175 51	-.012 51	-.131 51

4.2.4 Discussion

The goals of this study were twofold. First, it evaluated and compared visual reception and fine motor skills in preterm and full-term participants at 26 months. Second, it explored potential predictive relationships between measures of visual recognition memory at 12 months and measures of visual reception and fine motor skills at 26 months.

Results from the MVRS and the MFMS indicated that visual reception and fine motor skills were comparable across preterm and full-term participants. This means that any observable (and meaningful) differences in language outcomes across preterm and full-term groups are unlikely attributable to differences in visual reception or fine motor skills.

Interestingly, no significant correlations were found between novelty scores for abstract patterns and faces at 12 months and scores on the MVRS and MFMS at 26 months. The lack of predictive relationships between these cognitive proxies could have

resulted from a number of things. First, measures of immediate visual recognition memory, visual reception, and fine motor skills were collected at different ages, and the absence of correlational relationships between them could reflect a discontinuity in cognitive development. That is, the development of visual recognition memory, which was measured earlier in life, does not affect the development of either visual reception or fine motor skills. A second, more plausible explanation for the lack of significant relationships between novelty scores, visual reception scores, and fine motor scores is that each of these measures is tapping into a unique aspect of cognition. This would be beneficial should significant relationships be found between any of these measures of cognitive skill and measures of language outcome. Specifically, the presence of such relationships (or lack thereof) could provide insight on what aspects of cognition are important for typical language learning, and could help establish the pathways from cognition to language that are especially at risk for atypical development in preterm children (as well as other children at risk for poor language outcomes). Knowledge on these matters is important for developing intervention strategies that tap into and focus on resolving the source of the problem.

PART II

MEASURES OF LANGUAGE OUTCOMES IN THE

SECOND AND THIRD YEARS OF LIFE

Chapter 5: Measurements of Child Language Outcomes

Measures of language outcomes were collected at two time points using three assessment tools: the MacArthur-Bates Communicative Development Inventories- Words and Gestures (WG-MCDI) at 14.5 months, and the MacArthur-Bates Communicative Development Inventories- Words and Sentences (WS-MCDI), and the Mullen Receptive Language Subscale (MRLS) at 26 months. The WG- and WS-MCDI are parental reports, whereas the MRLS is composed of a set of tasks administered in the lab by an experimenter or clinician. Both the WG- and WS-MCDI are widely used and highly validated tools in the field of child language acquisition. The MRLS and other subscales of the Mullen Scales of Early Learning AGS Edition are primarily used by educators and clinicians to evaluate language comprehension and production skills, as well as cognitive and motor development in infancy and early childhood. Each of these assessment tools is further specified below. Correlations of the numerous language outcome measures are reviewed in the final portion of this section.

5.1 Language Outcome Assessment Tools

5.1.1 Parental Surveys

5.1.1.1 WG-MCDI

The WG-MCDI is a checklist primarily used to assess language comprehension and production in children between the ages of 8 and 16 months. This tool is composed of two major sections: Early Words, and Actions and Gestures.

Early Words is divided into four subsections: (1) First Signs of Understanding (i.e., does the child respond to (a) her name, (b) “no-no”, and (c) “there’s mommy/daddy”?), (2) Phrases Understood, (3) Starting to Talk (i.e., does the child (a) imitate others’ speech, and (b) label objects?), (4) Vocabulary Checklist (i.e., quantification of words (a) understood, and (b) produced).

Actions and Gestures measures infants’ early and later gestures, and is divided into three subsections: (1) Early Gestures, (2) Later Gestures, and (3) Total Gestures (i.e., a composite of early and later gestures). Early gestures include first communicative gestures (e.g., pointing, nodding head to indicate “yes”, shaking head to indicate “no”, requesting something by extending arm, shrugging to indicate “all gone”, etc.), and games and routines (e.g., playing peekaboo, singing, dancing, etc.). Later gestures include actions with objects (e.g., eating with a spoon or fork, putting on a sock or shoe, putting telephone to ear, throwing a ball, stirring pretend liquid in a cup or pan with a spoon, etc.), pretending to be a parent (e.g., putting a doll to bed, feeding a doll with a bottle, talking to a doll, etc.), and imitating other adult actions (e.g., sweeping with a broom, watering plants, writing with a pen or pencil, etc.).

5.1.1.2 WS-MCDI

The WS-MCDI is a checklist primarily used to assess language production at the later stages of language acquisition, between the ages of 16 to 30 months. Unlike the WG-MCDI this tool does not measure the number of words understood by the child. The WS-MCDI is composed of two sections: Words Children Use, and Sentences and Grammar. Words Children Use specifically assesses the number of words produced by

the child, whether or not she talks about past events, future events, and objects and people that are not present, and understands requests for objects that are not present. Sentences and Grammar measures the mean length of the child's three longest utterances (M3L), and use of word endings (e.g., plural “-s”, possessive “ ’s”, progressive “-ing” and past tense “-ed”), irregular noun and verb forms, overregularization of nouns and verbs, and grammatical complexity (e.g., “*I no do it*” vs. “*I can ’t do it*”).

5.1.2 Lab Assessment

5.1.2.1 MRLS

The MRLS is one of the five subscales comprising the Mullen Scales of Early Learning. This standardized tool, which was also used to evaluate participants' visual reception and fine motor skills, can be used to test children between the ages of 1 to 45 months. The MRLS specifically measures language comprehension via 33 test items that evaluate children's knowledge on topics that are likely encountered at home and everyday life. Like the MVRS and the MFMS, the items on the MRLS are arranged in order of progressive difficulty. Topics tested include: body part recognition (e.g., “*where is the baby's mouth?*”), picture identification (e.g., “*can you point to the doll?*”), auditory spatial awareness (e.g., “*can you put the teddy next to the table?*”), following one or more commands (e.g., “*can you get that box and bring it to me?*” vs. “*can you give me the crayon and put the (toy) car on the chair?*”), color identification (e.g., “*which one is the red crayon?*”), comparative concepts (e.g., “*can you point to the two pictures that are the same?*”), etc. The MRLS also has sliding start and end points: older, more developed children start with increasingly difficult test items, and testing continues

until zeros are scored on three consecutive test items. Children tested in the present study started testing on item 18 (i.e., recognition of body parts). A maximum of 48 points can be scored on the MRLS.

5.1.3 Participants

5.1.3.1 WG-MCDI

Surveys were received from 20 preterm (11 females, mean corrected age= 444.7 days, range= 427-488 days) and 39 full-term (19 females, mean age= 445.1 days, range= 426-463 days) toddlers. Data were collected from an additional eight full-term participants, but were excluded from analyses because, by the final lab visit at 26 months, these children had delayed speech and/or were receiving early intervention services for language (n= 6), were at risk for autism and had extremely delayed speech (n= 1), or were diagnosed with a brain malformation (n= 1). On average, preterm participants were 69.8 days older than their full-term counterparts (mean chronological age= 514.9 days, range= 460-563 days).

5.1.3.2 WS-MCDI

Reports were collected from 20 preterm (13 females, mean corrected age= 790.9 days, range= 777-814 days) and 36 full-term (16 females, mean age= 787.2 days, range= 770-805 days) participants. An additional six surveys were completed among full-term participants, but excluded from analyses because these children had delayed speech and/or were receiving early intervention services for language development (n= 6).

Chronologically, preterm participants were 72.3 days older than full-term participants (mean chronological age= 860.2 days, range= 832-895 days).

5.1.3.3 MRLS

Participants included 15 preterm (9 females, mean corrected age= 796.1 days, range= 784-817 days) and 33 full-term (16 females, mean age= 793.4 days, range= 778-812 days) children. Five additional full-term participants were tested but excluded from analyses because they had delayed speech and/or were receiving early intervention services for language (n= 5). Chronologically, preterm participants were 68.7 days older than their full-term counterparts (mean chronological age= 862.1 days, range= 814-925 days).

5.1.4 Method

5.1.4.1 WG-MCDI

The WG-MCDI was given to parents during their children's 14.5-month visit to the lab. During this visit, children were tested on a receptive language experiment that assessed their ability to learn similar-sounding words; the results to this additional experiment are not reported in this paper. Parents were instructed to complete the language survey at home and to mail it back to the lab no later than two weeks after their visit to the lab. In addition to the WG-MCDI, parents were given a set of guidelines for filling out the language survey. The experimenter verbally reviewed these instructions with parents to ensure they understood how to assess their child's language comprehension and production (see Appendix A for the set of guidelines provided).

5.1.4.2 *WS-MCDI*

The WS-MCDI was sent to parents a couple of weeks prior to their children's 26-month visit to the lab. Parents were asked to complete the survey as close as possible to the day of their appointment and to bring the completed form with them at the time of their visit. The decision to give them the survey *before* (rather than during) their final visit to the lab was made in an effort to increase the chances of receiving the survey in a timely fashion. Parents were again provided with guidelines for filling out the WS-MCDI (see Appendix B for the set of instructions provided).

5.1.4.3 *MRLS*

The MRLS was conducted at 26 months as part of the final set of studies for the longitudinal project. The decision to test toddlers' receptive language skills in the lab was made in an effort to explore the extent to which lab observations would corroborate parental reports. Additionally, the MRLS provides a measure of receptive language, which the WS-MCDI, unlike the WG-MCDI, does not provide. Thus, the MRLS made it possible to also have a language comprehension outcome measure at 26 months.

Testing was conducted in a small room with limited visual stimulation to help focus participants' attention. Children sat at a table across from the experimenter, either next to or on their parents. Given the numerous tasks participants were tested on at 26 months, the MRLS was typically completed over the course of two visits to lab.

5.1.5 Results

5.1.5.1 WG-MCDI

Mean scores and standard deviations for the eight language outcome measures (i.e., First Signs of Understanding, Phrases Understood, Imitation and Labeling, Words Understood, Words Produced, Early Gestures, Later Gestures and Total Gestures) collected at 14.5 months are summarized on Table 7. Independent-samples t-test analyses comparing full-term and preterm participants indicated no significant group differences in any of the mean outcome scores (see Table 7 for t-test results). That said, full-term infants did have a tendency to outscore preterm infants despite being chronologically younger and as a result having less exposure to language. Full-term participants had higher mean scores on six of the eight language measures, including Phrases Understood, Imitation and Labeling, Words Understood, Words Produced, Later Gestures, and Total Gestures.

Many of these group mean differences were fairly small, which is unsurprising given that most of the measures are scored on very few items. For instance, scores for First Signs of Understanding, Phrases Understood, and Imitations and Labeling are each only based on 3, 28, and 2 items, respectively. Scores based on such small numbers of items may not be sufficient to fully capture the variability in language development across full-term and preterm participant groups, especially at 14.5 months when a lot of variability is already expected in typically developing populations. This held true among this study's full-term and preterm groups who showed lots of within group variability on scores for Words Understood and Words Produced. Unlike other subsections of the WG-MCDI, these measures are each based on almost 400 items, making them better suited to

capture variability across subjects. To specify, the score ranges for Words Understood were 29 to 317 and 30 to 242¹⁴ for full-term and preterm participants, respectively, and those for Words Produced were 0 to 105 and 0 to 113 for full-term and preterm children, respectively. A combination of measures with small maximum scores and great within group variability are likely contributing to the lack of statistically significant differences observed across full-term and preterm groups.

Table 7. Summary of the means, standard deviations, and t-test results comparing full-term and preterm participants' scores on the eight language outcomes measured by the WG-MCDI at 14.5 months. Measures corresponding to children's Early Words appear in light green columns, whereas those for Actions & Gestures appear in dark green columns. The total possible score for each outcome measure is provided in parentheses.

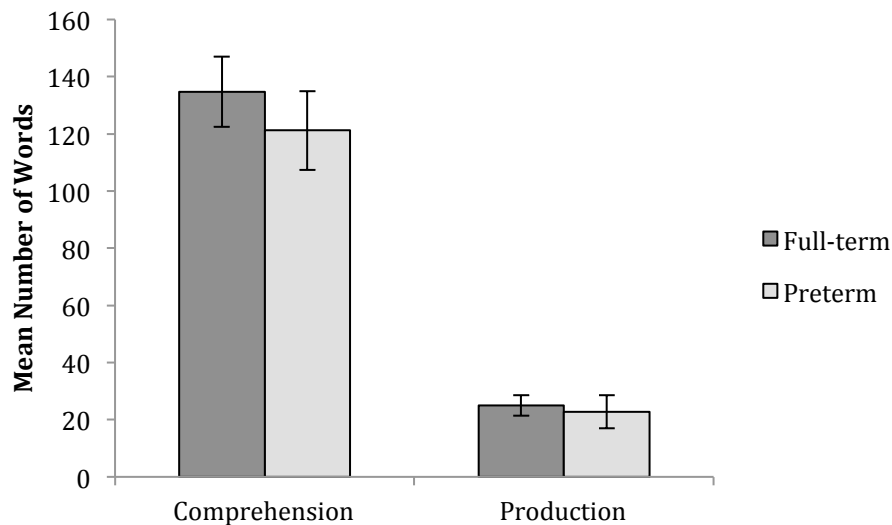
	EARLY WORDS					ACTIONS & GESTURES		
	1st Signs of Understanding (3 pt. max.)	Phrases Understood (28 pt. max.)	Imitation, Labeling (2 pt. max.)	Words Understood (396 pt. max.)	Words Produced (396 pt. max.)	Early Gestures (18 pt. max.)	Later Gestures (45 pt. max.)	Total Gestures (63 pt. max.)
Full-term	2.97 SD= $\pm$.160	17.5 SD= $\pm$ 7.05	1.51 SD= $\pm$.683	134.8 SD= $\pm$ 76.70	24.9 SD= $\pm$ 22.3	13.2 SD= $\pm$ 2.02	21.4 SD= $\pm$ 8.14	34.6 SD= $\pm$ 9.11
Preterm	3.00 SD= $\pm$ 0	16.6 ¹⁵ SD= $\pm$ 5.84	1.25 SD= $\pm$.851	121.2 ¹⁵ SD= $\pm$ 60.20	22.8 SD= $\pm$ 25.9	13.2 SD= $\pm$ 2.66	21.3 SD= $\pm$ 7.52	34.5 SD= $\pm$ 9.03
T-Test Results	$t(57) = -.713$ $p = .479$	$t(56) = .471$ $p = .639$	$t(57) = 1.29$ $p = .204$	$t(56) = .675$ $p = .503$	$t(57) = .331$ $p = .742$	$t(57) = .006$ $p = .995$	$t(57) = .051$ $p = .960$	$t(57) = .046$ $p = .964$

¹⁴ When accounting for the eight full-term and one preterm participants excluded from statistical analyses, the score ranges for Words Understood were 4 to 317 and 30 to 357 for full-term and preterm groups, respectively.

