Waves of Change: Longitudinal Growth Profiling of Bilingual (Spanish-English) Language Development

---

A Dissertation
Submitted to
The Temple University Graduate Board

---

in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

---

by
Raúl Rojas
January, 2011

Examining Committee Members:

Aquiles Iglesias, Adivsory Chair, Communication Sciences and Disorders
Brian Goldstein, Communication Sciences and Disorders
Aneta Pavlenko, Curriculum, Instruction, and Technology in Education
Jennifer Cromley, External Member, Psychological Studies in Education

ABSTRACT

Although the research literature supports the notion of language growth trajectories, primarily in monolingual English children, the shape and direction of English-language learners' (ELLs) language growth trajectories are largely unknown. The present study examined the shape of ELLs' language growth trajectories by estimating the initial status and the growth rates of specific oral language skills (mean length of utterance in words (MLUw), number of different words (NDW), and words per minute (WPM)) in each language during the first 3 years of formal schooling. This study was framed from the perspective of language as a dynamic system, composed of linguistic subsystems that change over time.

This study utilized secondary data from a larger project, the Bilingual Language Literacy Project (BLLP), which collected narrative retell language samples produced in Spanish and English from ELL children. The final longitudinal dataset used in this study consisted of 12,248 oral narrative language samples (6,516 Spanish; 5,732 English) that were produced by 1,723 ELLs. This study examined the effect of three predictors on language growth: academic semester (metric of time), gender, and schooling. Growth curve model (GCM) testing was used to profile the longitudinal growth of the ELLs' oral language skills in Spanish and English over time.

This study had a number of important findings regarding change over time, intra- and inter-individual variability, and the impact of initial status on growth. With regard to change over time: MLUw, NDW, and WPM demonstrated growth over time in Spanish

and English; the shapes of Spanish (curvilinear, non-monotonic, and continuous) and English growth (linear, non-monotonic, and discontinuous) were similar within-language; language growth in Spanish was predicted by academic semester and gender; and language growth in English was predicted by academic semester, gender, and schooling.

With regard to intra- and inter-individual variability: significant intra-individual differences in the growth of all the oral language measures, across each wave of measurement, were found for both languages; significant intra-individual differences in the initial status of participants for all the oral language measures were found for both languages; significant inter-individual differences in the growth rates were found for WPM-Spanish; and significant inter-individual differences in the growth rates were found for all the oral language measures in English.

With regard to the impact of initial status on growth: the growth of MLUw-Spanish was systematically related to initial status (lower performers at initial status may not catch up to higher performers); the growth of NDW- and WPM-Spanish were unrelated to its initial status (lower performers at initial status may, or may not catch up to higher performers); and the growth of MLUw-, NDW-, and WPM-English was systematically related to initial status (lower performers at initial status may catch up to higher performers).

With regard to the co-development of interconnected subsystems, qualitative observations (non-empirically tested) based on visual inspection and GCM estimates

provided initial insight into the possible co-development occurring within- and across-languages.

The present study broke new ground by specifying the shape of growth for MLUw, NDW, and WPM in the Spanish and English of ELLs during their first 3 years of formal schooling. The study had a number of methodological limitations that will guide and motivate future work on the language growth of ELLs.

ACKNOWLEDGEMENTS

This study has been supported in part by grant R305U01001 “Biological and Behavioral Variation in the Language Development of Spanish-speaking Children”, funded jointly by the National Institute of Child Health and Human Development, National Institutes of Health, and the US Department of Education’s Institute of Education Sciences.

I owe a debt of unending gratitude to the members of my Dissertation Committee. Aquiles, thank you for being you…relentless and unwavering. Amidst all the madness from 1997 until today, thank you for believing in me and for making all of this possible. I will never forget that I was “your Doc”. Brian, thank you for always being willing to help, and most of all, for listening. Aneta, thank you for your insight and candor regarding my work. Jennifer, thank you for teaching me everything I did not know about growth, change, and the wonders of longitudinal data.

To Cindy, it has been a long, arduous, yet also incredible 4 years. Thank you for being there every step of the way, and for allowing me to lean on you when I most needed it. You are my rock.

A mis padres y a mi hermano, gracias por todo el apoyo, las porras, y el cariño. Yo soy una reflección de ustedes, y la culminación de este Doctorado es un testamento de nuestra familia.

TABLE OF CONTENTS

Page

ABSTRACT iii

ACKNOWLEDGMENTS vi

LIST OF TABLES xii

LIST OF FIGURES xv

CHAPTER

1. THEORETICAL FRAMEWORK 1

Dynamic Systems Theory (DST) 2

Language: A Dynamic System 4

Change (Growth) Over Time 5

Intra- and Inter-Individual Variability 13

Impact of Initial Status on Growth 17

Co-Development of Interconnected Subsystems 21

Analyzing Language Growth 25

Purpose 30

2. METHODOLOGY 34

Parent Study and Bilingual Language Literacy Project (BLLP) 34

BLLP Narrative Language Sampling Protocol 36

BLLP Longitudinal Sampling 38

Secondary Data Used for the Present Study 40

Outcome Measures: Expressive Language Skills 44
Covariates: Academic Semester, Schooling, and Gender 46
Analytic Approach 47
Time Metric and Time-Centering 48
Exploratory Analyses 50
Growth Curve Model Testing 51
Goodness-of-Fit Indices 51
Unconditional Means (UM) Model 52
Unconditional Growth (UG) Models 53
Conditional Growth (CG) Models 55
Prototypical Growth Trajectory 58
Summary 59
3. RESULTS 60
Descriptive Statistics 61
Preliminary Steps 61
Growth Curve Modeling (GCM) 63
Mean Length of Utterance in Words (MLUw) – Spanish 65
Unconditional Means (UM) Model 65
Unconditional Growth (UG) Model Testing 66
Conditional Growth (CG) Model Testing 68
MLUw-Spanish: Final Growth Curve Model (GCM) 73
MLUw-Spanish: Prototypical Growth Trajectory 80

Number of Different Words (NDW) – Spanish 82
Unconditional Means (UM) Model 82
Unconditional Growth (UG) Model Testing 82
Conditional Growth (CG) Model Testing..... 84
NDW-Spanish: Final Growth Curve Model (GCM) 89
NDW-Spanish: Prototypical Growth Trajectory 95
Words per Minute (WPM) – Spanish..... 97
Unconditional Means (UM) Model 97
Unconditional Growth (UG) Model Testing 97
Conditional Growth (CG) Model Testing..... 98
WPM-Spanish: Final Growth Curve Model (GCM) 103
WPM-Spanish: Prototypical Growth Trajectory 111
Mean Length of Utterance in Words (MLUw) – English 112
Unconditional Means (UM) Model 113
Unconditional Growth (UG) Model Testing 113
Conditional Growth (CG) Model Testing..... 116
MLUw-English: Final Growth Curve Model (GCM)..... 121
MLUw-English: Prototypical Growth Trajectory..... 127
Number of Different Words (NDW) – English..... 128
Unconditional Means (UM) Model 129
Unconditional Growth (UG) Model Testing 129
Conditional Growth (CG) Model Testing..... 131

NDW-English: Final Growth Curve Model (GCM) 135
NDW-English: Prototypical Growth Trajectory 141
Words per Minute (WPM) – English 142
Unconditional Means (UM) Model 143
Unconditional Growth (UG) Model Testing 144
Conditional Growth (CG) Model Testing 146
WPM-English: Final Growth Curve Model (GCM) 150
WPM-English: Prototypical Growth Trajectory 156
Summary 157
4. DISCUSSION 168
Dynamic Systems Theory (DST): Elements 169
Change (Growth) Over Time 169
Growth in Spanish 170
Growth in English 179
Summary of Growth in Spanish and English 185
Intra- and Inter-Individual Variability 185
Impact of Initial Status on Growth 190
Co-Development of Interconnected Subsystems 194
Across-Subsystem Co-Development Within a Language 196
Within-Subsystem Co-Development Across a Language 197
Limitations 199
Future Directions 200