¹⁵ One preterm subject's scores for Phrases Understood and Words Understood were excluded from analysis because these were outliers. The parent reported that the child understood all 28 phrases and 357 of the 396 words listed on the survey even though she also reported that the child produced zero words.

Results from follow-up 2 (Words Understood and Words Produced) by 2 (birth status) repeated measures ANOVA confirmed that, as expected at this age, both full-term and preterm participants understood significantly more words than what they produced, main effect $F(1, 56) = 123.5, p < .001$, interaction $F(1, 56) = .458, p = .501$ (see Figure 20).

Figure 20. Mean number of words comprehended and produced by full-term (dark grey bars) and preterm (light grey bars) participants at 14.5 months. Error bars represent standard error.



5.1.5.2 WS-MCDI

Mean scores and standard deviations for each of the seven outcome measures (i.e., Words Produced, Use of Words to Express the Past, Future, and Absence, Use of Word Endings to Express Plurals, Possessives, Progressives, and the Past Tense, Use of Irregular Words, Overregularized Word Production, Grammatical Complexity, and M3L) evaluated by the WS-MCDI at 26 months are outlined in Table 8. These showed that full-

term toddlers outscored preterm toddlers on six of these measures, including Words Produced, Use of Words to Express the Past, Future, and Absence, Use of Word Endings to Express Plurals, Possessives, Progressives, and the Past Tense, Use of Irregular Words, Grammatical Complexity, and M3L. The one measure that was higher for preterm toddlers, Overregularized Word Production, is a representation of errors produced. Thus, full-term participants outperformed their preterm peers on all measures of the WS-MCDI. However, independent-samples t-tests showed that the only measure to approach a statically significant difference across groups was M3L (see Table 8 for t-test results). That is, preterm toddlers produced fewer morphemes per utterance than their full-term counterparts.

Table 8. Summary of the means, standard deviations, and t-test results comparing full-term and preterm participants' scores on the seven language outcome measures assessed by the WS-MCDI at 26 months. Measures corresponding to Words Children Use and Sentences & Grammar appear in light and dark green columns, respectively. Where appropriate, the total possible score for each outcome measure is provided in parentheses.

	WORDS CHILDREN USE		SENTENCES & GRAMMAR				
	Words Produced (680 pt. max.)	Past, Future, Absence Words (5 pt. max.)	Word Endings (4 pt. max.)	Irreg. Word Production (25 pt. max.)	Overreg. Word Production (45 pt. max.)	Grammatical Complexity (37 pt. max.)	M3L
Full-term	408.3 SD= $\pm$ 154.8	4.78 SD= $\pm$.59	3.19 SD= $\pm$ 1.31	8.83 SD= $\pm$ 6.30	1.72 SD= $\pm$ 2.95	16.9 SD= $\pm$ 11.2	6.03 SD= $\pm$ 2.87
Preterm	368.1 ¹⁶ SD= $\pm$ 137.3	4.40 SD= $\pm$.99	2.80 SD= $\pm$.95	6.75 SD= $\pm$ 3.39	2.25 SD= $\pm$ 2.61	11.8 SD= $\pm$ 10.1	4.73 SD= $\pm$ 2.10
T-Test Results	$t(53)= .965$ $p= .339$	$t(26.6)= 1.55$ $p= .132$	$t(54)= 1.19$ $p= .241$	$t(54)= 1.61$ $p= .113$	$t(54)= -.667$ $p= .508$	$t(54)= 1.72$ $p= .092$	$t(54)= 1.77$ $p= .083$

¹⁶ The Words Produced score for one preterm subject was an outlier, and thus excluded from analysis. The parent reported that the child produced 614 of the 680 words listed on the survey. Prior to removal of this data point, the mean for Words Produced by preterm participants was 380.4 words (SD= $\pm$ 144.6 words).

Though the observed cross group differences in mean scores were not statistically significant, the WS-MCDI outcome data collectively suggest a tendency for poorer language outcomes in preterm toddlers (see Figures 21, 22, 23, 24, and 25). Though expected, this tendency is interesting given that preterm participants were chronologically more than two months older than their full-term peers, and thus, should be outperforming full-term toddlers on at least some language measures. More importantly, this finding suggests that, it may not be appropriate to stop correcting for age after the second birthday, a practice that is typically used by medical professionals when evaluating preemies.

The lack of statistically significant differences across full-term and preterm group outcomes may be attributed to a couple of different factors. First, the number of preterm participants surveyed was small, and as a group, much more heterogeneous than their full-term counterparts. Second, like the WG-MCDI, the outcome measures collected with the WS-MCDI showed great variability, which may have subsequently overshadowed the measures' effect sizes. For example, the ranges for Words Produced for full-term and preterm participants were 59 to 657 and 74 to 548¹⁷, respectively. Nonetheless, the data are in the predicted direction.

¹⁷ When accounting for the six full-term and one preterm participants excluded from analyses, the ranges Words Produced were 31 to 657 and 74 to 614 for full-term and preterm toddlers, respectively.

Figure 21. Mean number of words produced by full-term (dark grey bars) and preterm (light grey bars) children at 26 months. The checklist consisted of 680 words. Error bars represent standard error.

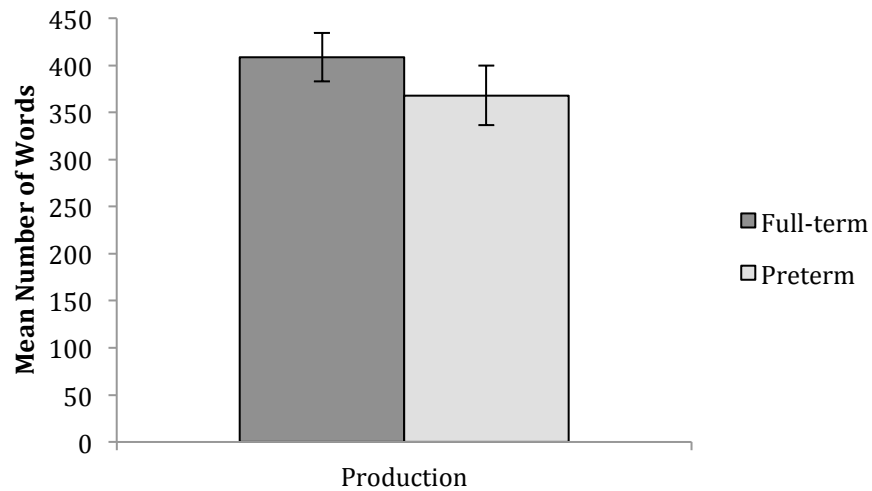


Figure 22. Mean use of words expressing Past, Future, and Absence, and mean use of Word Endings including plural *-s*, possessive *'s*, progressive *-ing*, and past tense *-ed* at 26 months. Participants could score up to five points on the former and four points on the latter. Error bars represent standard error.

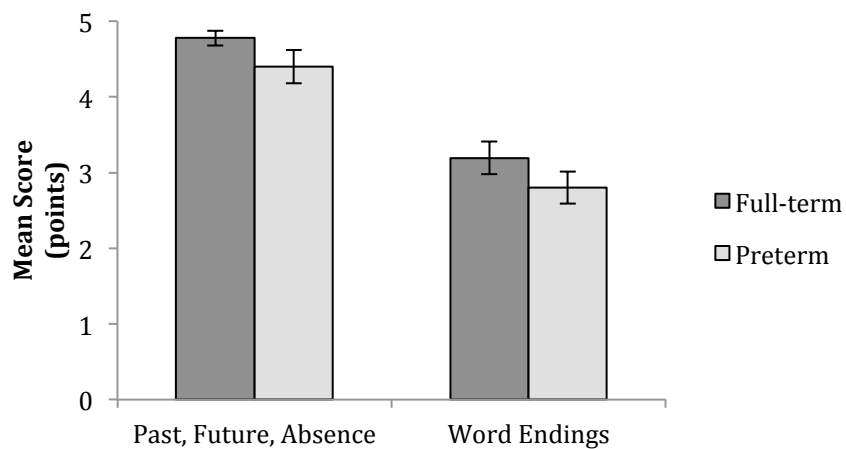


Figure 23. Means for Irregular Words and Overregularized Words used by full-term (dark grey bars) and preterm (light grey bars) toddlers. Children could be given credit for up to 25 words on the former and 45 words on the latter. Error bars represent standard error.

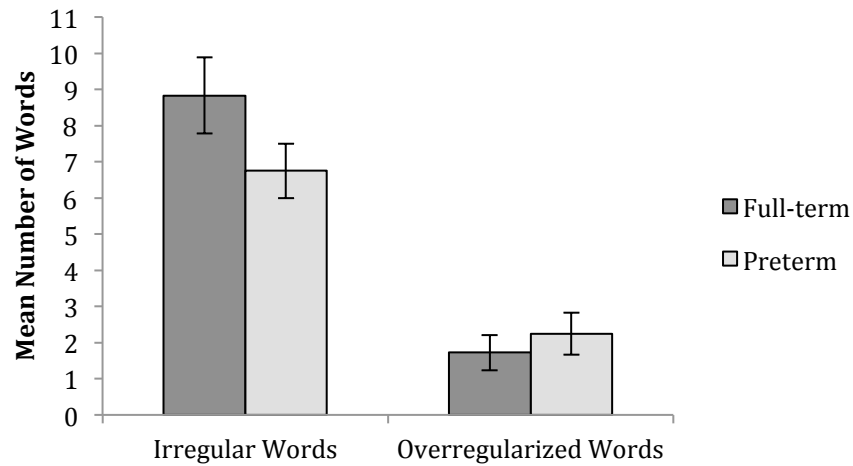


Figure 24. Mean number of more Grammatically Complex forms used by full-term (dark grey bars) and preterm (light grey bars) children. Parents were given 37 items, each with a simple and a more complex version of a sentence, and asked to pick which one sounded more like what their child might say. One point was awarded each time the more complex form was marked. Error bars represent standard error.

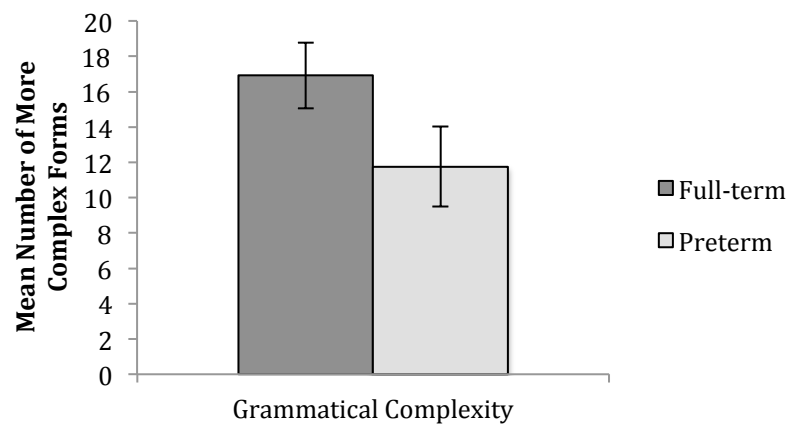
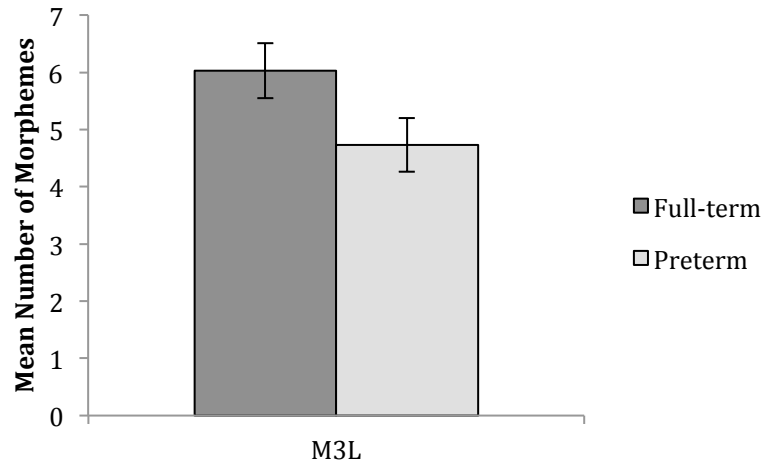


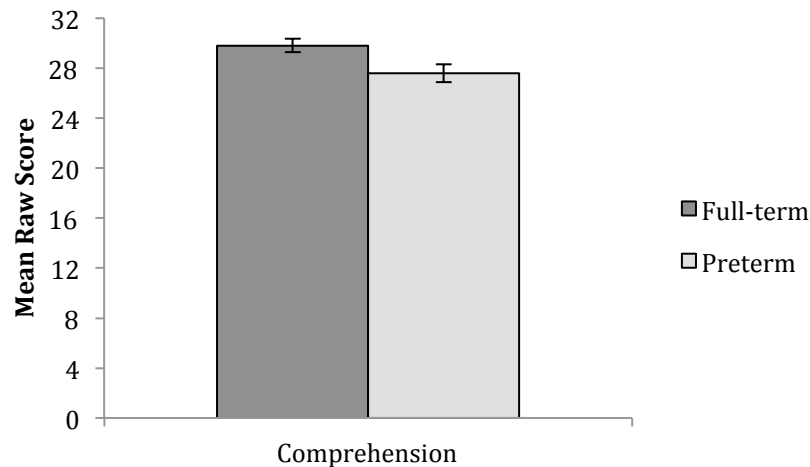
Figure 25. Mean Length of the Three Longest Utterances produced by full-term (dark grey bars) and preterm (light grey bars) subjects. Utterance length was measured as a function of the number of morphemes making up the utterance. Error bars represent standard error.



5.1.5.3 MRLS

The data show that full-term toddlers (mean= 29.8 points, SD= ± 3.07 points) outscored preterm toddlers (mean= 27.6 points, SD= ± 2.75 points) on the MRLS by 2.2 points (see Figure 26). An independent-samples t-test indicated that this difference in mean raw scores was statistically significant, $t(46)= 2.40$, $p= .021$. Cross group mean differences remained significant even after the data were reanalyzed with the five full-term participants (mean_{all full-term}= 29.5 points, SD= ± 3.05 points) that had delayed speech or were receiving early intervention services, $t(51)= 2.10$, $p= .041$. Despite the observed language delays among some full-term participants, the MRLS was still able to capture differences in language comprehension abilities across full-term and preterm participants; as expected, preterm participants had lower comprehension scores than their full-term peers.