REFERENCES CITED 204

APPENDICES

A. GLOSSARY OF GREEK TERMS (CROMLEY, 2008) 227

B. THE UNCONDITIONAL MEANS (UM) MODEL 228

C. THE UNCONDITIONAL GROWTH (UG) MODEL. 230

D. THE CONDITIONAL GROWTH (CG) MODEL 232

LIST OF TABLES

Table Page

1. BLLP Longitudinal Data Sampling Procedure 39

2. Language Sample and Chronological Age Descriptive Statistics per Semester 42

3. Means and Standard Deviations of Outcome Measures per Semester 62

4. Comparisons of Fixed and Randomly Varying Slopes for Unconditional Growth (UG) Models of Mean Length of Utterance in Words-Spanish 67

5. Comparisons of Gender *x* Slope Interactions for Conditional Cubic Growth Models with Gender (CG-Cb+G) of Mean Length of Utterance in Words-Spanish 70

6. Comparisons of Fixed and Randomly Varying Slopes for Conditional Discontinuous Growth (Dc-CG) Models of Mean Length of Utterance in Words-Spanish 71

7. Comparisons of Gender *x* Slope Interactions for Conditional Discontinuous Growth Models with Gender (Dc-CG+G) of Mean Length of Utterance in Words-Spanish 73

8. Comparisons of Growth Curve Model Parameter Estimates for Mean Length of Utterance in Words-Spanish 74

9. Comparisons of Fixed and Randomly Varying Slopes for Unconditional Growth (UG) Models of Number of Different Words-Spanish 83

10. Comparisons of Gender *x* Slope Interactions for Conditional Cubic Growth Models with Gender (CG-Cb+G) of Number of Different Words-Spanish 85

11. Comparisons of Fixed and Randomly Varying Slopes for Conditional Discontinuous Growth (Dc-CG) Models of Number of Different Words-Spanish 87

12. Comparisons of Gender *x* Slope Interactions for Conditional Discontinuous Growth Models with Gender (Dc-CG+G) of Number of Different Words-Spanish 88

13. Comparisons of Growth Curve Model Parameter Estimates for Number of Different Words-Spanish 90

14. Comparisons of Fixed and Randomly Varying Slopes for Unconditional Growth (UG) Models of Words per Minute-Spanish 99

15. Comparisons of Gender *x* Slope Interactions for Conditional Cubic Growth Models with Gender (CG-Cb+G) of Words per Minute-Spanish 101

16. Comparisons of Fixed and Randomly Varying Slopes for Conditional Discontinuous Growth (Dc-CG) Models of Words per Minute-Spanish 102

17. Comparisons of Gender *x* Slope Interactions for Conditional Discontinuous Growth Models with Gender (Dc-CG+G) of Words per Minute-Spanish 103

18. Comparisons of Growth Curve Model Parameter Estimates for Words per Minute-Spanish 105

19. Comparisons of Fixed and Randomly Varying Slopes for Unconditional Growth (UG) Models of Mean Length of Utterance in Words-English 115

20. Comparisons of Gender *x* Slope Interactions for Conditional Cubic Growth Models with Gender (CG-Cb+G) of Mean Length of Utterance in Words-English 117

21. Comparisons of Fixed and Randomly Varying Slopes for Conditional Discontinuous Growth (Dc-CG) Models of Mean Length of Utterance in Words-English 119

22. Comparisons of Gender *x* Slope Interactions for Conditional Discontinuous Growth Models with Gender (Dc-CG+G) of Mean Length of Utterance in Words-English 120

23. Comparisons of Growth Curve Model Parameter Estimates for Mean Length of Utterance in Words-English 122

24. Comparisons of Fixed and Randomly Varying Slopes for Unconditional Growth (UG) Models of Number of Different Words-English 130

25. Comparisons of Gender *x* Slope Interactions for Conditional Cubic Growth Models with Gender (CG-Cb+G) of Number of Different Words-English 132

26. Comparisons of Fixed and Randomly Varying Slopes for Conditional Discontinuous Growth (Dc-CG) Models of Number of Different Words-English 134

27. Comparisons of Gender *x* Slope Interactions for Conditional Discontinuous Growth Models with Gender (Dc-CG+G) of Number of Different Words-English..... 135

28. Comparisons of Growth Curve Model Parameter Estimates for Number of Different Words-English..... 136

29. Comparisons of Fixed and Randomly Varying Slopes for Unconditional Growth (UG) Models of Words per Minute-English..... 145

30. Comparisons of Gender *x* Slope Interactions for Conditional Cubic Growth Models with Gender (CG-Cb+G) of Words per Minute-English..... 147

31. Comparisons of Fixed and Randomly Varying Slopes for Conditional Discontinuous Growth (Dc-CG) Models of Words per Minute-English..... 148

32. Comparisons of Gender *x* Slope Interactions for Conditional Discontinuous Growth Models with Gender (Dc-CG+G) of Words per Minute-English 149

33. Comparisons of Growth Curve Model Parameter Estimates for Words per Minute-English 151

34. Final Growth Curve Model Parameter Estimates for MLUw, NDW, and WPM in Spanish and English 159

LIST OF FIGURES

Figure Page

1. Taxonomy of Growth Trajectories (Elman et al., 1999) 12

2. Linear Regression Fitted Growth Plots for Mean Length of Utterance in Words (MLUw), Number of Different Words (NDW), and Words Per Minute (WPM) in Spanish and English 64

3. Prototypical Growth Trajectories by Gender for Mean Length of Utterance in Words-Spanish 81

4. Prototypical Growth Trajectories by Gender for Number of Different Words-Spanish 96

5. Prototypical Growth Trajectories by Gender for Words per Minute-Spanish 112

6. Prototypical Growth Trajectories by Gender for Mean Length of Utterance in Words-English 128

7. Prototypical Growth Trajectories by Gender for Number of Different Words-English 143

8. Prototypical Growth Trajectories by Gender for Words per Minute-English 157

9. Prototypical Growth Trajectories by Gender for Mean Length of Utterance in Words (MLUw), Number of Different Words (NDW), and Words Per Minute (WPM) in Spanish and English 163

10. Prototypical Growth Trajectories by Gender for Mean Length of Utterance in Words (MLUw), Number of Different Words (NDW), and Words Per Minute (WPM) Collapsed Across Measures and Across Languages 164

CHAPTER 1

THEORETICAL FRAMEWORK

Language development research has demonstrated that as monolingual children get older, their overall language development follows a continuous and orderly trajectory (Gleason & Ratner, 2009; Hoff, 2005). This is not often the case for children in the US who come to school speaking a language other than English and whose English language skills are insufficient to function successfully in English-only classrooms. The children, classified as English-language learners (ELLs), speak languages that follow a range of varying developmental (growth) trajectories. The growth trajectories of both languages may simultaneously increase or one trajectory may increase at the expense of the other, as in the case of native language attrition (Anderson, 2004; Kohnert, 2008). The rate of growth or decline in the domains of each language is also variable. For some children, the rate of growth of the second language might not be sufficient to ensure that the child is sufficiently competent in English to be successful in school (Abedi, 2006; Genesee, Lindholm-Leary, Saunders, & Christian, 2005; Hopstock & Stephenson, 2003; Swanson, 2009).

Although the research literature supports the notion of language growth trajectories, primarily in monolingual English children (Bauer, Goldfield, & Reznick, 2002; Beitchman et al., 2008; Bornstein, Hahn, & Hayes, 2004; Farkas & Beron, 2004; Hadley & Holt, 2006; Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991; Law, Tomblin, & Zhang, 2008; Pan, Rowe, & Snow, 2005; Rice, Hoffman, & Wexler, 2009;

Rice, Redmond, & Hoffman, 2006; Rice, Wexler, & Hershberger, 1998; Smit, Hand, Freilinger, Bernthal, & Bird, 1990), the shape and direction of ELLs' language growth trajectories are largely unknown (Hammer, Lawrence, & Miccio, 2008b). To understand the growth trajectories of ELLs' languages and the factors that may influence those trajectories, we must identify how each language and their individual subsystems *change across time*. Measuring the beginning and ending points of language development as absolute values is not ideal because the results of two time-point design studies provide an unreliable measure of change, and contribute limited information with respect to the differing trajectories of growth that occur along the way (Huttenlocher, et al., 1991). In essence, what must be measured in order to understand language development in the ELL population is the growth, over multiple observations, of a continuously changing and complex system composed of two languages and their individual linguistic subsystems.

The purpose of this study was to examine the growth trajectories of ELLs' language by estimating the initial status and the growth rates of specific oral language skills in each language during the first 3 years of formal schooling. Specifically, this study focuses on the language growth in Spanish and English of ELLs during their first 3 years in school. This study is framed from the perspective of language as a dynamic system, composed of linguistic subsystems that change over time.

Dynamic Systems Theory (DST)

Dynamic systems are self-organizing entities that are termed *dynamic* because their unifying characteristic is that they change over time. Dynamic systems are also

complex, typically composed of subsystems whose interactions result in the overall behavior of the system. These subsystems also change over time as they interact with one another and the environment. Dynamism and complexity characterize much of (if not all) human behavior which, through advances in mathematics and the developmental and cognitive sciences in the 1990s, gave rise to Dynamic Systems Theory (DST) (Kelso, 1994; Port & van Gelder, 1995; Thelen & Smith; 1994; Thelen & Ulrich, 1991; van Geert, 1994).

DST outlines four principal elements that characterize nonlinear complex dynamic systems (De Bot, Lowie, & Verspoor, 2007a; Elman et al., 1999; Heredina & Jessner, 2002; Howe & Lewis, 2005; Kohnert, 2008; Larsen-Freeman, 2007; Marin & Peltzer-Karpf, 2009; Smith & Thelen, 2003; van Geert, 2008; van Gelder, 1998): (1) the emphasis of dynamic systems is on *change over time,* the overarching and unifying principle of DST, not on any one static state; (2) *intra- and inter-individual variability* is an intrinsic characteristic of dynamic systems that provides critical information regarding the ongoing changes in the system, which is represented by *individual differences*; (3) *initial states (status) impact(s) the developmental trajectory of dynamic systems* (i.e., known as the sensitivity to initial conditions), where small differences at the beginning may result in major changes over time, which in turn are representative of the non-linearity of dynamic systems; and (4) dynamic systems are complex in that they possess interconnected *subsystems that co-develop* over time, where the changes of one dynamic subsystem are likely to impact the changes of another subsystem, resulting in states of

parallel, dynamic change (connected growers), or ongoing fluctuations between steady states and states of dynamic change or opposing states of positive and negative change (phase transitions).

Language: A Dynamic System

Language, like any dynamic system is both dynamic and complex, composed of several linguistic subsystems (linguistic domains) that develop over time, (De Bot et al., 2007a; De Bot, Lowie, & Verspoor, 2007b; Elman, 1995; Evans, 2001; Hohenberger & Peltzer-Karpf, 2009; Thelen & Bates, 2003; & van Geert, 1994, 1998, 2007, 2008). Language, conceptualized as a dynamic system, consists of multiple linguistic subsystems. Although a number of linguistic subsystems (morphology, syntax, lexicon, phonology, pragmatics) have been examined in the language development literature, two of the most frequently studied, from a developmental perspective in monolingual and bilingual speakers, are morphosyntax and the lexicon (e.g., Beitchman et al., 2008; Bornstein, Hahn, & Haynes, 2004; Huttenlocher et al., 1991; Kohnert, Kan, & Conboy, 2010; Miller, 1987; Miller & Chapman, 1981; Simon-Cereijido & Gutiérrez-Clellen, 2009; Weismer, Murray-Branch, & Miller, 1994). The linguistic subsystems of morphosyntax and the lexicon can be represented by numerous oral language measures. Two oral language measures commonly used in language development research to capture the two subsystems are mean length of utterance in words (MLUw) and number of different words (NDW) (Carias & Ingram, 2006; Dollaghan et al., 1999; Eisenberg, Fersko, & Lundgren, 2001; Gutiérrez-Clellen, Restrepo, Bedore, Peña, & Anderson,

2000; Miller et al., 2006; Kohnert et al., 2010; Yip & Matthews, 2006). The global integration of multiple linguistic subsystems is often captured by verbal fluency in research based on language sample analysis (LSA), measured in words per minute (WPM) (e.g., Heilmann et al., 2008; Price, Hendricks, & Cook, 2010; Tilstra & McMaster, 2007). As noted by Miller et al. (2006):

> [verbal fluency] reflects all aspects of a speaker's task, developing the sequence of though to be spoken, selecting the words, grammatical structures, and sequencing the utterances into a coherent narrative all the while monitoring the listener for clues to understanding and revising on the spot. (p. 34)

These three oral language measures (MLUw, NDW, and WPM), which have been positively associated with grade and reading achievement in ELLs (Miller et al., 2006), provide a composite (yet not comprehensive) image of a language system.

What follows is a summary of the existing literature on language growth, focused on the language growth of ELL children and framed within each of the four principal elements of DST.

Change (Growth) Over Time

Studies conducted prior to the advent of advanced statistical techniques specialized for longitudinal data analyses, such as growth curve modeling, have attempted to answer questions about language development by identifying starting and/or ending levels of the children's language development, or exploring factors that could impact these absolute values which positively develop with age (e.g., Brown, 1973; de

Villiers & de Villiers, 1973; McCarthy, 1930; Miller, 1987; Miller & Chapman, 1981; Shriner, 1969; Templin, 1957; Velten, 1943; Weismer, Murray-Branch, & Miller, 1994). Contemporary work that uses specialized methods for longitudinal data analyses (e.g., Farkas & Beron, 2004; Kan, 2010; Rice et al., 2009), attempts to answer questions about language development by describing and predicting the change of language over time (language growth), rather than focusing on absolute values occurring at static points in time. Regardless of the analytical approach, however, past and present studies on language growth have dealt with the first DST element: change over time.

A common characteristic of studies that have examined the change of language over time is their need for predictors of language growth; otherwise, their predictive conclusions are limited. Predictors of change over time can be time-invariant, in that they do not change over time (e.g., ethnicity), or they can be time-varying (e.g., annual income), in that they change over time (Singer & Willett, 2003). Gender is often used as a time-invariant predictor in studies of early language development, and significant gender-based differences have been identified in cross-sectional and longitudinal studies of language development (Bauer, Goldfield, & Reznick, 2002; Bornstein, Hahn, & Haynes, 2004; Bouchard, Trudeau, Sutton, Boudreault, & Deneault, 2009; Fenson et al., 1994; Hammer, Farkas, & Maczuga, 2010; Huttenlocher et al., 1991). In a large-scale, cross-sectional monolingual English study (Fenson et al., 1994) using the MacArthur Communicative Development Inventories (CDIs) (Fenson et al., 1993), girls tended to show higher levels of lexical, grammatical, and gestural development, relative to boys,

across infants of ages 8-to-16 months, and toddlers ages 16-to-30 months. Similarly, in a cross-sectional study of language development of monolingual French Canadian infants and toddlers (Bouchard et al., 2009) of 8-to-30 months of age, girls' language skills outperformed boys' language skills (produced more words; used more grammatical forms; demonstrated more syntactic complexity) at all age ranges. Bouchard et al. (2009) contemplate the need for developing gender-based norms for the overall language skills of monolingual Quebec French speaking children in order to prevent systematically overestimating (girls) or underestimating (boys) communication skills.