Figure 26. Mean raw score on the MRLS for full-term (dark grey bars) and preterm (light grey bars) toddlers surveyed at 26 months. Error bars represent standard error.



5.1.5.4 Correlations Between Language Outcomes

Bivariate correlations were run to assess relationships between the different language outcome measures collected for participants across ages. Results indicated that many of the WG-MCDI measures collected at 14.5 months were highly correlated with many of the WS-MCDI measures collected at 26 months. Specifically, Imitation and Labeling at 14.5 months correlated with Word Production, Use of Words to Express the Past, Future, and Absence, Use of Word Endings to Express Plurals, Possessives, Progressives, and the Past Tense, Use of Irregular Words, and Grammatical Complexity at 26 months; Comprehension at 14.5 months correlated with Word Production at 26 months; Word Production at 14.5 months correlated with Word Production, Use of Word Endings to Express Plurals, Possessives, Progressives, and the Past Tense, Use of Irregular Words, Grammatical Complexity, and M3L at 26 months; and Early Gestures at

14.5 months correlated with Word Production, Use of Irregular Verbs, and Grammatical Complexity at 26 months. All statistically significant correlations were positive, indicating that higher scores at 14.5 months were predictive of higher scores at 26 months. For a complete summary of all possible correlations between WG-MCDI measures at 14.5 and WS-MCDI measures at 26 months refer to Table 9. A high number of correlations between language outcomes at 14.5 and 26 months suggests that parental reports were consistent across ages. Furthermore, these correlations indicate that measures of language early in development are predictive of later language outcomes, despite the high level of variability observed both within and across the sampled groups.

*Table 9. Summary of correlations between language outcome measures collected with the WG-MCDI at 14.5 months and the WS-MCDI at 26 months. All correlations are Spearman correlations. Correlations flagged with * are significant at the 0.05 level (2-tailed) and those with ** are significant at the 0.01 level (2-tailed).*

			26-Month WS-MCDI						
			Words Produced	Past, Future, Absence Words	Word Endings	Irreg. Words	Overreg. Words	Grammatical Complexity	M3L
14.5-Month WG-MCDI	1 st Signs of Understanding	<i>r</i> <i>N</i>	-.162 59	-.069 59	-.131 59	-.089 59	.117 59	-.135 59	-.135 59
	Phrases Understood	<i>r</i> <i>N</i>	.117 59	.197 59	.045 59	.124 59	.009 59	.047 59	.034 59
	Imitation, Labeling	<i>r</i> <i>N</i>	.316* 59	.347** 59	.286* 59	.257* 59	.079 59	.286* 59	.213 59
	Words Understood	<i>r</i> <i>N</i>	.263* 59	.135 59	.125 59	.247 59	.061 59	.190 59	.082 59
	Words Produced	<i>r</i> <i>N</i>	.436** 59	.207 59	.297* 59	.288* 59	-.036 59	.318* 59	.264* 59
	Early Gestures	<i>r</i> <i>N</i>	.333** 59	.193 59	.122 59	.287* 59	.006 59	.296* 59	.171 59
	Later Gestures	<i>r</i> <i>N</i>	.218 59	.065 59	.141 59	.189 59	-.044 59	.151 59	.065 59
	Total Gestures	<i>r</i> <i>N</i>	.251 59	.085 59	.135 59	.215 59	-.064 59	.190 59	.093 59

Correlations between measures on the WG-MCDI and the MRLS score revealed that the only 14.5-month outcome to significantly correlate with comprehension at 26-months was Total Gestures. This correlation was positive, indicating that a greater number of Total Gestures was predictive of a higher MRLS score. Results for all correlations between the eight language measures collected at 14.5 months and MRLS score at 26 months are summarized in Table 10. The low number of significant correlations between measures collected through the WG-MCDI and the MRLS was somewhat unexpected. At the very least, a significant predictive relationship was expected between the language comprehension measure at 14.5 months and that at 26 months. Perhaps this relationship did not surface because Words Understood at 14.5 months was not an accurate representation of what participants were truly understanding at that age. After all, it is much more difficult to decide whether your child understands a word compared to deciding whether your child produces a word; the latter is based on an overt behavior, whereas the former may require some guesswork.

*Table 10. Summary of correlations between language outcome measures collected with the WG-MCDI at 14.5 months and the MRLS at 26 months. All correlations are Spearman correlations. Correlations flagged with * are significant at the 0.05 level (2-tailed).*

			26-Month MRLS Score
14.5-Month WG-MCDI	1 st Signs of Understanding	<i>r</i> <i>N</i>	.000 52
	Phrases Understood	<i>r</i> <i>N</i>	.170 52
	Imitation, Labeling	<i>r</i> <i>N</i>	.163 52
	Words Understood	<i>r</i> <i>N</i>	.234 52
	Words Produced	<i>r</i> <i>N</i>	.221 52
	Early Gestures	<i>r</i> <i>N</i>	.245 52
	Later Gestures	<i>r</i> <i>N</i>	.265 52
	Total Gestures	<i>r</i> <i>N</i>	.286* 52

Finally, correlational analyses between production and comprehension measures concurrently collected at 26 months revealed a number of statistically significant relationships between the WS-MCDI and the MRLS. Specifically, the MRLS score significantly correlated with Words Produced, Use of Words to Express the Past, Future, and Absence, Use of Word Endings to Express Plurals, Possessives, Progressives, and the Past Tense, Use of Irregular Words, Grammatical Complexity, and M3L. These correlations were all highly significant, with *r* values ranging from 0.300 to 0.530 (see Table 11 for a complete summary of correlations across outcome measures). Besides underscoring a high degree of interrelatedness between the WS-MCDI and the MRLS, these results are indicative of a couple of things. First, they show that language

production skills, as assessed by the WS-MCDI, are highly predictive of concurrent language comprehension, as evaluated by the MRLS. Second, they suggest that parental observations of their children's abilities are reliable. Otherwise, there would not be so many significant correlations between the WS-MCDI, which was filled out by parents, and the MRLS, which was administered in the lab by an unbiased experimenter.

*Table 11. Summary of correlations between language outcome measures collected with the WS-MCDI and the MRLS at 26 months. All correlations are Spearman correlations. Correlations flagged with * are significant at the 0.05 level (2-tailed) and those with ** are significant at the 0.01 level (2-tailed).*

		26-Month WS-MCDI						
		Words Produced	Past, Future, Absence Words	Word Endings	Irreg. Words	Overreg. Words	Grammatical Complexity	M3L
26-Month MRLS Score	<i>r</i> <i>N</i>	.530** 53	.300* 53	.464** 53	.449** 53	.138 59	.520** 53	.518** 53

5.1.6 Discussion

Like previous research, the present data set indicated that preterm children tend to be outscored by their full-term counterparts on measures of language outcome. This was evident both at the early (14.5 months) and later (26 months) stages of language acquisition, with differences being especially prevalent later in development. The present data set specifically indicates that preterm participants have a syntactic disadvantage, producing shorter and less complex utterances than full-term participants. Moreover, as indexed by the fewer number of test items completed on the MRLS, preemies' comprehension skills do not seem to be as advanced as those of their full-term

counterparts. It is critical to note that even though many of the cross-group differences in outcome measures were not statistically significant preterm participants were more than two months older than full-term controls, and as a result were exposed to language significantly longer. On this age account, preterm participants should have outperformed full-term participants, given that, for a typically developing toddler, a two-month span is enough time to make significant strides in vocabulary acquisition. Thus, if compared to chronologically age-matched full-term peers, the present sample of preterm participants would likely show significant delays in language skills.

A high degree of predictive value was found across measures collected with the WG-MCDI at 14.5 months and those collected with the WS-MCDI at 26 months. This is unsurprising, as the WS-MCDI, which was designed to assess language in older children, is essentially just an extension of the WG-MCDI. Moreover, both of these surveys rely on parental observations, thus parental consistency across ages makes sense and should be expected. The 14.5-month data showed that the measures with the highest degree of relatedness to later outcomes were Imitation and Labeling, Words Produced, and Early Gestures. Collectively, these three early measures represented six of the seven measures surveyed by the WS-MCDI at 26 months, including Words Produced, Use of Past, Future, and Absence Words, Use of Word Endings, Use of Irregular Words, Grammatical Complexity, and M3L. Specifically, Imitation and Labeling and Words Produced each correlated with five 26-month measures, and Early Gestures with three measures; all of these three 14.5-month measures correlated with Words Produced, Use of Irregular Words, and Grammatical Complexity. Words Produced at 26 months was the best predicted outcome measure, with a total of four 14.5-month measures strongly correlating

with it (i.e., Imitation and Labeling, Words Understood, Words Produced, and Early Gestures).

Despite the high number of correlations between the WG-MCDI at 14.5 months and the WS-MCDI at 26 months, the former was, as a whole, a poor predictor of comprehension at 26 months. Of the eight WG-MCDI measures collected at 14.5 months, only Total Gestures significantly correlated with the 26-month MRLS score. This predictive relationship was unexpected: superficially, there is no obvious connection between these two language outcome measures. However, further inspection of the types of gestures evaluated at 14.5 months and the test items completed at 26 months suggests a high level of overlap between them, and this is likely the source of the correlational relationship observed between scores on Total Gestures and MRLS. Many of the assessed gestures, which involve games and routines children engage in with their caregivers and imitation of parental behaviors, are indirectly referenced in the MRLS. For instance, the WG-MCDI asks parents if their child eats with or pretends to feed his/her doll with a fork or spoon and the MRLS asks children to identify which object, out of a chair, spoon, pair of scissors, and toy car, is used to eat. Another possible explanation for the correlational relationship between Total Gestures and the MRLS score is that 14.5-month-olds with a greater propensity for playing games and imitating parental behaviors will also be more likely to cooperate with the experimenter and complete more of the game-like items on the MRLS at 26 months.

Lastly, language outcome data showed a significant degree of relatedness between concurrent measures of language production and comprehension at 26 months. Of the seven measures collected on the WS-MCDI, all but Use of Overregularized Words

significantly correlated with the MRLS score. Though unsurprising, these findings are exciting, as they help to validate parental reports, which some argue are less reliable than measures collected in the lab.

PART III

**RELATIONS BETWEEN LANGUAGE, COGNITIVE AND ENVIRONMENTAL
PREDICTORS, AND LATER LANGUAGE OUTCOMES**

Chapter 6: Connecting the Dots Through Structural Equation Modeling

This section conceptualizes and tests predictive relationships between the latent constructs early native-language Sound-Structure Sensitivities, later Word Form Representations, Cognitive Skill, developmental environment, birth status, and Language Outcomes using structural equation modeling (SEM). The initial goal was to explore all possible relationships between the various indicator variables assessing each of these constructs (see Figure 1). However, the study's small sample size limited the number of variables that could be concurrently evaluated within a single unified model. Thus, model testing was conducted in an incremental fashion, starting with a base model comprised only of the three language latent constructs, Sound-Structure Sensitivities, Word Form Representations, and Language Outcomes; variables representing Cognitive Skill, birth status, and developmental environment were gradually aggregated as model covariates. This chapter summarizes and examines the outcomes of model testing using SEM. In the following sections use of SEM as a statistical tool for analysis of this data set is validated despite the study's less than ideal sample size.

6.1 Why Use SEM Instead of Multiple Regression Analysis?

SEM is a combination of exploratory factor analysis and multiple regression analyses. What makes it a unique and powerful tool is that it allows for the exploration of relationships between *multiple* independent (predictor) variables and *multiple* dependent (outcome) variables. While regression analysis does permit concurrent use of multiple predictor variables, it can only account for their relationship with one outcome variable at

a time. Additionally, SEM looks at relationships between multiple observed variables, or indicators, through the latent constructs, or factors, that they represent. Indicators under a single latent construct group together because they are highly correlated and are assumed to have the same (or related) underlying cause(s), which is captured by the latent construct. Since SEM focuses on relationships between latent constructs, rather than between indicator variables, it is not as susceptible to collinearity issues as multiple regression; in the latter, multiple highly correlated predictors can yield collinearity issues that result in a misestimate of the variance accounted for in the model. Moreover, direct analysis of the latent variables instead of the observed indicators is also advantageous as it reduces measurement error. Finally, SEM can be used to test the mediating role of a variable, and thus, is not limited to additive models. In short, SEM answers the same questions that regression can, but it is a more powerful tool as it can concurrently take into account interactions, nonlinearities, correlations between independent variables, correlations between error terms, and multiple latent factors with multiple indicators.

6.2 Data Analysis With SEM

Given the extensive number of predictor and outcome variables included in this longitudinal project and all the potentially meaningful relationships between these variables, SEM presented itself as the ideal tool for assessing the present data set. However, one caveat to consider in the following results is the limited number of participants on which analyses were based; typically, SEM is used with much larger sample sizes, with more complex models requiring more participants (Ullman, 2006). As reviewed by Thompson (1998), some have suggested that the minimum sample size

required by SEM is 100, and that at least 10 subjects are needed per indicator variable measured. In an effort to address the present study's sample size issue, preliminary analyses were undertaken to reduce the number of indicator variables in data set. Factor analysis (FA) was run on these variables to see how well they fit within the latent constructs they were assigned to; variables with poor factor loadings were eliminated from their corresponding latent factors, and thus, excluded from SEM analyses.

Prior to FA, four latent constructs were conceptualized from the data collected for this study. These included Sound-Structure Sensitivities (i.e., early receptive language abilities), Word Form Representations (i.e., later receptive language skills), Cognitive Skill, and Language Outcomes. Sound-structure sensitivities encompassed four indicator variables, including sensitivity to stress at 4 and 6 months, and sensitivity to phonotactic probabilities at 6 and 9 months. Word Form Representations comprised four indicator variables, including sensitivity to mispronunciations at 19 months and three measures of word processing accuracy following target noun onset at 26 months¹⁸. Cognitive skill was represented by visual recognition memory at 12 months, Mullen visual reception scores at 26 months, and Mullen fine motor scores at 26 months. Finally, Language Outcomes was indexed by the eight WG-MCDI scores collected at 14.5 months, the seven WS-MCDI scores collected at 26 months, and Mullen receptive language scores at 26 months.

Collectively, the four latent constructs were represented by 12 indicator variables, which were assessed with a grand total of 28 measurements. Table 12 provides a summary of the latent constructs, their corresponding indicator variables, and the

¹⁸ RT measures were excluded from analyses on latent constructs because RT is a less direct measure of linguistic abilities. RT data is typically thought to represent the time it takes to process and react to any type of stimulus, regardless of the modality it is presented in. Thus, it has a cognitive rather than a linguistic basis.

measurements selected to represent each of these variables. FA and SEM were performed with Mplus software, which uses maximum likelihood estimates to resolve model convergence and Chi-square, root mean square error of approximation (RMSEA), and comparative fit index (CFI) tests to assess goodness-of-fit.