Multiple studies with monolingual children have confirmed the pattern of girls exhibiting advanced linguistic development compared to boys, which seems to attenuate around the age of 3;0 (e.g., Bauer et al., 2002; Bornstein et al., 2004; Fenson et al., 1994; Huttenlocher et al., 1991), with the pattern potentially reappearing between the ages of 10;0 to 11;0 (Coates, 1993). Longitudinal work on the language growth of ELLs, however, suggests different gender-based patterns of language development. A 3-year, longitudinal study investigating the impact of maternal language use on the vocabulary and early literacy skills of native Spanish speaking children (Hammer, Davison, & Miccio, 2009), found that gender did not significantly impact vocabulary or early literacy skill growth in either Spanish or English. The lack of gender-based differences across languages was unexpected, as Hammer et al. (2009) hypothesized that girls would demonstrate higher initial levels and more growth of Spanish language skills over time. This hypothesis was based on prior findings (Hammer, Lawrence, Davison, & Miccio,

2007) of mothers predominantly using Spanish with their daughters, and on the research literature (De Von Figueroa-Moseley, Ramey, Keltner, & Lanzi, 2006; McHale, Updegraff, Shanahan, Crouter, & Killoren, 2005; Zentella, 1997) suggesting that Hispanic families often socialize Hispanic children according to traditional gender roles: girls are expected to become mothers that stay home and raise children; boys are expected to become financial providers that work away from home. Therefore during childhood, Hispanic girls are more immersed in Spanish at home, while boys are encouraged to spend time outside the home (less immersed in Spanish at home). Hammer et al. (2009) concluded that the lack of gender-based differences, particularly in Spanish, might have been associated with the participants' immersion in English as they attended English-only Head Start and kindergarten programs.

The work of Uchikoshi (2006b) on the longitudinal English vocabulary development of ELLs during kindergarten provides contrasting gender-based findings. Among a wide range of predictors of the English receptive and expressive growth of ELLs, Uchikoshi (2006b) found that boys demonstrated higher initial levels and more growth of English receptive and expressive vocabulary skills, relative to girls. Additional contributing factors included higher levels of initial Spanish receptive vocabulary, being exposed to books at home, and attending preschool. Uchikoshi speculates that the receptive and expressive vocabulary advantage of boys may be related to boys interacting more often with English speakers in the community, and of Hispanic parents stressing more to boys the importance of learning English in school.

The pattern of girls exhibiting advanced early linguistic development compared to boys found in monolinguals, and the inconclusive gender-based differences of language growth in bilinguals, establish gender as a critical time-invariant factor that should be considered as a predictor of language skills that change over time.

Time-varying predictors change over time, and they can impact the overall trajectory (shape) of language growth over time, sometimes resulting in discontinuous growth (e.g., Bleise, Chan, & Ployhart, 2007; Perkins, Brutten, & Grass, 1996; Wu, West, & Hughes, 2008). One such time-varying predictor is whether participants have not had schooling in between academic semesters (summer vacation). When change over time (growth) is studied in children during their years in school, it is important to consider the potential effect of not having attended school during the summer months between fall and spring semesters. Prior research on the change over time (growth) in reading and math skills (e.g., Alexander, Entwisle, & Olson, 2001; meta-analysis by Cooper, Nye, Charlton, Lindsay, & Greathouse, 1996; McCoach, O'Connell, Reis, & Levitt, 2006) has demonstrated that not having had schooling in between academic semesters (summer vacation) tends to negatively impact the growth of math and reading skills differentially, with math skills being more negatively affected than reading. The work of Hammer et al. (2008a,b) suggests a more complex relationship between the effect of having or not having had schooling in between academic semesters (the effect of schooling). In a longitudinal study of ELL preschool children enrolled in Head Start programs for 2 years, Hammer et al. (2008a) found that ELLs exposed to Spanish and English before entering

Head Start had significantly higher receptive language skills (measured as raw scores) in English than ELLs exposed exclusively to Spanish. Participants who entered Head Start with prior exposure to the two languages maintained their English raw score advantage over the duration of the study. Conversely, ELLs exposed exclusively to Spanish before Head Start entry had significantly higher receptive language skills (measured as raw scores) in Spanish than ELLs exposed to Spanish and English, and they maintained their Spanish raw score advantage over the duration of the study (Hammer et al., 2008a).

Although linear growth in raw scores was predicted for both groups in both languages, Hammer et al. (2008a) noted wide variations in standard scores, including a period of negative growth (decline) on receptive language measures in Spanish. This variation in growth patterns prompted Hammer et al. (2008b) to investigate the potential effect of summer vacation on the receptive language skills of ELLs during the same 2 years in Head Start. The results indicated that summer vacation positively impacted the receptive language skill growth in both languages (i.e., standard score growth between the spring and fall semesters) for participants that had demonstrated negative growth of their languages during the first year, and negatively impacted the growth in Spanish and English (i.e., standard score decline between the spring and fall semesters) for those that had previously demonstrated positive growth of their languages. However, these contrasting growth effects were limited to the summer vacation, as the preceding patterns of growth during the fall and spring semesters of the first year continued during the second year. Thus, participants that had demonstrated negative growth of their languages

during the first year, continued to show negative growth of their languages during the second year after a period of positive growth during the summer; participants that had demonstrated positive growth of their languages during the first year, continued to show positive growth of their languages during the second year after a period of negative growth during the summer (Hammer et al., 2008b). These contrasting patterns of change over time resulted in growth trajectories with periods of discontinuous growth during the summer. In sum, although the receptive language skills of ELLs demonstrated either overall growth or decline during the 2 years of Head Start, the shape of this change over time was discontinuous for Spanish and English.

The work of Hammer et al. (2008b) and others (e.g., Hadley & Holt, 2006; Huttenlocher et al., 1991; Kan, 2010; Rice et al., 2006, 2009), illustrates the importance of considering not only the overall growth rates of language subsystems, but also the shape of their growth trajectories. Elman et al. (1999) developed a growth trajectory taxonomy focused on the *shape* of change within different growth patterns. This taxonomy of growth trajectories specified three nested dimensions that could account for varying and complex growth patterns across time: linearity, direction, and continuity. Figure 1 outlines Elman et al. (1999) taxonomy, which specifies the dimensions of growth trajectories. The dimensions (linearity, direction, and continuity) are nested in that they describe trajectories of growth that integrate each of the three dimensions. For example, the well-known vocabulary explosion in young toddlers (e.g., Gleason, & Ratner, 2009; McMurray, 2007; Paul, 2006) could be described as a curvilinear,

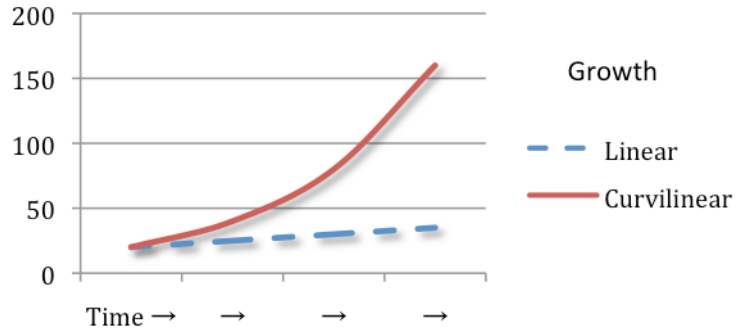
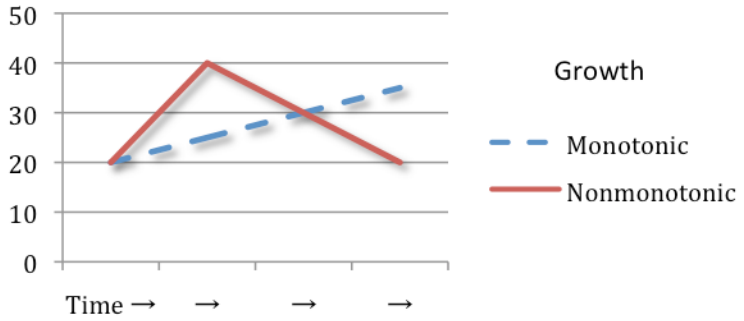
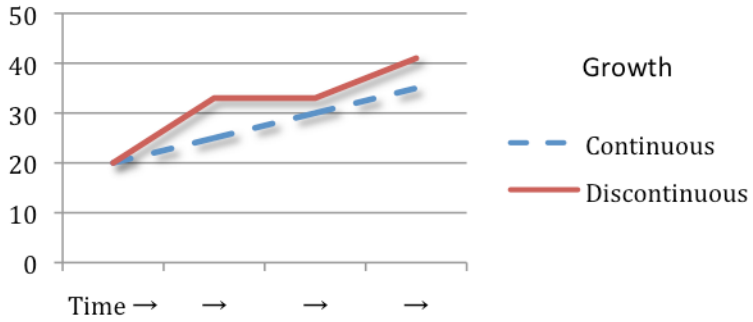
Dimension	Range	Example
Linearity	Linear Vs. Curvilinear	
Direction	Monotonic Vs. Non-monotonic	
Continuity	Continuous Vs. Discontinuous	

Figure 1. Taxonomy of Growth Trajectories (Elman et al., 1999)

monotonic, and continuous growth trajectory. On the other hand, the impact of lack of academic instruction during the summer months on receptive language skills (e.g., Hammer et al., 2008b) can be described as a linear, non-monotonic, and discontinuous growth trajectory. Each of these dimensions categorizes a specific aspect of growth.

Linearity specifies whether growth is gradual, periodic, and steady (i.e., linear) or whether it is characterized by instantaneous, exponential accelerations or decelerations (i.e., curvilinear). Direction specifies whether growth is consistently increasing or decreasing (i.e., monotonic) or whether it is characterized by alternating periods of positive and negative growth (i.e., non-monotonic). Continuity specifies whether quantitative changes in growth are consistent (i.e., continuous) or whether there are periods characterized by sudden shifts of inconsistent positive or negative growth (i.e., discontinuous). The three nested dimensions of linearity, direction, and continuity of Elman et al. (1999) growth trajectory taxonomy can be used to categorize the growth from diverse developmental domains, including trajectories of language growth.

Intra- and Inter-Individual Variability

A second critical element of DST is variability at the intra- and inter-individual level, which are representative of individual differences in growth over time. Although the overall child development research literature has traditionally underemphasized the importance of variability and individual differences (Siegler, 2002; van Geert & van Dijk, 2002), a number of studies have cast a spotlight on the importance of studying individual differences in language growth. Fenson et al. (1994) seminal study based on a large-scale

cross-sectional sample of MacArthur CDIs (Fenson et al., 1993) clearly illustrates the role of individual differences in the early communicative development of English speaking infants (8-16 months) and toddlers (8-30 months). One of the hallmarks of the MacArthur CDI study was the significant variability identified across children in the development of their lexical, grammatical, and gestural linguistic domains, which challenged the notion of the "modal" child (Bates & Carnevale, 1993; Bates, Dale, & Thal, 1995; Fenson et al., 2000; Fenson et al., 1994). Bates et al. (1995) stressed, "there are enormous differences individual differences in the onset time and rate of growth" (p. 96), when discussing the range of individual differences in linguistic domains indentified in Fenson et al. (1994) MacArthur CDI study. For example, the *SD*s for vocabulary comprehension nearly match the means at 10 and 11 months, after which point the *SD*s decrease. Although the same *SD*-mean matching pattern applies to sentence complexity, it does not occur until 22 months, and it continues at 23, 24, and 25 months (Fenson et al., 2000).

Contemporary studies of language development have begun to address intra- and inter-individual variability of individual differences by using growth curve modeling (GCM) to analyze longitudinal data. Rather than assigning variance to be entirely measurement error, GCM partitions variance into within- and between-subject variance. In addition, depending on the specific growth curve model in question, variance can be independently attributed to the onset of growth (initial status), the rate(s) of growth, and to their covariance (Hammer et al., 2008a; Singer & Willett, 2003); covariance explains

how initial status can impact the growth rate, which addresses the third element of DST. The majority of studies of bilingual language growth that have used GCM (e.g., Kan, 2010; Uchikoshi, 2006b), however, have primarily focused on interpreting change over time (i.e., expected average onset of growth and growth rates) instead of also interpreting findings of the intra- and inter-individual variability of individual differences. Notable exceptions can be found in the work of Hadley and Holt (2005) with monolingual children, the work of Rice and colleagues (Rice et al., 1998, 2009) with monolingual children diagnosed with specific language impairment (SLI), and in the work of Hammer et al. (2008a,b) with ELLs.

Hadley and Holt's (2005) 1-year longitudinal study investigated individual differences in the onset of verb tense marking of slowly developing monolingual language learners of English. The results, which emphasized the inter-individual variance components of their data, found significant inter-individual differences in the onset of tense marking, and significant inter-individual differences in the growth rates of onset tense marking. Although Hadley and Holt (2005) determined that a positive history of speech, language, or learning disability and MLUm accounted for the variability in onset of tense marking, they recognized that their restricted sample size (N = 22) limited the potential variability present in the population of slowly developing monolingual language learners. Rice et al. (1998, 2006, 2009) work on the delayed morphosyntactic (verb marking; MLUm level) skills of monolingual SLI children, found differential patterns of inter-individual differences based on specific outcome measures. Rice et al.

(1998) found significant inter-individual differences in the growth rates of marking third-person singular *–s* in young (followed from 4;6 to 8;9) SLI children, but no significant intra-individual differences in growth or inter-individual differences in initial status. Rice et al. (2006) found significant intra-individual differences in the growth of each participant's MLUm level at each measurement point, significant inter-individual differences in initial status, and significant inter-individual differences in the growth rates of MLUm. Rice et al. (2009) found significant intra-individual differences in the growth of each participant's ability to judge grammatical verb tense marking in sentences with distinct question types at each measurement point, significant inter-individual differences in the initial status of judging grammatical verb tense marking in sentences with *Yes/No* and *Do* questions, and significant inter-individual differences in the initial status *and* the growth rates of judging tense marking in sentences with *Wh-* questions in the same SLI children when they were older (followed from 8;0 to 15;0). The intra- and inter-individual differences revealed by the findings of Rice et al. (1998, 2006, 2009) confirmed the inconsistent patterns of development that characterize the fractional language development of SLI children (Bedore &Leonard, 1998; Leonard, 2000), which has been compared to the fractional language skills of ELLs undergoing native language attrition (Anderson, 2004; Kohnert, 2008). The work of Hammer et al. (2008a) on ELLs' receptive language growth in Spanish and English, which did not set out to focus on individual differences, resulted in findings of remarkable variation that warranted a further study (Hammer et al., 2008b). Group membership criterion was determined by

the patterns of variation detected during the first year of Head Start (positive growth versus negative growth of language skills).

In summary, the work of Hadley and Holt (2005) is of note because its central aim was to study individual differences. The work of Rice et al. (1998, 2006, 2009) is of particular import because of the range of differential findings regarding intra- individual variability in growth, and inter-individual variability in initial status and growth rates, that were found across different measures of the morphosyntactic subsystem within the *same* participants over time. The work of Hammer et al. (2008a,b) is of note because it represents a series of related studies linked by unexpected findings of individual differences. Considering the role of variability in studies of language growth answers the growing call (Bates et al., 1995; Fenson et al., 1994; Hadley and Holt, 2005; Siegler, 2002) for no longer neglecting individual differences and for examining the heterogeneity of language learners.