Table 12. Summary of the initial latent constructs (4), their corresponding measurement variables (12), and the values used to represent each of these variables (28). Construct validity was evaluated using FA.

LATENT CONSTRUCT	INDICATOR VARIABLE	MEASUREMENT USED
Sound-Structure Representations (Receptive Lang.)	1a. Sensitivity to stress at 4 months 1b. Sensitivity to stress at 6 months 2a. Sensitivity to phonotactics at 6 months 2b. Sensitivity to phonotactics at 9 months	1a. Difference score (TROCHAIC-IAMBIC) 1b. Difference score (TROCHAIC-IAMBIC) 2a. Difference score (HIGH-LOW) 2b. Difference score (HIGH-LOW)
Word Form Representations (Receptive Lang.)	3. Sensitivity to mispronunciations at 19 months 4. Word processing accuracy at 26 months	3. Difference score (CORRECT-MISPRO) 4. Mean proportion looking to the target following target noun onset for I-verb, I-adj, and the composite of Vanilla, U-Verb, and U-adj trials
Cognitive Skill	1. Visual recognition memory at 12 months 2. Visual reception skills at 26 months 3. Fine motor skills at 26 months	1a. Abstract patterns composite novelty score (FAMILIAR-NOVEL) 1b. Faces composite novelty score (FAMILIAR-NOVEL) 2. MVRs Score 3. MFMS score
Language Outcomes	1. WG-MCDI at 14.5 months 2. WS-MCDI at 26 months 3. Receptive language skills at 26 months	1. Scores for 1 st Signs of Understanding, Phrases Understood, Imitation and Labeling, Words Understood, Words Produced, Early Gestures, Later Gestures, and Total Gestures 2. Scores for Words Produced, use of Past/Future/Absence Words, Word Endings, Irregular Words, Overregularized Words, Grammatical Complexity, and M3L 3. MRLS score

6.2.1 Participants

As previously stated in Chapter 1, a total of 62 (19 preterm and 43 full-term) participants were eligible for inclusion in SEM analyses. In order to be retained, participants had to (1) have at least one indicator measure for each of the latent language constructs described below (i.e., Sound-Structure Sensitivities, Word Form Representations, and Language Outcomes) AND (2) have information on gestational age (GA), performance on the Mullen Visual Reception Subscale (MVRs Score), and maternal education (MaEd). Missing data points were imputed with Mplus, which uses a Bayesian Estimator calculator.

6.2.2 Results

6.2.2.1 FA Results

FA results demonstrated that most indicator variables loaded significantly onto their assigned latent factors. The exception was the four variables selected to represent Cognitive Skill. Model conversion could not be achieved for this latent factor, suggesting that novelty scores for abstract patterns, novelty scores for faces, MVRs scores, and MFMS scores, have different underlying causes, and thus cannot be grouped under one factor. Consequently, Cognitive Skill could not be used as a latent factor in the models subsequently assessed with SEM. However, instead of completely eliminating Cognitive Skill indicators from the analyses, they were each tested as model covariates.

Of the remaining three latent factors, Word Form Representations had the greatest number of well-fitting indicator variables. All representative variables had statistically significant factor loadings; loadings were 0.259 ($p = .038$) for sensitivity to

mispronunciations at 19 months, 0.987 ($p < .001$) for the composite measure of word processing accuracy following target noun onset for Vanilla, U-verb, and U-adj trials at 26 months, 0.444 ($p = .001$) for word processing accuracy following target noun onset for I-verb trials at 26 months, and 0.451 ($p = .001$) for word processing accuracy following target noun onset for I-adj trials at 26 months.

Indicators representing the Sound-Structure Sensitivities factor did not load nearly as well as those for later Word Form Representations. Of the four indicators, only two had potentially meaningful factor loadings; these were sensitivity to lexical stress at 4 months with a loading of 0.388 ($p = .067$) and sensitivity to lexical stress at 6 months with a loading of -0.733 ($p = .020$). Loadings for the remaining indicators, sensitivity to phonotactic probabilities at 6 months and at 9 months, were -0.216 ($p = .277$) and 0.223 ($p = .334$), respectively. Despite having statistically insignificant factor loadings, for phonotactic sensitivity indicators were retained these indicators as part of the model because SEM requires latent factors to have a minimum of three indicator variables. Furthermore, removing the most poorly fitting variable, sensitivity to phonotactics at 9 months, resulted in no model convergence.

The FA model comprised of all 16 language outcome indicators did not converge. Subsequent testing involved stepwise removal of these variables until a good model fit was achieved. The best fitting model was one that only included nine indicator variables. Factor loadings for these variables were 0.361 ($p = .002$) for Imitation and labeling at 14.5 months, 0.337 ($p = .004$) for Comprehension at 14.5 months, 0.498 ($p < .001$) for Production at 14.5 months, 0.911 ($p < .001$) for Production at 26 months, 0.369 ($p = .001$) for Use of past, Future and Absence Words at 26 months, 0.892 ($p < .001$) for Use of

Irregular Words at 26 months, 0.193 ($p = .130$) for Use of Overregularized Words at 26 months, 0.850 ($p < .001$) for Grammatical Complexity at 26 months, and 0.403 ($p < .001$) for MRLS scores at 26 months. To further minimize the number of free parameter estimates in subsequent SEM analyses, Production at 14.5 months and Use of Overregularized Words at 26 months were also removed from the list of indicators representing Language Outcomes. The former was removed because it is a redundant measure: words produced are also understood, thus Words Understood at 14.5 months also encompassed Production at 14.5 months. Use of Overregularized Words at 26 months was excluded because, of the remaining indicators, it had the poorest factor loading on the Language Outcomes construct.

6.2.2.2 SEM Results

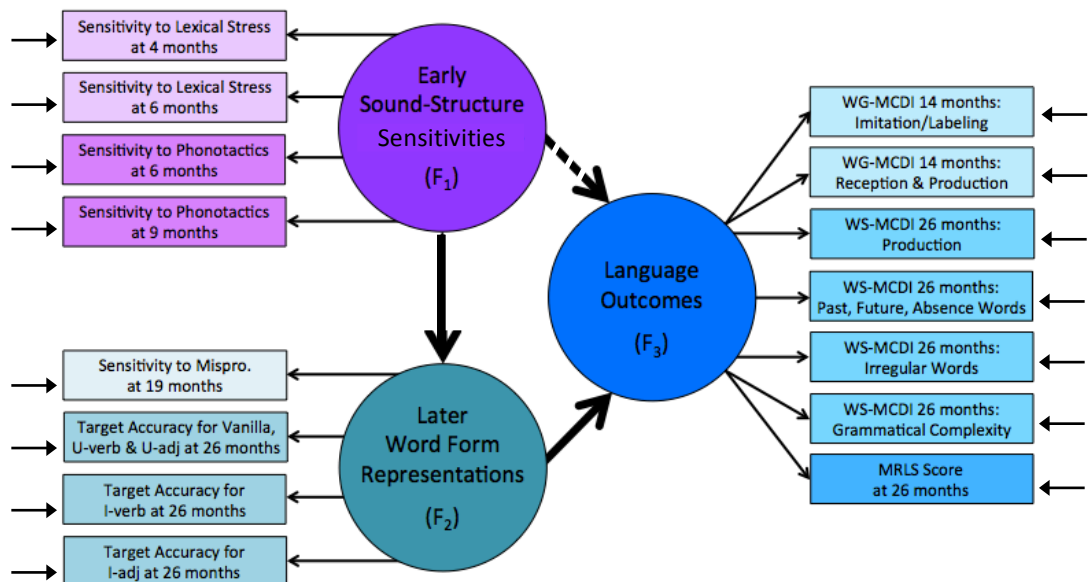
Model 1: Single-Group Model With No Covariates

As a general rule of thumb, model testing with SEM starts with the simplest possible model. Once adequate model fit is attained, covariates can be added to the model to investigate how these impact relationships across latent constructs. The simplest model of interest that can be generated with the present data set is one that assumes a single group of participants, and is only interested in relationships between the latent language constructs, Sound-Structure Sensitivities, Word Form Representations, and Language Outcomes (see Figure 27).

In SEM diagrams latent factors are depicted as circles and are named by a capital F followed by a number subscript; solid single- (i.e., $F_1 \rightarrow F_2 = F_1$ predicts F_2) or double-headed (i.e., $F_1 \leftrightarrow F_2 = F_1$ correlates with F_2) arrows connecting circles mark

relationships between latent factors. For example, the simplest model tested in this section denotes F_1 (i.e., Sound-Structure Sensitivities) as a predictor of F_2 (i.e., Word Form Representations), and F_2 as a predictor of F_3 (i.e., Language Outcomes). Additionally, this model tests an indirect predictive pathway between Sound-Structure Sensitivities and Language Outcomes, as marked by the dashed arrow pointing from F_1 to F_3 . Indicator variables are represented by rectangles, and single-headed arrows pointing from circles to rectangles denote indicators' loadings on their corresponding latent constructs. In this model, F_1 and F_2 each have four indicators, and F_3 has seven indicators.

Figure 27. Structural equation model of the three latent language constructs and their corresponding indicator variables. Circles represent latent constructs and rectangles represent indicator variables. Solid one-headed arrows from F_1 to F_2 and from F_2 to F_3 denote predictive relationships between Sound-Structure Sensitivities and Word Form Representations, and between Word Form Representations and Language Outcomes, respectively. The dashed arrow between F_1 and F_3 symbolizes the indirect predictive relationship between Sound-Structure Sensitivities and Language Outcomes. Single-headed arrows connecting circles to rectangles represent indicators' factor loadings on their corresponding latent factors. Small single-headed arrows pointing to the indicators are residual variances.



Model test results indicated that Sound-Structure Sensitivities significantly predicts Word Form Representations, $R = .457, p = .058$, and that the predictive relationship between Word Form Representations and Language Outcomes is marginally significant, $R = .218, p = .094$ (see Appendix C, Model 1 for Mplus code). Despite these relationships, the indirect predictive pathway between Sound-Structure Sensitivities and Language Outcomes was not significant, $R = .100, p = .233$. Goodness-of-fit indices for this model indicated that it has a mediocre fit, $\chi^2(88, N = 62) = 137.2, p < .001$, RMSEA = .095, and CFI = .801¹⁹, which is unsurprising given the model's high number of parameter estimates and comparable small sample size (i.e., $47:62 = .76$).

As previously mentioned, the current model represents the simplest set of predictive relationships between Sound-Structure Sensitivities, Word Form Representations, and Language Outcomes. It does not take into account how premature birth, cognitive abilities, or developmental environment factors (i.e., maternal education) potentially modulate the relationships between these three latent constructs. Subsequent SEM testing focused on investigating the roles of these variables by adding them to the model as covariates.

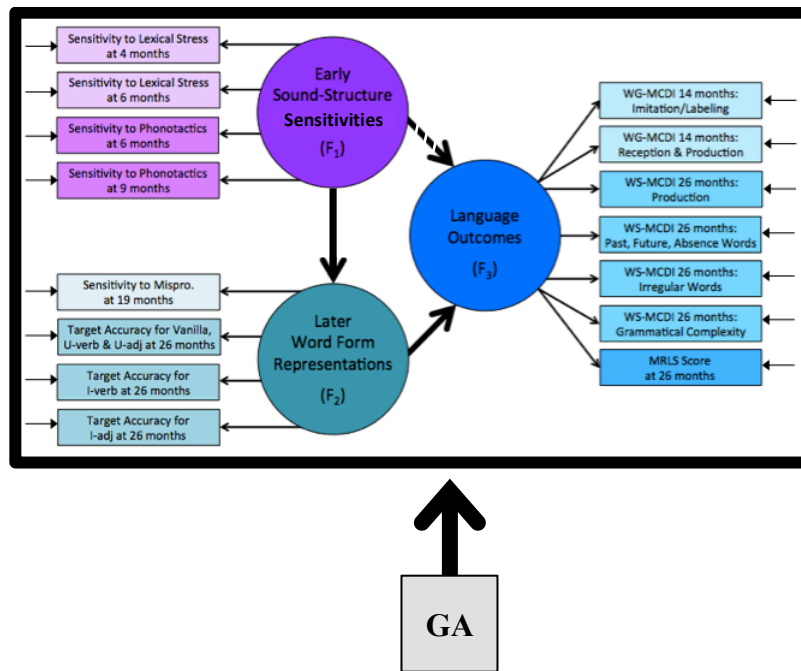
Model 2: Single-Group Model With GA as a Covariate

This model explored the impact of birth status on Sound-Structure Sensitivities, Word Form Representations, and Language Outcomes, and the relationships between these constructs that were established in Model 1 (see Figure 28). Instead of dichotomizing birth status and testing a two-group (i.e., full-term vs. preterm) model, this

¹⁹ Excellent model fit is denoted by RMSEA and CFI values approaching .01 and 1, respectively.

variable was tested as a continuous covariate in the form of GA. As a continuous variable, GA provides greater statistical power, which is especially advantageous in light of the present study's small sample size.

Figure 28. Structural equation model of F_2 on F_1 , F_3 on F_2 , and F_3 indirect on F_1 , with GA as a covariate. Latent factor relationships specified in Model 1 have been maintained in the present model.



Model test results suggest that GA does impact at least some of the three latent language constructs and the relationships established between them (see Appendix C, Model 2 for Mplus code). Specifically, the regression coefficients for Sound-Structure Sensitivities and for Language Outcomes on GA were statistically significant, $R = -.415$, $p = .015$, and $R = .235$, $p = .058$, respectively. These findings show that GA does predict infant sensitivities to native-language sound structures as well as early childhood

language outcomes. Interestingly, GA did not predict Word Form Representations, $R = .227, p = .191$.