Impact of Initial Status on Growth

Another important element of DST is the impact that initial status can have on developmental growth trajectories, where differences at the onset of growth can result in systematically different growth trajectories over time. This phenomenon of growth is detected by the intercept-slope covariance in longitudinal data, which estimates the strength and direction of the relationship between the initial status of growth (intercept) and the rate of growth (slope), after controlling for group or program membership (Singer & Willett, 2003). The work of Seltzer, Choi, and Thum (2003) provides a helpful

summary of the relationship between where growth begins and how it eventually grows or declines. Intercept-slope covariances can be positive and significant, negative and significant, or non-significant (unrelated). A positive and significant intercept-slope covariance indicates that higher onsets of growth (initial status) result in growth trajectories that increase at faster rates, on average, relative to the growth trajectories that result from lower onsets of growth. In addition, a positive and significant intercept-slope covariance indicates that differences in initial status will become more pronounced over time. In contrast, a negative and significant intercept-slope covariance indicates that higher onsets of growth (initial status) result in growth trajectories that decrease at faster rates, on average, relative to the growth trajectories that result from lower onsets of growth. In addition, a negative and significant intercept-slope covariance indicates that differences in initial status will become less pronounced over time (Seltzer et al., 2003). Positive and significant intercept-slope covariances are representative of scenarios where the growth of initially lower-performing participants does *not catch up* to the growth of initially higher-performing participants (e.g., Rice et al., 1998), and negative and significant intercept-slope covariances are representative of scenarios where the growth of initially lower-performing participants *catches up* to the growth of initially higher-performing participants (e.g., Dotterer, McHale, & Crouter, 2009). However, the initial status-growth rate covariance is not always significant (e.g., Rice et al., 2009). Non-significant (unrelated) intercept-slope covariances indicate the lack of a systematic relationship between the onset of growth (intercept) and the rate of growth over time (slope) (Seltzer et al., 2003). Therefore, the growth rates may vary (or may be similar)

among participants (irrespective of their group or program membership), regardless of their initial status of growth. In addition, a non-significant intercept-slope covariance indicates that differences in initial status are expected to hold steady over time.

The impact of initial status on growth trajectories is often found in the educational research literature (e.g., Aarnoutse & van Leeuwe, 2000; Ding & Davison, 2005). A study that compared the growth in math skills in low versus high achieving students (Ding & Davison, 2005) found a negative and significant intercept-slope covariance between math skills at initial status and the growth of math skills over a 4-year period in two different cohorts of primarily monolingual children. This finding suggested that low achieving students at initial status demonstrated faster growth rates, on average, than high achievers. The finding also suggested that that the math skills of initially lower achieving students would eventually catch up to the math skills of high achievers at initial status, and that the initial differences in achievement would diminish over time. Examples of intercept-slope covariance data can also be found in the language development literature, particularly among studies (Rice et al., 1998, 2006, 2009) that compare the growth of typically developing (TD) children and children with specific language impairment (SLI).

Rice et al. (1998) carried out what is arguably one of the earliest studies in communication sciences and disorders to use GCM, focusing on the longitudinal growth of tense acquisition by monolingual TD children and monolingual children with SLI. This pioneering study found that the ability to mark verb tense was significantly higher for TD children than SLI children at the beginning of data collection, and that this

difference in initial status determined parallel growth trajectories for both groups of children, albeit the tense marking skills of SLI children never caught up to the skills of TD children. Although Rice et al. (1998) did not report the intercept-slope covariance data, the findings lead one to believe that a positive and significant covariance between initial status and growth rate was present. In a follow-up study (Rice et al., 2009), SLI participants from Rice et al. (1998) were followed for several years in order to determine if they would continue to demonstrate difficulty with finiteness marking in sentences into later childhood and adolescence, or whether they would eventually catch up to their MLUm-matched TD peers. The results of this follow-up study were strikingly similar to the findings of Rice et al. (1998). The TD and SLI participants again demonstrated parallel growth trajectories, but their initial levels of finiteness marking in sentences were disparate, with the TD participants demonstrating significantly higher success in making appropriate judgments of finiteness marking at the beginning of data collection. Rice et al. (2009) conclude that older SLI participants continue to show slower growth of finiteness marking mastery well into adolescence, and that SLI participants would perhaps never catch up to their MLUm-matched TD peers—even into adulthood. Of import here is that the intercept-slope (i.e., initial status-growth rate) covariance reported by Rice et al. (2009) was *negative* and *non-significant* (i.e., the initial status and growth trajectories were not systematically related). Therefore, the conclusions of Rice et al. (2009) can be extended to speculate that although SLI participants may not catch up to TD participants, they may not fall further behind either. The same conclusions can be extended to yet another study (Rice et al., 2006) that used the SLI participants from Rice

et al. (1998), which found another negative and significant intercept-slope covariance for MLUm in English. Nevertheless, the work of Rice et al. (1998) demonstrates the importance of considering the impact of initial status on growth trajectories, and of reporting (Rice et al., 2006, 2009) intercept-slope covariance data to gauge this relationship.

Co-Development of Interconnected Subsystems

One of the key elements of a dynamic system is the complexity of having multiple interconnected subsystems that co-develop over time, where the change of one subsystem can have a dynamic impact on other subsystems and vice versa. Bates et al. (1995) argue for the importance of considering, "the degree of association or dissociation in the rate of development" (p. 2), that can occur between linguistic subsystems. In particular, Bates et al. (1995) emphasize the *temporal asynchrony* that occurs between lexical and morphosyntactic functions during early language development, when children do not begin to use morphological inflections until they have amassed sufficient words in their lexicon to surpass the single-word stage of language production. Based on these arguments, Bates et al. (1995) were essentially describing the need to consider the co-development of interconnected subsystems from a temporal perspective – the perspective of language growth. Although there is an emerging body of research on the relationships between the distinct linguistic subsystems in ELLs (Bedore, Fiestas, Peña, & Nagy, 2006; Carias & Ingram, 2006; Castilla, Restrepo, & Perez-Leroux, 2009; Conboy & Thal, 2006; Hammer, Lawrence, & Miccio, 2007; Kohnert, Kan, & Conboy, 2010; Miller et al., 2006;

Simon-Cereijido et al., 2009), not all of this work is framed within the temporal perspective of language growth.

In a cross-linguistic study comparing the use of mazes (filled pauses, repetitions, revisions) in Spanish and English narrative language samples, Bedore et al. (2006) found a positive association between maze use and oral language measures of morphosyntactic (MLUw) and lexical (NDW) subsystems across Spanish-English bilingual and functional monolingual (Spanish; English) children of varying ages. MLUw was positively correlated with overall maze use in Spanish and with filled pauses in English; NDW was positively correlated with grammatical revisions in Spanish and with filled pauses in English. Similar results were found in another study (Carias & Ingram, 2006) comparing the disfluencies (repetitions, revisions, insertions, and prolongations) of ELLs in Spanish and English. Carias and Ingram (2006) found a positive relationship between MLUw and disfluency. ELLs with the highest levels of MLUw demonstrated the most disfluencies. Although the findings of Bedore et al. (2006) and Carias and Ingram (2006) provide insight into the relationships between the linguistic subsystems of two different language systems and verbal fluency, they are not from the temporal perspective of language growth. The work of Kohnert et al. (2010) and Simon-Cereijido et al. (2009) examined the potential relationships between the linguistic subsystems of morphosyntax (MLUw) and the lexicon (NDW) in distinct types of ELL children; Kohnert et al. (2010) studied Hmong-English ELLs and Simon-Cereijido et al. (2009) studied Spanish-English ELLs. The findings of both studies were similar with regard to finding positively significant

cross-domain relationships between MLUw and NDW *within* each language, but neither study detected significant cross-domain relationships *across* languages. As with Bedore et al. (2006), Kohnert et al. (2010) and Simon-Cereijido et al. (2009) suggest that their findings provide evidence for related (interconnected) linguistic subsystems within each language system. The temporal perspective on language growth is necessary, however, in order to make conclusions about the dynamic co-development of interconnected linguistic subsystems.