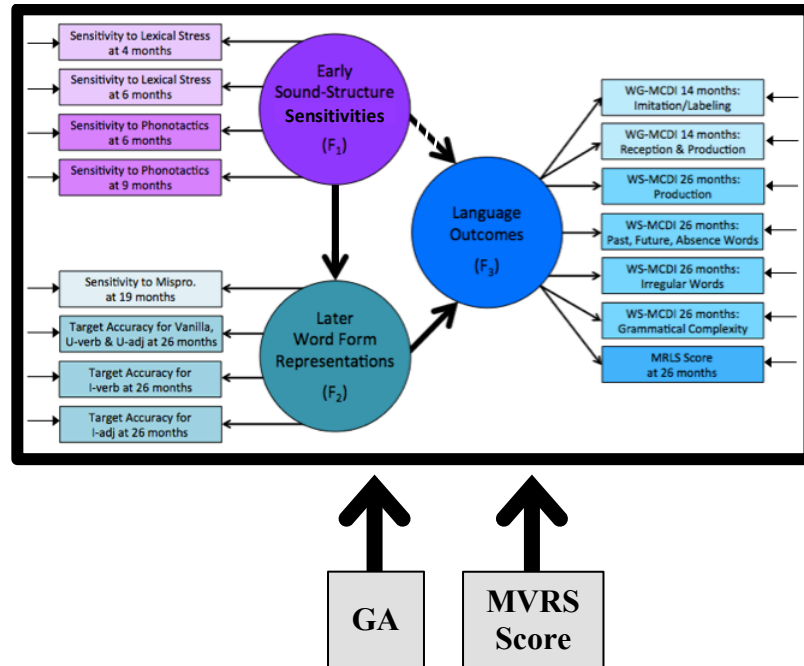
Adding GA as a model covariate also strengthened the predictive relationship between F_1 and F_2 that was established in Model 1, (i.e., Model 1: $R = .457, p = .058$ *versus* Model 2: $R = .512, p = .015$). This suggests that the predictive value of Sound-Structure Sensitivities for Word Form Representations is at least partially modulated by GA. GA did not significantly change Model 1's relationships between F_2 and F_3 , $R = .212, p = .098$, and F_1 and F_3 , $R = .109, p = .194$. Goodness-of-fit indices for the present model were again mediocre, though not as good as those for Model 1, $\chi^2(100, N = 62) = 178.2, p < .005$, RMSEA = .112, and CFI = .726. Again, this is unsurprising given the high ratio of parameter estimates to sample size (i.e., $50:62 = .81$).

The following model assesses the role of cognitive abilities on latent language constructs through the addition of MVRS Score as a covariate.

Model 3: Single-Group Model With GA and MVRS Score as Covariates

Since Cognitive Skill indicators did not load onto a single latent construct, the role of cognition on language constructs had to be explored by individually adding Cognitive Skill indicators as model covariates. This model specifically examined the relationship between visual reception scores on the Mullen at 26 months (MVRS Score) and Sound-Structure Sensitivities, Word Form Representations, and Language Outcomes (see Figure 29).

Figure 29. Structural equation model of F_2 on F_1 , F_3 on F_2 , and F_3 indirect on F_1 , with GA and MVRS Score as covariates. Latent factor relationships specified in Model 1 have been maintained in the present model.



Model test results indicated that while MVRS Score at 26 months significantly associates with the Language Outcomes construct, $R = .291, p = .016$, it is not a significant predictor of either Sound-Structure Sensitivities, $R = .035, p = .840$, or Word Form Representations, $R = .212, p = .177$ (see Appendix C, Model 3 for Mplus code). Adding MVRS Score as a model covariate did not significantly impact the predictive relationships between F_1 and F_2 and between F_1 and F_3 that were initially established in Model 1 and maintained in Model 2: Sound-Structure Sensitivities remained a significant predictor of Word Form Representations, $R = .549, p = .017$, and continued to show no significant associations with Language Outcomes, $R = .088, p = .282$. However, the

addition of MVRS Score did wipe out the relationship between F_2 and F_3 : Word Form Representations no longer predicted Language Outcomes, $R = .161$, $p = .221$.

Covariance relationships seen in Model 2 between GA and the three latent language constructs remained fairly consistent with the addition of MVRS Score: GA continued to show a strong relationship with Sound-Structure Sensitivities, $R = -.436$, $p = .016$, but not with Word Form Representations, $R = .223$, $p = .260$. However, the covariance relationship between GA and Language Outcomes weakened, changing from significant to only marginally significant, Model 2: $R = .235$, $p = .058$ *versus* Model 3: $R = .206$, $p = .087$.

Goodness-of-fit indices for the current model were once again mediocre, $\chi^2(112, N = 62) = 229.3$, $p < .005$, RMSEA = .130, and CFI = .664, which was expected given the model's growing ratio of parameter estimates to sample size (i.e., $53:62 = .85$). The fourth and final model assessed with SEM considered the role of maternal education as an additional covariate.

Model 4: Single-Group Model With GA, MVRS Score, and MaEd as Covariates

Since the present study's small sample size precluded adding a latent construct encompassing measures of developmental environment to SEM model testing, MaEd was chosen as a proxy for developmental environment and was incorporated as a covariate into the model assessed in this section (see Figure 30). Of the various developmental environment indicators collected (e.g., paternal education, number of siblings, number of people living with the participant, hours of television watched per week, etc.), maternal

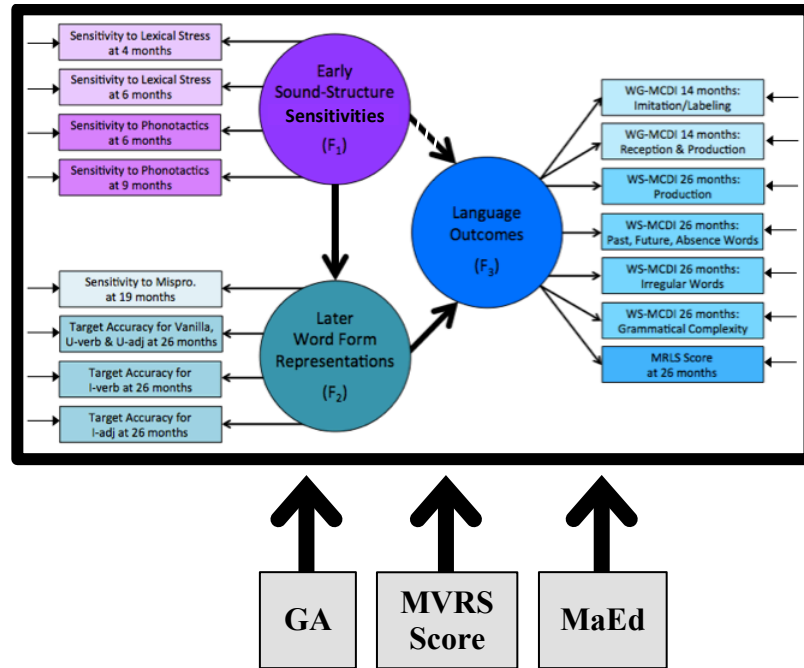
education was selected because it has consistently been demonstrated to predict children's linguistic outcomes. For example, Dollaghan and colleagues (1999) showed statistically significant linear trends across levels of maternal education and various measures of 3-year-olds' spontaneous speech, including mean utterance length in morphemes, number of different words used, and total number of words produced. In all cases, higher levels of maternal education were associated with better scores on these three measures of spontaneous speech.

Maternal education is an important component of socioeconomic status (SES), which has been shown to highly correlate with cognitive and linguistic outcomes (e.g., Hackman & Farah, 2009), structural and functional brain development (e.g., Tomalski et al., 2013), and educational achievement (e.g., Sirin, 2005). The premise behind these relationships is that SES ultimately defines a child's developmental environment by influencing the types of experiences and resources that are available to that child, and these ultimately impact developmental processes and outcomes. Specifically, lower SES has been linked to less cognitive (Cates et al., 2012) and linguistic (Hart & Risley, 1995) stimulation, which enable poorer outcomes. For example, Hart and Risley reported that, on average, children living on welfare hear half the number of words per hour as children from working-class families, and only a third of those heard by children from professional families. Furthermore, these researchers calculated that, by age 4 years, there is an astounding 30 million word gap in language exposure between children of low SES and professional families, with consequences for intellectual development and school performance.

Ultimately, MaEd was selected as the model covariate representative of developmental environment because it is SES's most stable component, and because SES plays an important role in defining developmental environment. Furthermore, compared to other SES components including household income and home resources, maternal education is a socially innocuous and easily collected measure. Finally, besides showing strong associations with language abilities, maternal education likely correlates with many of the other developmental environment measures collected during this study, thus making it an ideal representative of these measures.

Based on the discussion above, MaEd was expected to significantly associate with Language Outcomes, which is the only latent construct that includes measures of participants' production skills. If maternal education influences any of the behaviors underlying the indicators for Sound-Structure Sensitivities and Word Form Representations, then MaEd was also expected to significantly associate with these latent constructs in as much as their corresponding behaviors impact the Language Outcomes construct.

Figure 30. Structural equation model of F_2 on F_1 , F_3 on F_2 , and F_3 indirect on F_1 , with GA, MVRS Score, and MaEd as covariates. Latent factor relationships specified in Model 1 have been maintained in the present model.



Model 4 SEM results revealed that MaEd associated significantly with Sound-Structure Sensitivities, $R = .301$, $p = .053$, but not with Word Form Representations, $R = .046$, $p = .778$, or Language Outcomes, $R = .176$, $p = .196$ (see Appendix C, Model 4 for Mplus code). Latent construct relationships discussed for Model 3 were not significantly altered by adding MaEd as a model covariate: Sound-Structure Sensitivities continued to predict Word Form Representations, $R = .478$, $p = .037$, but not Language Outcomes, $R = .011$, $p = .422$, and, again, no significant predictive relationship was found between Word Form Representations and Language Outcomes, $R = .132$, $p = .317$.

For the most part, covariance relationships between GA, MVRS Score, and the three latent language constructs in this model were consistent with those observed in

Model 3. One exception was the association between GA and Language Outcomes, which went from being marginally significant to statistically insignificant, Model 3: $R = .206, p = .087$ versus Model 4: $R = .140, p = .284$. Relationships between GA and Sound-Structure Sensitivities and between MVRS Score and Language Outcomes remained significant in this model, $R = -.505, p = .003$ and $R = .256, p = .039$, respectively. All other associations remained insignificant, $R = .156, p = .449$ for F_2 on GA, $R = -.019, p = .900$ for F_1 on MVRS Score, and $R = .192, p = .201$ for F_2 on MVRS Score.

As expected, Model 4 goodness-of-fit indices were mediocre, $\chi^2(124, N = 62) = 259.6, p < .005$, RMSEA = .133, and CFI = .612. The number of parameter estimates for this last model was 56, bringing the final ratio of parameter estimates to sample size to .90. This value is .24 greater than it was for Model 1.

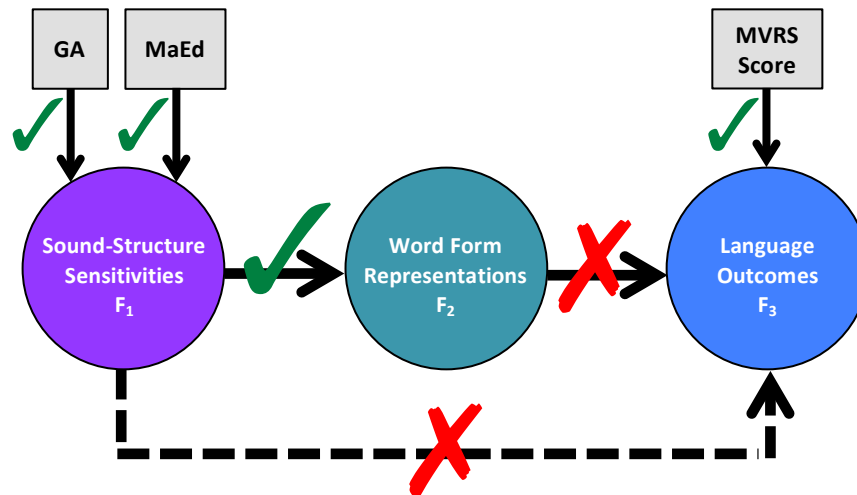
6.2.3 Discussion

The purpose of SEM analyses was to explore how measures of receptive language abilities, birth status, cognitive skill, and developmental environment relate to one another as well as to investigate whether these measures can predict language outcomes in the second and third years of life. At the most basic level, SEM was used to test whether Sound-Structure Sensitivities in infancy predict Word Form Representations in toddlerhood, and whether Word Form Representations in toddlerhood predict Language Outcomes. Of interest was also examining whether the relationships between these latent constructs are modulated by preterm birth, visual reception abilities, and maternal education. These issues are important to examine because preterm children are at a greater risk for language delays and deficits than their full-term counterparts, and

knowing whether there are possible indicators in infancy and early toddlerhood of future language delay is crucial for the development and execution of successful intervention strategies.

Four key findings emerged throughout the model testing process. First, when limited to relationships between the three latent language constructs (i.e., Model 1), SEM results showed that Sound-Structure Sensitivities is a strong predictor of Word Form Representations, and that Word Form Representations marginally predicts Language Outcomes. Second, Sound-Structure Sensitivities remains a significant predictor of Word Form Representations even after GA, MVRs Score, and MaEd are added as model covariates. Together, these first two findings suggest that any predictive value Word Form Representations has on Language Outcomes is likely modulated by some, if not all, of the tested covariates, GA, MVRs Score, and MaEd. The third finding of interest indicated that GA and MVRs Score each show strong covariance relationships with Sound-Structure Sensitivities: lower gestational age and higher maternal education are associated with higher sound-structure sensitivities. Lastly, MVRs Score at 26 months significantly co-varies with Language Outcomes: higher MVRs values were linked with better language outcomes. Figure 31 provides a visual summary of these four key findings.

Figure 31. Summary of key findings established through SEM analyses.



Interestingly, when only GA and MVRS Score were considered as model covariates, GA was marginally associated with Language Outcomes. This relationship was lost when MaEd, which did not significantly co-vary with Language Outcomes, was included in the model. What is the significance of this result? One possibility is that MaEd modulates the association between GA and Language Outcomes. Additional analyses need to be conducted to investigate the appropriateness of this interpretation.

Interestingly, neither of the receptive language constructs, Sound-Structure Sensitivities nor Word Form Representations, were predictive of Language Outcomes. What implications does this have for the original expectations proposed for this study? As previously reviewed in this manuscript, past research has suggested that receptive language skills in BOTH infancy and toddlerhood are predictive of language outcomes. For example, better discriminability of native phonetic contrasts at 7 months is linked with greater vocabulary growth in toddlerhood (Kuhl et al., 2005), and greater speed and

accuracy in word recognition at 18 and 25 months predicts accelerated vocabulary growth, with 18-month measures continuing to predict vocabulary development through age 30 months (Fernald & Marchman, 2012; Fernald, Perfors, & Marchman, 2006). Yet, the present results suggest that sensitivities to lexical stress and phonotactics in infancy, and sensitivities to mispronunciations and word processing accuracy in toddlerhood have no impact on the developmental trajectories and outcomes of language. What is the source of this discrepancy between the present and previous research findings? One possibility is that the receptive language skills assessed in the present study encapsulate aspects of language development that are not reflected in any of the measures collected with the WG-MCDI at 14.5 months, the WS-MCDI at 26 months, or the MRLS at 26 months.

A more plausible explanation for the observed differences across study results lies in the types of outcome measures selected to serve as the predicted variables as well as whether or not these variables were grouped under a single latent construct during analyses. For instance, Kuhl et al. (2005) limited their analyses to correlations between performance on a phonetic discrimination task at 7 months and each of four language outcome measures collected with the WS-MCDI, including number of vocabulary items produced at 18 and 24 months, mean length utterance at 24 months, and grammatical complexity at 24 months. Fernald and colleagues (2006) focused their analyses on hierarchical growth models looking at the relationship between changes in vocabulary size assessed by the MCDI from early to later toddlerhood as a function of speed and accuracy of word recognition at 25 months. Unlike Kuhl et al.'s (2005) and Fernald et al.'s (2006) studies, the present study used seven outcome variables grouped under a single latent construct that was then used as the predicted variable- that is, what was

predicted was a composite variable rather than multiple individual variables. Another difference between this study's and previous studies' outcome variables is that these were collected from different sources, the MCDI and the Mullen Scales of Early Learning, rather than from a single source. Additionally, the selected variables assessed a variety of linguistic skills, ranging from comprehension and production to morphological and syntactic abilities. Perhaps the broad nature of the Language Outcomes construct used in this study made it difficult to find predictive relationships between infant and toddler receptive language measures and language outcomes. Future work needs to be done to develop latent constructs that capture specific types of language outcomes (e.g., syntactic outcomes *versus* morphological outcomes)²⁰.

In addition to the reasons specified above, the restricted variability in scores for many of the language outcome indicators likely contributed to the lack of predictive relationships between the receptive language constructs and the Language Outcomes construct. In correlational (and related) analyses, lack of variability in the data makes it difficult to find meaningful and statistically significant relationships between variables.

Even though the receptive language measures collected for this study did not predict language outcomes in the second and third years of life, the present set of SEM analyses do suggest that there is some degree of continuity between early and later receptive language skills whereby sound-structure sensitivities in infancy predict word form representations in toddlerhood.

²⁰ An effort was made in the present study to separate and group language outcome indicators according to the specific type of language skill they represented. However, model convergence could not be achieved when multiple types of language outcome constructs were used.

PART IV

PROJECT SUMMARY AND SIGNIFICANCE

Chapter 7: General Discussion and Conclusions

7.1 Summary of Measures of Receptive Language Abilities, Cognitive Skill, and Language Outcomes

The key take away message from this study is that preterm children are, even at the earliest stages of development, processing the world around them differently than their full-term counterparts. This theme prevailed throughout many of the experiments reviewed in this manuscript, and was evident in both language and non-language (i.e., cognitive) related contexts.

In the language domain, experiments assessing early sensitivities to native-language sound structures showed that preterm infants are attending to patterns of lexical stress and phonotactics differently than their age matched full-term counterparts. Specifically, we observed that full-term and preterm 6-month-olds have opposite preferences for trochaically and iambically stressed words: while full-terms tend to listen longer to words with trochaic stress, preemies tend to listen longer to words with iambic stress. By 9 months opposite preference patterns also emerged in participants' sensitivities to phonotactic probabilities, with full-term infants showing a significant preference for words with low probabilities and preterm infants tending to prefer words with high probabilities. The fact that preference did not reach statistical significance for the preterm group could be an indication of a few things: preterm infants (1) had a difficult time with the experimental task, (2) do not show a robust preference effect as a group because some infants behave like their full-term counterparts, and/or (3) are

delayed relative to their full-term counterparts (if full-term behavior is accepted as typical).

Data on sensitivities to native-language sound structures indicate that preterm infants do have the ability to process native-language sound patterns. However, they start off with different listening biases than their full-term peers, and these can potentially impact the information being extracted and learned from the linguistic input received. Early differences in how linguistic input is processed can have cascading effects on subsequent language development, first altering the formation of native-language sound-structure representations, and then affecting lexical representations acquired during word-learning. SEM findings from the present project indirectly corroborate this notion, showing that a predictive relationship exists between the constructs representing early Sound-Structure Sensitivities and later Word Form Representations.

The continued effects of prematurity on the development of receptive language skills were also observed in our assessments of word form representations. Data showed that 19-month-old preterm toddlers do not reliably distinguish between correctly pronounced and mispronounced target referents. While full-term subjects looked longer at a target object when it was correctly pronounced, preterm participants spent the same amount of time looking at a target object regardless of whether it was pronounced correctly or not. Though these findings also show behavioral differences between full-term and preterm participants, they are unlike the results on infant sound-structure sensitivities in that they suggest preterm children lack an ability. While data on sound-structure sensitivities indicated that preterm infants possess some ability to distinguish between patterns of lexical stress and phonotactic probabilities, the mispronunciation data

show that, by the time they reach toddlerhood, these same children are incapable of differentiating between correct and incorrect word forms. This suggests that preemies' representation of known words is not as mature as that of their full-term counterparts. If early differences in infant language processing skills do impact later representations of word forms, then it is unsurprising that preterm participants were insensitive to mispronunciations as these were the same children who, as infants, showed atypical patterns of sensitivity to lexical stress and phonotactic probabilities.

Data on word processing skills at 26 months showed that preterm children are just as quick and accurate as their full-term counterparts at target noun recognition when the target word is presented without contextual cues. However, when contextual cues are available in the speech stream, preterm children are not primed by them as readily and do not use them to identify and orient to the target object. Specifically, reaction time data showed that full-term children look at a target more quickly than preterm children when they hear a verb that helps disambiguate the identity of the target before it is enunciated (e.g., “***Read*** the book” versus “***Take*** the book”). Like the mispronunciation data, these data suggest that preterm children are linguistically disadvantaged compared to their full-term counterparts.

Despite the differences in receptive language skills observed between full-term and preterm infants, significant group differences in language outcomes did not emerge until late in the word-learning process, at 26 months. Outcome measures collected at 14.5 months were variable both within and across groups, making it difficult to differentiate between full-term and preterm participants and to singularize measures potentially indicative of future deficits or delays. By 26 months, our data do show meaningful

differences between full-term and preterm participants. Specifically, preterm toddlers showed decreased comprehension abilities, scoring significantly less on the Receptive Language Subscale of the Mullen Scales of Early Learning. Additionally, these children tended to produce shorter and less grammatical complex utterances than their full-term counterparts. Additional follow-up should be conducted to determine whether any of the children participating in this study are eventually diagnosed with language delays and/or deficits.

Evidence supporting differential attention to stimuli between full-term and preterm children was also observed in our visual recognition memory data. Like experiments looking at infant sensitivities to lexical stress and phonotactics, the Rose Task experiment showed that preterm children can distinguish a novel image from a familiarized one, showing the developmentally expected novelty preference. What differentiated preterm from full-term participants was the extent to which they showed novelty preferences for different types of stimuli. While full-term infants preferred novel faces to novel abstract patterns, preterm infants preferred novel abstract patterns to novel faces.

The linguistic and cognitive data reviewed thus far speak to a critical principle about development: small differences in early perceptual sensitivities can have cascading effects on developmental processes that ultimately can result in significantly different outcomes. The idea here is that infant perceptual sensitivities are the building blocks that ultimately define development, guiding attention to input stimuli and impacting what is extracted and learned from this input. When first embarking on this project we were convinced that substantial performance differences would emerge between preterm and

full-term participants at the earliest ages tested, and that such differences would become more pronounced with increasing age. Specifically, we thought preterm participants would be incapable of doing many of the linguistic tasks typically developing full-term infants have been tested on. Given that the poor developmental outcomes associated with preterm birth are numerous and well documented, we assumed that their underlying causes must also be substantial and entrenched early in developmental processes. However, we were proved wrong, as preterm participants were capable of performing the experimental tasks they were tested on in infancy. What differentiated preterm infants from their full-term counterparts was *how* they responded to the stimuli rather than whether they *could* respond to it. On the basis of these findings, we propose that small differences in how preterm infants pay attention to the world around them set the stage for the poor developmental outcomes that are often reported among this special population.

7.2 Summary of SEM Analyses

Though we hypothesized that receptive language abilities in infancy and toddlerhood would predict later language outcomes, SEM analyses suggested that neither infant sound-structure sensitivities nor toddler word form representations are significant language outcome predictors. However, these findings do not mean that the receptive language skills explored in this study are not important for successful language acquisition; instead, they suggest that the measures we selected to represent the latent constructs Sound-Structure Sensitivities and Word Form Representations are not associated with the indicators comprising the Language Outcomes construct. That is,

sensitivities to lexical stress, phonotactic probabilities, and mispronunciations, and word recognition speed and accuracy may be predictive of language outcome measures not included in the Outcomes construct conceptualized for this study.

On this account, a reconceptualization of the Language Outcomes latent construct may yield predictive links between infant and toddler receptive language skills and language outcomes. For instance, instead of grouping outcome indicators under a single Outcomes construct, it may have been more appropriate to group them under two latent constructs, one representing *early* language outcomes and the other representing *later* language outcomes. Specifically, the former would encompass measures from the WG-MCDI collected at 14.5 months, while the latter would encompass those from the WS-MCDI and the MRLS collected at 26 months.

Another reason why this study did not find predictive relationships between receptive language skills and language outcomes may be that many of the indicators included in the Language Outcomes construct had values with small ranges that were not statistically different between full-term and preterm groups. For example, measures such as Imitation and Labeling at 14.5 months and Comprehension and Production at 14.5 months were not only not meaningfully different across groups, but also were grouped together simply because FA results suggested that they loaded well onto a single latent construct. To complicate matters, some of the indicators that were meaningfully different across full-term and preterm groups were not included in the Outcomes latent construct (e.g., M3L at 26 months) tested in this study. Future work should consider repeating SEM analyses with an Outcomes construct that only contains those indicators that were statistically different between full-term and preterm participants. Additionally, this data

set should also be analyzed using multiple regression models that consider the predictive value of those receptive language measures that significantly diverged across full-term and preterm groups on individual language outcome measures that were also statistically different across these groups. For example, one such model could test whether sensitivities to lexical stress at 6 months, phonotactic probabilities at 9 months, and mispronunciations at 19 months collectively and individually predict MRLS score at 26 months.

Despite the lack of significant relationships observed between this study's receptive language and language outcome measures, SEM analyses did show a highly significant link between the latent constructs representing early and later receptive language skills. Specifically, Sound-Structure Sensitivities was highly predictive of Word Form Representations, thereby suggesting some degree of developmental continuity between early, less abstract (i.e., detection of sound-pattern differences) measures of receptive language and later, more complex (i.e., word processing) measures of receptive language. Furthermore, the relationship between Sound-Structure Sensitivities and Word Form Representations could suggest that these constructs embody some unknown language outcome measure that was not captured by our Language Outcomes latent factor.

7.3 Study Limitations (and Merits)

Sample size, particularly within the preterm group, was the biggest challenge faced throughout this project. Its effects were cascading, impacting numerous facets of this study. Issues resulting from a limited sample size are further discussed below.

This project is only one of two longitudinal studies ever conducted in our lab, and it is the first to pursue preterm infant recruitment. Given that previous research endeavors have always focused on studying language development in healthy, full-term infants, the lab has a solid protocol in place that has assured successful subject recruitment. Typically, recruitment is done by calling families who have recently had babies and asking if they are interested in coming to the lab for a study. Though this method is effective for full-term recruitment, it cannot specifically target the recruitment of preterm infants. This left us with limited options for preterm recruitment.

A host of reasons made preterm recruitment for this study extremely difficult. First, as previously mentioned, potential recruits were initially contacted at the Women and Infants Hospital Neonatal Follow-up Clinic in Providence, RI. Given the lack of proximity between the clinic and our testing facilities, only a few hours a week could be devoted to informing parents about the research project. Once at the clinic, it was also not guaranteed that patients would show up for their appointment or that they would be eligible to participate in the study. Furthermore, parental interest in the project at the time of their clinic visit did not assure participation in the study. When first approached at the clinic potential subjects were too young to begin testing procedures, and by the time these infants were old enough to come to the lab, many parents were no longer interested in taking part in the study.

Recruitment was also difficult because many parents were hesitant to have their preterm infants partake in a project that required numerous non-medical visits to a lab and had no obvious benefits for the well-being of their children. Typically, in their first year of life preterm infants have numerous medical follow-ups and early intervention

appointments to check on their general health and to ensure that they are meeting developmental milestones. We suspect that many of the parents contacted at the Clinic already felt overwhelmed by all these appointments, and thus were reluctant to add additional commitments to their schedules.

Lastly, recruitment was challenging because some parents are exceptionally protective of their preterm infants because they perceive them to be especially vulnerable. This is, of course, reasonable and expected given that many preterm infants are born with major health complications that jeopardize their survival and compromise their immune systems. Parents may have seen visits to our lab as an unnecessary risk to their infants' health.

One of the major issues arising from a small sample size is that there was no good way to deal with all the heterogeneity within the preterm group. Participants in this group covered a large range of gestational ages, birthweights, and overall health at birth, and because there were few participants, we were unable to bin and compare the data according to how preterm participants were classified at birth (i.e., MP, VP, or EP, and LB, VLBW, ELBW, or micropremature). Initially, we tried to avoid too much within group heterogeneity by enforcing strict recruitment criteria. For instance, participants were expected to have gestational ages under 33 weeks and weigh no more than 1500 g at birth. These children were also expected to be relatively healthy as newborns; neonates who were on ventilators and/or had IVH or PVL above grade I were initially not eligible to participate. However, we quickly realized that few participants met these recruitment criteria and even fewer were willing to participate in this study. Thus, in order to get the project off the ground, recruitment criteria had to be relaxed. This change augmented

sample variability, which likely contributed to the increased variability observed in experimental results.

Too much within group sample variability is problematic for this study for at least two reasons. First, it makes it more difficult to get statistically significant results, especially with infant data, which can be extremely variable even among typically developing samples. Too much sample variability yields data sets that can look flat and potentially hide meaningful effects. For instance, if, in a given experiment, half of the participants are showing one effect and the other half are showing the opposite effect, then the resulting mean is zero, essentially, a cancellation of both effects. Thus, sample variability might explain why we had such a hard time finding statistically significant performance differences between full-term and preterm participants in some of the studies reviewed throughout this paper. Second, a varied sample combined with a small sample size did not allow us to examine the graded effects of prematurity and birthweight on the developmental outcomes assessed. Despite most research suggesting that degree of prematurity in conjunction with birthweight significantly predict later developmental outcomes, all preterm participants in this study had to be grouped together, regardless of how early they were born and how much they weighed. Thus, data analyses for this project make the assumption that all preterm infants follow the same developmental trajectories and have similar outcomes.