A 12-month longitudinal study of the language development of Spanish-English bilingual children (Conboy et al., 2006) provided a temporal language growth perspective on the association between morphosyntactic and lexical subsystems. Although Conboy et al. (2006) study incorporated a key difference with regard to the participant population under investigation (i.e., participants were simultaneous bilinguals, consistently exposed to both languages in the home prior to 6 months of age), relative to the ELLs studied in Kohnert et al. (2010) and Simon-Cereijido et al. (2009), the associations between the growth of morphosyntactic and lexical subsystems were also positive and significant, and mostly restricted within-language. However, Conboy et al. (2006) found that utterance length in Spanish positively predicted growth in the number of English words, which was the only cross-language, cross-domain finding. A two time-point longitudinal study of ELLs in preschool (Castilla et al., 2009) found a range of relationships between grammatical and semantic measures in Spanish and English. Castilla et al. (2006) advance the notion of developmental interdependence, supported by finding that

grammatical and semantic measures in Spanish at Time 1 (beginning of preschool) significantly accounted for the variance in grammatical and semantic measures in English at Time 2 (end of preschool). In addition, Castilla et al. (2006) found both significant within-language, and across-language relationships between semantic and grammatical measures. Other studies have also provided evidence for the co-development of interconnected subsystems from a temporal language growth perspective in ELLs (e.g., Hammer et al., 2007; Laija-Rodriguez, Ochoa, & Parker, 2006; Miller et al., 2006; Uchikoshi, 2006a), but these have examined the relationship between oral language and reading skills. What is clear is that language growth over time must be considered in order to identify the dynamic co-development of interconnected language subsystems over time.

The existing literature on language growth demonstrates that the work to date can be characterized by an emphasis on: the language growth of monolingual children; change over time superseding individual differences, the impact of initial status on growth trajectories, and the dynamic co-development of interconnected subsystems; and studying the growth of one linguistic subsystem, rather than the growth of multiple linguistic subsystems. However, there is an emerging body of research on the language growth of ELL children, along with a growing need to contextualize the study of language growth from the perspective of language as a dynamic system. Clearly, a versatile analytic method is also needed in order to capture the four principal elements of language as a dynamic system.

Analyzing Language Growth

Longitudinal designs are the most appropriate methodological designs to analyze growth over time (Nagin, 2005; Rogosa, Brandt, & Zimowski, 1982; Singer & Willet, 2003; Wang & Bodner, 2007; Willet, 1989). A longitudinal design is characterized when measures of interest are obtained on more than one occasion from the same participant(s), over a prescribed period of time. Generally speaking, more than two instances of measurement over time are desirable in longitudinal studies. Otherwise, empirical claims addressing what happened in between the initial and final state (the actual change over time) cannot be made.

Although longitudinal designs are not a novelty in the study of language, many of the statistical techniques that are commonplace in contemporary research are primarily suitable to measure growth, but from a static perspective. Measuring the beginning and ending points of an outcome measure via two-time point longitudinal designs (e.g., pre- and post-treatment measures) is entirely possible using traditional statistical methods. The restrictions of traditional statistical methods, however, provide limited information regarding the growth patterns (i.e., shape and rate of growth at the intra- and inter-individual level) that occur along the way (Burchinal, Nelson, & Poe, 2006). Analysis of variance (ANOVA) is a widely used statistical method across disciplines that study human behavior (e.g., Psychology, Sociology, Communication Sciences and Disorders). ANOVA has multiple variants (e.g., multivariate ANOVA; factorial ANOVA) that are designed for specific analytical purposes. The strength of ANOVA is that it detects

significant differences between groups. Therefore, the analytic focus of ANOVA is between-person variation. However, when exploring change over time, questions about individual differences (i.e., within-person variation) are equally as important as group-level data. However, the analytic scope of ANOVA is limited in this regard.

Univariate and multivariate repeated measures ANOVA are two alternative approaches to analyze longitudinal data that account for intra- (within-person) and inter-individual (between-person) variation. Both of these approaches were originally designed to measure change over time (Burchinal, Nelson, & Poe, 2006). Although the univariate repeated measures approach measures individual differences in the intercept (i.e., initial status), it does not on the slope (i.e., rate of change). On the other hand, although the multivariate repeated measures approach allows for individual differences on the slope, it does not accept missing data (Tabachnick & Fidell, 2007). A shared limitation of univariate and multivariate repeated measures ANOVA is that participants with missing data, a common characteristic of longitudinal studies, are excluded from both measures (Burchinal, Nelson, & Poe, 2006). Finally, the univariate and multivariate repeated measures approaches only allow static predictors (e.g., gender) as covariates; they do not permit time-varying predictors (e.g., parental employment). Time-varying predictors can be important covariates, particularly if they are hypothesized to have a dynamic impact on development over time. An appropriate statistical method for measuring change over time should handle missing data and accommodate time-invariant

and time-varying predictors, which neither univariate nor multivariate repeated measures can accommodate.

Growth curve modeling (GCM), a specialized form of Hierarchical Linear Modeling (HLM), is a sophisticated and versatile statistical method poised to become an increasingly popular approach to study language development and disorders because it captures simple and complex patterns of growth and growth rates over time (Justice, 2009; Raudenbush & Bryk, 2002; Ram & Grimm, 2007; Singer & Willett, 2003). GCM associates the complex relationship of development, change, and time by analyzing measurement occasions nested within people and considering the starting point of growth (i.e., the intercept), the rate(s) of growth over time (i.e., the slope), and the shape/form of that growth over time (i.e., the trajectory). These factors allow GCM to appropriately capture developmental patterns of change at the intra-individual (i.e., within-person) level and to identify predictors of development at the inter-individual (i.e., between-person) level (Singer & Willet, 2003). The multifaceted quality of GCM permits a conceptual shift in terms of *how* different types questions can be asked about change, including language development (Hammer et al., 2007; Luke, 2004). GCM simultaneously integrates and answers questions such as: do participants begin around the same level (the intercept), or do participants begin at significantly different levels?; do participants demonstrate similar rates of growth (slopes), or do some participants show faster (or slower) growth rates than others?; do participants show positive growth only, or do some show negative growth (decay) or periods of stability?; are the shapes of the growth

trajectories strictly linear, or are some trajectories curvilinear?; are there individual differences in terms of growth over time, and can certain factors account for some of these differences?