This study's small sample size also limited the power and scope of the statistical techniques that could be applied to data processing. This was especially true for SEM, which is typically only considered appropriate for use with large sample sizes. Since the present sample was small relative to the tested models' complexity and high number of

parameter estimates, many of the variables we collected data on could not be included in SEM analyses. This was especially true of variables measuring cognitive abilities and developmental environment. As a result, the relationships between receptive language abilities, cognitive skills, developmental environment, and language outcomes that we initially sought to investigate could not be fully explore.

Despite its limitations, this project does have many merits. One of its principal strengths is that it followed developmental outcomes in the same two samples of children for a period of 22 months. During this time a lot of detailed information was collected on each participant, and the result was an exceptionally rich data set. Of course, with any longitudinal study, there are going to be missing data points and some degree of subject attrition. This study was no exception to these shortcomings. Not all participants were able to complete the full study, meaning that little to no information is available on these participants' later receptive language skills, cognitive abilities, and language outcomes. Furthermore, many of those who dropped out were some of the most preterm participants. These infants, who have the highest risk for poor developmental outcomes, also happened to drop out early on in the project; they only contributed to the first couple of experiments, which are the ones that showed pretty significant differences between full-term and preterm participants. Loss of these preterm infants could further explain why many of the group differences found in toddler receptive language experiments and language outcome measures were not large or statistically significant.

Another important contribution made by this project is that it focuses on more than just language outcomes. Previous research showing that preterm children are more likely to develop linguistic deficits and delays has always been based on outcome

measures in toddlerhood; little is known about what is causing these linguistic issues or what is going on in infancy that might predict these later problems. This project is unique in that it looks at language outcomes in a holistic manner, taking into account infant and toddler receptive language abilities, cognitive skills, and environmental factors that potentially guide and define language outcomes.

The multiple facets of this study along with the richness of the resulting data set have given me the opportunity to think about developmental processes and data analyses from a number of different perspectives. I have gained a greater appreciation for the complexities of developmental trajectories and for how different types of variables can affect their outcomes. Furthermore, having a longitudinal data set allows the researcher to better develop a “big picture” perspective and to draw unexpected connections between seemingly unrelated variables. Longitudinal data are powerful as they can really challenge and foster interpretations of developmental processes.

7.4 Future Directions

The present work is an exploration of relationships between infant and toddler receptive language skills and language outcomes, and how these are modulated by birth status (i.e., preterm vs. full-term). These relationships were examined using latent constructs, each comprised of a number of indicators directly assessing receptive language skills and language outcomes. Since analyses were conducted over the latent constructs, the role played by these indicators in defining language outcomes was only assessed indirectly. Future work should focus on examining the predictive value of specific indicators on specific outcome measures. Knowing what receptive language

skills best predict future outcomes could inform diagnostic and early intervention strategies.

In its current state, this body of work only captures language acquisition in progress, and does not say much about its end state. A potential follow-up to this project would be to continue testing full-term and preterm participants on language skills after age 2 years. Of particular interest would be skills that have been linked to future academic success, such as articulatory abilities, phonological awareness, and other pre-reading skills. Assessing relationships between the receptive language measures collected in this study and language outcomes after age 2 would help gain a better understanding of the continued role played by infant and toddler skills in defining developmental trajectories and outcomes.

Another issue to be addressed by future work is that of sample size, particularly among preterm participants. As previously mentioned, the preterm sample tested throughout this project was relatively small and heterogeneous. One solution would be to continue testing preterm infants on the many experiments discussed in this dissertation. Ideally, all levels of prematurity would be equally represented in the new test sample. This would allow for a better understanding of the graded effects gestational age is assumed to have on developmental processes and outcomes.

Finally, though many of the results discussed in this paper are consistent with the More is Less Hypothesis (MiLH), none of the experiments directly test it. MiLH posits that excessive and premature multisensory stimulation interferes with the trajectory of developmental processes and outcomes. Future experiments could test this hypothesis more directly by examining preterm newborns' sensitivity to native-language rhythmic

patterns as a function of their level of prematurity. If *in utero* exposure to low-pass filtered speech is guiding the earliest stages of language acquisition, then preterm newborns who spend more time *in utero* after their cochleae become functional are more likely to show similar patterns of native-language recognition as their full-term counterparts. Specifically, compared to newborns with shorter gestational ages, those with greater gestational ages are expected to be more likely to distinguish their native language from rhythmically different languages.

References

- Aylward, G.P. (2002). Cognitive and Neuropsychological Outcomes: More Than IQ Scores. *Mental Retardation and Developmental Disabilities Research Reviews*, **8**, 234-240.
- Baron, I.S., & Rey-Casserly, C. (2010). Extremely Preterm Birth Outcome: A Review of Four Decades of Cognitive Research. *Neuropsychological Review*, **20**, 430-452.
- Boenker, G. Premature Birth Statistics. (2008). Quint Boenker Preemie Survival Foundation. Retrieved April 2013, from <http://www.preemiesurvival.org/info/>.
- Bosch, L. (2011). Precursors to language in preterm infants: Speech perception abilities in the first year of life. In O. Braddick, J. Atkinson & G. Innocenti (Eds.), *Progress in Brain Research* (Vol. 189, pp. 240-257). Elsevier B.V.
- Cates, C.B., Dreyer, B.P., Berkule, S.B., White, L.J., Arevalo, J.A., & Mendelsohn, A.L. (2012). Infant Communication and Subsequent Language Development in Children from Low Income Families: The Role of Early Cognitive Stimulation. *Journal of Developmental and Behavioral Pediatrics*, **33**(7), 577-585.
- DeCasper, A.J., & Fifer, W.P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science*, **208**(6), 1174-1176.
- Dollaghan, C.A., Campbell, T.F., Paradise, J.L., Feldman, H.M., Janosky, J.E., Pitcairn, D.N., & Kurs-Lasky, M. (1999) Maternal Education and Measures of Early Speech and Language. *Journal of Speech, Language, and Hearing Research*, **42**, 1432-1443.
- Elman, J.L. (1993). Learning and development in neural networks: the importance of starting small. *Cognition*, **48**, 71-99.
- Fernald, A., & Marchman, V.A. (2012). Individual Differences in Lexical Processing at 18 Months Predict Vocabulary Growth in Typically Developing and Late-Talking Toddlers. *Child Development*, **83**(1), 203-222.
- Fernald, A., Perfors, A., & Marchman, V.A. (2006). Picking Up Speed in Understanding: Speech Processing Efficiency and Vocabulary Growth Across the 2nd Year. *Developmental Psychology*, **42**(1), 98-116.
- Foster-Cohen, S., Edgin, J.O., Champion, P.R., & Woodward, L.J. (2007). Early delayed language development in very preterm infants: Evidence from the MacArthur-Bates CDI. *Journal of Child Language*, **34**, 655-675.

- Fouchée, R.D., & Lickliter, R. (2002). Early visual experience affects postnatal auditory responsiveness in bobwhite quail (*Colinus virginianus*). *Journal of Comparative Psychology*, **116**(4), 369-380.
- Grissmer, D., Grimm, K.J., Aiyer, S.M., Murrah, W.M., & Joel, S. (2010). Fine motor skills and early comprehension of the world: Two new school readiness indicators. *Developmental Psychology*, **46**(5), 1008-1017.
- Hackman, D.A., & Farah, M.J. (2009). Socioeconomic status and the developing brain. *Trends in Cognitive Sciences*, **13**(2), 65-73.
- Hart, B., & Risley, T. (1995). *Meaningful Differences in the Everyday Experience of Young American Children*. Baltimore, MD: Paul H. Brookes Publishing Co., Inc.
- Herold, B., Höhle, B., Walch, E., Weber, T., & Obladen, M. (2008). Impaired word stress pattern discrimination in very-low-birthweight infants during the first 6 months of life. *Developmental Medicine & Child Neurology*, **50**, 678-683.
- Houston-Price, C., & Nakai, S. (2004). Distinguishing Novelty and Familiarity Effects in Infant Preference Procedures. *Infant and Child Development*, **13**, 341-348.
- Hunter, M. & Ames, E. (1988). A multifactor model of infant preferences for novel and familiar stimuli. In *Advances in Infancy Research*, Rovee-Collier, C. & Lipsitt, L. (eds.), vol. 5. Ablex: Stamford; 69-95.
- Johnson, M.H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, **20**(1 and 2), 1-19.
- Jusczyk, P.W., Houston, D.M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, **39**, 159-207.
- Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infants' Sensitivity to Phonotactic Patterns in the Native Language. *Journal of Memory and Language*, **33**(5), 630-645.
- Jusczyk, P.W., Friederici, A.D., Wessels, J., Svenkerud, V.Y., & Jusczyk, A.M. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory & Language*, **32**, 402-420.
- Kapellou, O., Counsell, S., Kennea, N., Dyet, L., Saeed, N., Stark, J., et al. (2006). Abnormal cortical development after premature birth shown by altered allometric scaling of brain growth. *PLoS Med*, **3**(8), 1382-1390.

- Kenny, P. A., & Turkewitz, G. (1986). Effects of unusually early visual-stimulation on the development of homing behavior in the rat pup. *Developmental Psychobiology*, *19*(1), 57-66.
- Kern, S., & Gayraud, F. (2007). Influence of preterm birth on early lexical and grammatical acquisition. *First Language*, *27*(2), 159-173.
- Kuhl, P.K., Conboy, B.T., Padden, D., Nelson, T., & Pruitt, J. (2005). Early Speech Perception and Later Language Development: Implications for the “Critical Period”. *Language Learning and Development*, *1*(3&4), 237-264.
- Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by six months of age. *Science*, *255*, 606-608.
- Kuhl, P.K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, *9*(2), F13-F21.
- Le Normand, M.T., & Cohen, H. (1999). The delayed emergence of lexical morphology in preterm children: the case of verbs. *Journal of Neurolinguistics*, *12*, 235-246.
- Marston, L., Peacock, J.L., Calvert, S.A., Greenough, A., & Marlow, N. (2007). Factors affecting vocabulary acquisition at age 2 in children born between 23 and 28 weeks’ gestation. *Developmental Medicine & Child Neurology*, *49*, 591-596.
- Mehler, J., Jusczyck, P., Halsted, N., Bertoni, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*(2), 143-178.
- Moon, C., Panneton-Cooper, R., & Fifer, W.P. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development*, *16*, 495-500.
- Newport, E.L. (1990). Maturation Constraints on Language Learning. *Cognitive Science*, *14*(1), 11-28.
- Ortiz-Mantilla, S., Choudhury, N., Leevors, H., & Benasich, A.A. (2008). Understanding language and cognitive deficits in very low birth weight children. *Developmental Psychobiology*, *50*(2), 107-126.
- Peña, M., Werker, J.F., & Dehaene-Lambertz, G. (2012). Earlier Speech Exposure Does Not Accelerate Speech Acquisition. *Journal of Neuroscience*, *32*(33), 1159-1163.
- Peterson, B., Vohr, B., Kane, M., Whalen, D., Schneider, K., et al. (2002). A Functional Magnetic Resonance Imaging Study of Language Processing and Its Cognitive Correlates in Prematurely Born Children. *Pediatrics*, *110*(6), 1153-1162.

- Peterson, B., Vohr, B., Staib, L., Cannistraci, C., Dolberg, A., Schneider, K., et al. (2000). Regional brain volume abnormalities and long-term cognitive outcome in preterm Infants. *American Medical Association*, **284**, 1939-1947.
- Pritchard, V., Clark, C., Liberty, K., Champion, P., Wilson, K., & Woodward, L. (2009). Early school-based learning difficulties in children born very preterm. *Early Human Development*, **85**(4), 215-224.
- Ren, J., & Morgan, J.L. (2012). The devil is in the details: Underspecification in infants' early lexical representations. Proceedings of the 36th Annual Boston University Conference on Language Development (pp. 500-511). Somerville, MA: Cascadilla Press.
- Rose, S.A., Feldman, F.F., & Jankowski, J.J. (2001). Attention and Recognition Memory in the 1st Year of Life: A Longitudinal Study of Preterm and Full-term Infants. *Development Psychology*, **37**(1), 135-151.
- Rose, S.A., Feldman, F.F., & Jankowski, J.J. (2009). A Cognitive Approach to the Development of Early Language. *Child Development*, **80**(1), 134-150.
- Rushe, T., Rifkin, L., Stewart, A., Townsend, J., Roth, S., Wyatt, J., et al. (2001). Neuropsychological outcome at adolescence of very preterm birth and its relation to brain structure. *Developmental Medicine and Child Neurology*, **43**(4), 226-233.
- Sansavini, A., Guarini, A., Alessandrini, R., Faldella, G., Giovanelli, G., & Salvioli, G. (2007). Are early grammatical and phonological working memory abilities affected by preterm birth? *Journal of Communication Disorders*, **40**, 239-256.
- Santos, A., Duret, M., Mancini, J., Busuttill, M., & Deruelle, C. (2009). Does preterm birth affect global and configural processing differently? *Developmental Medicine & Child Neurology*, **52**, 293-298.
- Sirin, S.R. (2005). Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research. *Review of Educational Research*, **75**(3), 417-453.
- Stahl, S.A., & Murray, B.A. (1994). Defining phonological awareness and its relationship to early reading. *Journal of Educational Psychology*, **86**(2), 221-234.
- Stolt, S., Klippi, A., Launonen, K., Munck, P., Lehtonen, L., Lapinleimu, H., Haataja, L., & The Pipari Study Group. (2007). Size and composition of the lexicon in prematurely born very-low-birth- weight and full-term Finnish children at two years of age. *Journal of Child Language*, **34**(2), 283- 310.

- Thompson, B. (1998). The Ten Commandments of Good Structural Equation Modeling Behavior: A User-friendly, Introductory Primer on SEM. Invited paper presented at the Annual Meeting of the U.S. Department of Education, Office of Special Education Programs (OSEP) Project Directors' Conference, Washington, DC.
- Tomalski, P., Moore, D.G., Ribeiro, H., Axelsson, E.L., Murphy, E., Karmiloff-Smith, A., Johnson, M.H., & Kushnerenko, E. (2013). Socioeconomic status and functional brain development – associations in early infancy. *Developmental Science*, **16**(5), 676-687.
- Turkewitz, G., & Kenny, P.A. (1982). Limitations on input as a basis for neural organization and perceptual development: A preliminary theoretical statement. *Developmental Psychobiology*, **15**(4), 357-368.
- Ullman, J.B. (2006). Structural Equation Modeling. In B.G. Tabachnick & L.S. Fidell, L.S. (Eds.), *Using Multivariate Statistics 5th Edition* (pp. 676-780). Boston, MA: Pearson.
- Vitevitch, M.S. & Luce, P.A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, and Computers*, **36**, 481-487
- Vohr, B.R. Garcia-Coll, C., & Oh, W. (1988). Language development of low-birthweight infants at two years. *Developmental Medicine & Child Neurology*, **30**(5), 608-615.
- Vohr, B.R., Garcia-Coll, C., & Oh, W. (1989). Language and neurodevelopmental outcomes of low- birthweight infants at three years. *Developmental Medicine & Child Neurology*, **31**(5), 582-590.
- Vohr, B.R., Wright, L.L., Dusick, A.M., Mele, L., Verter, J., Steichen, J.J., Simon, N.P., Wilson, D.C., Broyels, S., Bauer, C.R., Delaney-Black, V., Yolton, K.A., Fleisher, B.E., Papile, L., & Kaplan, M.D. (2000). Neurodevelopmental and Functional Outcomes of Extremely Low Birth Weight Infants in the National Institute of Child Health and Human Development Neonatal Research Network, 1993-1994. *Pediatrics*, **105**(6), 1216-1226.
- Vouloumanos, A., & Werker, J.F. (2004). Tuned to the signal: the privileged status of speech for young infants. *Developmental Science*, **7**(3), 270-276.
- Werker, J.F., & Tees, R.C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development*, **25**, 121-133.
- White, K.S., & Morgan, J.L. (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language*, **59**, 114-132.

- Wolke, D., & Meyer, R. (1999). Cognitive status, language attainment, and prereading skills of 6-year-old very preterm children and their peers: the Bavarian Longitudinal Study. *Developmental Medicine & Child Neurology*, **41**, 94-109.
- Wolke, D., Samara, M., Bracewell, M., & Marlow, N. for the EPICure Study Group. (2008). Specific Language Difficulties and School Achievement in Children Born at 25 Weeks of Gestation or Less. *The Journal of Pediatrics*, **152**(2), 256-262.
- Yoshida, K.A., Fennell, C.T., Swingley, D., & Werker, J.F. (2009). Fourteen-month-old infants learn similar-sounding words. *Developmental Science*, **12**(3), 412-418.
- Young, G.S., Merin, N., Rogers, S. J., & Ozonoff, S. (2009). Gaze behavior and affect at 6 months: predicting clinical outcomes and language development in typically developing infants and infants at risk for autism. *Developmental Science*, **12**(5), 798-814.

Appendix A

Guidelines for completing the MacArthur-Bates Communicative Development Inventories: Words and Gestures (WG-CDI)

The WG-MCDI helps us get our first look into your child's early language comprehension and production. Please read the following instructions before filling it out. If you have any questions or concerns feel free to contact Glenda Molina via phone at 401-863-2377 or e-mail at glenda_molina@brown.edu. If you choose to fill out the WG-MCDI at home, please try to complete and return it within a week of having received it; you can mail it in the post-marked envelope provided to you during your visit.

- 1) WG-MCDIs are used with children from a wide range of ages; your child is **not expected to understand or produce all the items** listed in the WG-MCDI.
- 2) To “Understand” a word means that your baby has knowledge of the word as an object or an action.
- 3) Please **do not mark “imitations”**.
 - Only mark words you know your child uses spontaneously. Do not ask your child to repeat words on the form to figure out whether or not s/he can say or understand them.
- 4) We want to give your child credit for what s/he understands and says. Children often have their own special words for things, people or actions. These words are acceptable substitutes for items listed on the Vocabulary Checklist. Consider the following when filling out the WG-MCDI:
 - a. Give credit for **mispronounced or “baby talk” words** (for example, saying “nanna” instead of “banana” or “blanky” instead of “blanket”).
 - b. Give credit for **signed words**. If your baby produces the sign, please include it under the “Says and Understands” column; if they only understand it when someone else produces it, include it under the “Understands” column. If the word is only signed please indicate that it is signed rather than spoken.
 - c. Give credit for **words s/he says or understands that have the same meaning** as a word on the WG-MCDI (for example, “doggie” instead of “dog”).
 - d. Take into account **words indicating the names of people** (for example, “Tati” or “Oma” instead of “Grandma”) or **words that are specific to your dialect or region** (for example, “owie” instead of “booboo”, “sack” instead of “bag”, “pop” instead of “soda”)

Please write down the word used by your child next to the corresponding word on the WG-MCDI.

- 5) When appropriate, **ask others**.
- In some cases, parents may only have access to a portion of the situations in which children's language abilities may be demonstrated- for example, when children are attending a childcare program. Thus, it may be appropriate to have multiple individuals complete the form. Please list the names of all the individuals who completed the WG-MCDI on the front of the form.
- 6) Please try to use a **No. 2 pencil** to fill in the bubbles. If you don't have one, please use a **black or blue pen**.

Please try to complete and return the WG-MCDI within 2 weeks of your visit.
Thank you for your time and continued interest in our work!

Appendix B

Guidelines for completing the MacArthur-Bates Communicative Development Inventories: Words and Sentences (WS-MCDI)

The WS-MCDI helps us get a sense of your child's language production and grammatical development after his/her second birthday. Please read the following instructions before filling out the form. If you have any questions or concerns feel free to contact Glenda Molina via phone at 401-863-2377 or e-mail at glenda_molina@brown.edu. Please try to return the completed WS-MCDI on the day of your appointment. If this is not possible, please mail the form back in the post-marked envelope we will provide you with on the day of your visit.

- 1) Since the WS-MCDI is different from other language surveys you may have filled out for us, please be sure to **carefully read instructions under each of its two major sections**, Part I: Vocabulary Checklist and Part II: Sentences and Grammar.
- 2) The WS-MCDI only focuses on **words produced**. **Words that are only understood should not be checked off on this form**.
- 3) The WS-MCDI is used with children from a wide range of ages; your child is **not expected to produce all the items** listed on this form.
- 4) Please **do not mark "imitations"**.
 - Only mark words you know your child uses spontaneously. Do not ask your child to repeat words on the form to figure out whether or not s/he can say them.
- 5) We want to give your child as much credit as possible for what s/he can say. Children often have their own special words for things, people or actions. These words are acceptable substitutes for items listed in the WS-MCDI. Consider the following when filling out the form:
 - a. Give credit for **mispronounced or "baby talk" words** (for example, saying "nanna" instead of "banana" or "blanky" instead of "blanket").
 - b. Give credit for **words signed**.
 - c. Give credit for **words s/he says that have the same meaning** as a word on the WS-MCDI (for example, "doggie" instead of "dog").
 - d. Take into account **words indicating the names of people** (for example, "Tati" or "Oma" instead of "Grandma") or **words that are specific to your dialect or region** (for example, "owie" instead of "booboo", "sack" instead of "bag", "pop" instead of "soda")

Please write down alternate words used by your child next to the corresponding word on the WS-MCDI.

- 6) In Part II, D:
- a. List examples just as your child would produce them; **do not correct for grammar**, etc. For example, if your child says, “look at back cat” instead of “look at the black cat”, please write “look at back cat”.
 - b. Sentences that are clearly part of routines (for example, songs, nursery rhymes, poems, etc.) **should not be included** as examples of the longest sentences produced by your child.
- 7) When appropriate, **ask others**.
- In some cases, parents may only have access to a portion of the situations in which children’s language abilities may be demonstrated- for example, when children are attending a childcare program. Thus, it may be appropriate to have multiple individuals complete the form. Please list the names of all the individuals who completed the WS-MCDI on the front of the form.
- 8) Please make sure to **complete pages 1-10** of the WS-MCDI. **Do not forget to put your child’s name and date of form completion** on the first page of the form.
- 9) Please try to use a **No. 2 pencil** or a **black or blue pen** to fill in the bubbles.

**Please try to return the completed WS-MCDI on the day of your visit.
Thank you for your time and continued interest in our work!**

Appendix C

Mplus Codes for SEM Analyses

MODEL 1: F₂ on F₁, F₃ on F₂, and F₃ on_{ind} F₁

TITLE: SEM with two predictor latent variables (Sound-Structure Sensitivities and Later Word Form Representations) and no covariates

DATA: FILE = all_data_imputed.dat;

VARIABLE: NAMES = ID BS SEX GA BW BO MAED PAED TRIM4 TRIM6
HILO6 HILO9 MISPRO19 VANPLN26 IVERBN26 IADJN26
PATTC12 FACEC12 MVRS26 MFMS26 CDIA14 CDIB14 CDIC14
CDID14 CDIE14 CDIF14 CDIG14 CDIH14 CDIA26 CDIB26
CDIC26 CDID26 CDIE26 CDIF26 CDIG26 MRLS26;

USEVARIABLES = TRIM4 TRIM6 HILO6 HILO9 MISPRO19 VANPLN26
IVERBN26 IADJN26 CDIC14 CDID14 CDIA26 CDIB26 CDID26
CDIG26 MRLS26;

OUTPUT: SAMPSTAT;
STANDARDIZED;
TECH1;

MODEL: f1 BY TRIM4 TRIM6 HILO6 HILO9;
f2 BY MISPRO19 VANPLN26 IVERBN26 IADJN26;
f3 BY CDIC14 CDID14 CDIA26 CDIB26 CDID26 CDIG26 MRLS26;
f2 ON f1;
f3 ON f2;

MODEL INDIRECT: f3 IND f2 f1;

!f1=Sound-Structure Sensitivities, f2=Word Form Representations, f3=Language
!Outcomes

MODEL 2: F₂ on F₁, F₃ on F₂, F₃ on_{ind} F₁, and GA_{cov}

TITLE: SEM with two predictor latent variables (Sound-Structure Sensitivities and Later Word Form Representations) and GA as covariate

DATA: FILE = all_data_imputed.dat;

VARIABLE: NAMES = ID BS SEX GA BW BO MAED PAED TRIM4 TRIM6
HILO6 HILO9 MISPRO19 VANPLN26 IVERBN26 IADJN26
PATTC12 FACEC12 MVRS26 MFMS26 CDIA14 CDIB14 CDIC14
CDID14 CDIE14 CDIF14 CDIG14 CDIH14 CDIA26 CDIB26
CDIC26 CDID26 CDIE26 CDIF26 CDIG26 MRLS26;

USEVARIABLES = GA TRIM4 TRIM6 HILO6 HILO9 MISPRO19 VANPLN26
IVERBN26 IADJN26 CDIC14 CDID14 CDIA26 CDIB26 CDID26
CDIG26 MRLS26;

OUTPUT: SAMPSTAT;
STANDARDIZED;
TECH1;

MODEL: f1 BY TRIM4 TRIM6 HILO6 HILO9;
f2 BY MISPRO19 VANPLN26 IVERBN26 IADJN26;
f3 BY CDIC14 CDID14 CDIA26 CDIB26 CDID26 CDIG26 MRLS26;
f2 ON f1;
f3 ON f2;
f1 f2 f3 ON GA;

MODEL INDIRECT: f3 IND f2 f1;

!f1=Sound-Structure Sensitivities, f2=Word Form Representations, f3=Language
!Outcomes

MODEL 3: F₂ on F₁, F₃ on F₂, F₃ on_{ind} F₁, GA_{cov}, and MVRS Score_{cov}

TITLE: SEM with two predictor latent variables (Sound-Structure Sensitivities and Later Word Form Representations) and GA and MVRS Score as covariates

DATA: FILE = all_data_imputed.dat;

VARIABLE: NAMES = ID BS SEX GA BW BO MAED PAED TRIM4 TRIM6
HILO6 HILO9 MISPRO19 VANPLN26 IVERBN26 IADJN26
PATTC12 FACEC12 MVRS26 MFMS26 CDIA14 CDIB14 CDIC14
CDID14 CDIE14 CDIF14 CDIG14 CDIH14 CDIA26 CDIB26
CDIC26 CDID26 CDIE26 CDIF26 CDIG26 MRLS26;

USEVARIABLES = GA TRIM4 TRIM6 HILO6 HILO9 MISPRO19 VANPLN26
IVERBN26 IADJN26 MVRS26 CDIC14 CDID14 CDIA26 CDIB26
CDID26 CDIG26 MRLS26;

OUTPUT: SAMPSTAT;
STANDARDIZED;
TECH1;

MODEL: f1 BY TRIM4 TRIM6 HILO6 HILO9;
f2 BY MISPRO19 VANPLN26 IVERBN26 IADJN26;
f3 BY CDIC14 CDID14 CDIA26 CDIB26 CDID26 CDIG26 MRLS26;
f2 ON f1;
f3 ON f2;
f1 f2 f3 ON GA MVRS26;

MODEL INDIRECT: f3 IND f2 f1;

!f1=Sound-Structure Sensitivities, f2=Word Form Representations, f3=Language
!Outcomes

MODEL 4: F₂ on F₁, F₃ on F₂, F₃ on_{ind} F₁, GA_{cov}, and MVRS Score_{cov}, and MaEd_{cov}

TITLE: SEM with two predictor latent variables (Sound-Structure Sensitivities and Later Word Form Representations) and GA, MVRS Score, and MaEd as covariates

DATA: FILE = all_data_imputed.dat;

VARIABLE: NAMES = ID BS SEX GA BW BO MAED PAED TRIM4 TRIM6
HILO6 HILO9 MISPRO19 VANPLN26 IVERBN26 IADJN26
PATTC12 FACEC12 MVRS26 MFMS26 CDIA14 CDIB14 CDIC14
CDID14 CDIE14 CDIF14 CDIG14 CDIH14 CDIA26 CDIB26
CDIC26 CDID26 CDIE26 CDIF26 CDIG26 MRLS26;

USEVARIABLES = GA MAED TRIM4 TRIM6 HILO6 HILO9 MISPRO19
VANPLN26 IVERBN26 IADJN26 MVRS26 CDIC14 CDID14 CDIA26
CDIB26 CDID26 CDIG26 MRLS26;

OUTPUT: SAMPSTAT;
STANDARDIZED;
TECH1;

MODEL: f1 BY TRIM4 TRIM6 HILO6 HILO9;
f2 BY MISPRO19 VANPLN26 IVERBN26 IADJN26;
f3 BY CDIC14 CDID14 CDIA26 CDIB26 CDID26 CDIG26 MRLS26;
f2 ON f1;
f3 ON f2;
f1 f2 f3 ON GA MVRS26 MAED;

MODEL INDIRECT: f3 IND f2 f1;

!f1=Sound-Structure Sensitivities, f2=Word Form Representations, f3=Language
!Outcomes