GCM offers numerous advantages, relative to traditional statistical methods (i.e., factorial analysis of variance; univariate and multivariate repeated measures ANOVA), for capturing simple and complex patterns of change over time with longitudinal data designs (Burchinal, Nelson, & Poe, 2006; Tabachnick, & Fidell, 2007). The methodological requirements for conducting a GCM-based study are minor: a minimum of three periods of observation/data collection, a metric for time (e.g., days, weeks, therapy sessions), and an outcome measure with values that change over time (e.g., number of different words; Singer & Willet, 2003). With regard to the actual measurement of change, GCM offers various advantages: the shape of the growth trajectory can be modeled to be linear or curvilinear (e.g., quadratic); the growth trajectory can indicate positive growth or decay (negative growth) over time; intra- and inter-individual patterns of change are calculated simultaneously; and descriptive and predictive information is provided. GCM also offers a high-degree of methodological flexibility. Covariates (i.e., independent variables) can be time-invariant (e.g., gender) or time-varying (e.g., effect of schooling), and both types can be included (Luke, 2004; Zwietering, Jongenburger, Rombouts, & Riet, 1990). GCM accepts time-structured (i.e., equally spaced over time) and time-unstructured (i.e., unequally spaced over time) data (Singer & Willet, 2003). Longitudinal datasets often have missing observations from

various participants at different points in time. While participants with missing data are excluded from other statistical methods (e.g., univariate and multivariate repeated measures), GCM uses data from all participants (even those with only one observation at any point in time) by using maximum likelihood (ML) estimation or Bayesian multiple imputation (MI); both ML and MI handle missing data remarkably well and are considered optimal estimation methods for dealing with missing data (Baraldi & Enders, 2010; Cheung, 2007; Fichman & Cummings, 2003; Graham, 2009; Schafer & Graham, 2002; Singer & Willet, 2003; Swasey Washington, 2009). GCM estimates individual growth curves (for each participant in a dataset) and uses the parameters from each individual growth curve to simultaneously estimate a group growth curve (Raudenbush & Bryk, 2002; Singer & Willett, 2003). Maximum likelihood estimation estimates individual growth curves by using weighted data from the entire sample and from each individual participant. Burchinal, Nelson, and Poe (2006) call this parameter weighting, *borrowing strength*, which can, "greatly increase the precision of parameter estimates and power to identify predictors of developmental patterns" (p. 78).

This abbreviated list of GCM's advantages is reminiscent of the four principal elements of DST: emphasis on change over time; intra- and inter-individual variability represented by individual differences; initial status impacts the developmental trajectory of non-linear dynamic systems; and the dynamic co-development of interconnected subsystems. Although studies that use GCM to analyze longitudinal data are not yet the norm in the field of child language development, arguably because of the mathematical

rigor involved or because of their recency (Howe & Lewis, 2005; Marin & Peltzer-Karpf, 2009), the possibilities made available through GCM are gradually expanding its fledgling popularity (Justice, 2009). GCM's specialized design to analyze longitudinal data, which nests time within the person, makes it one of the most well equipped statistical tools for investigating language growth in ELLs based on longitudinal data (Hammer et al., 2007; Uchikoshi, 2006b).

Purpose

The development of language is a dynamic and complex process, involving linguistic domains that change over time, which result in varying developmental (growth) trajectories. Conceptualized as a dynamic system, the growth of language can be contextualized according to four principal elements of DST: the change (growth) of language over time; individual differences in language growth are indicated by intra- and inter-individual variability; initial status of growth can significantly impact growth trajectories; and interconnected linguistic subsystems may demonstrate dynamic co-development over time. Although the existing literature overwhelmingly centers on the language growth of monolingual children, an emerging body of research has begun to also address the language growth of ELL children. Due to these efforts, we know that the stability of gender-based growth trajectories (girls outpace boys) found in monolingual children until around the age of 3;0 (e.g., Bauer et al., 2002; Bornstein et al., 2004; Fenson et al., 1994; Huttenlocher et al., 1991), may not apply to the disparate findings on gender-based growth trajectories in ELLs (Hammer et al., 2009; Uchikoshi, 2006b). We

also know that summer vacation has a contrasting effect on the growth of ELLs' receptive language skills during the summer only, which results in trajectories of language growth that are discontinuous in Spanish and English (Hammer et al., 2008a,b). The focus on group-based growth rates has resulted in a lack of focus on: the individual differences in the language growth of ELLs indicated by intra- and inter-individual variability; the impact that initial status may have on growth trajectories; or the dynamic co-development of interconnected subsystems from the perspective of language growth (but see Conboy et al., 2006, regarding the early language growth of simultaneous bilinguals).

The purpose of this study was to advance our knowledge of language growth in ELLs by estimating the initial status and the rates of growth of MLUw, NDW, and WPM in Spanish and English during the first 3 years of formal schooling, and determining the predictive roles of time, gender, and summer vacation on language growth. The findings of this study were contextualized from the perspective of language as a dynamic system, considering the four principal elements of DST.

The findings of this study advance our understanding of oral language development in Spanish and English of ELLs by focusing on the overall trajectory of oral language skills during the first 3 years of formal schooling. This focus on language growth can identify whether there are individual differences in the language growth rates of ELLs and to what extent initial levels of language predict growth over time in specific oral language skills. In addition, examining the shapes of growth of each language can

begin to answer how the growth in different oral language skills may be associated, and if the shape of growth is different across the languages.

Four research questions guided this longitudinal study of ELLs' oral language growth:

First, do measures associated with different subsystems (MLUw and NDW) and the global integration of subsystems (WPM) demonstrate change (growth) over time? If so, what is the shape of this growth in each language, and what predictive roles (if any) are played by time, gender, and/or schooling on language growth? Based on the existing literature on the language growth of ELLs, the linguistic subsystems and their global integration are hypothesized to demonstrate positive growth, and the shapes of growth are hypothesized to be similar within each language (Spanish and English). Based on prior findings, time, gender, and schooling are all expected to play a predictive role in the growth of Spanish and English.

Second, are there significant individual differences in the initial status of growth and in the growth rates for MLUw, NDW, and WPM in Spanish and English? If so, are these individual differences detected at the intra- and inter-individual levels? Although prior findings on the individual differences of ELLs' language growth are limited, varying degrees of inter- and intra-individual variance across measures and across languages are hypothesized.

Third, does the initial status of growth have an impact on the growth trajectories of MLUw, NDW, and WPM in Spanish and English? If so, what is the direction and significance of the impact of initial status on growth? The intercept-slope covariance data are largely absent from contemporary work on language growth, except for Rice et al. (2006, 2009), who found non-significant intercept-slope covariances in the morphosyntactic skills of TD and SLI children. Therefore, MLUw in Spanish and English are hypothesized to demonstrate non-significant intercept-slope covariances. By extension, NDW and WPM are hypothesized to also demonstrate non-significant intercept-slope covariances in both languages.

Fourth, does the growth of the different linguistic subsystems and their integration demonstrate dynamic co-development over time? Is so, is this co-development restricted across-subsystems and within each language, or is it also evident within-subsystems and across languages? Based on prior findings, the interconnected linguistic subsystems and their integration are hypothesized to demonstrate across-subsystem and within-language co-development.

CHAPTER 2

METHODOLOGY

The present study, which longitudinally profiled the language growth of ELLs' oral language skills in Spanish and English, utilized secondary data from a larger project. Prior to discussing the secondary data used in the present study, relevant aspects of the larger project are outlined first: the overall design of the larger project, the narrative language sampling protocol, and the longitudinal data sampling procedure. This is followed by a description of the language samples produced by participants selected for the present study, the outcome measures and covariates examined, and the analytic approach.

Parent Study and Bilingual Language Literacy Project (BLLP)

Data collected as part of the Biological and Behavioral Variation in the Language Development of Spanish-Speaking Children (BVLDSC) study [R305U01001] (Francis et al., 2005) were used in the present study. The BVSLDSC was a large-scale, multi-level study designed to investigate factors (e.g., community context, school context, academic instruction, intervention) that might influence the development of literacy and language skills of ELLs whose first language was Spanish. The BVLDSC consisted of two phases: an assessment tool development phase and a longitudinal data collection and analysis phase. The first phase focused on developing, refining, and evaluating a wide range of assessment tools that would be used in the second phase. The second phase used the

assessment tools developed in phase 1 of the project and examined, from a longitudinal perspective, the literacy and language skills of ELL children as they transitioned from kindergarten to second grade.

The participants, who were enrolled in one of 40 schools and 93 classrooms in California (Long Beach area) and Texas (Austin, Houston, and Brownsville), all met the following criteria:

- Designated by the school district as ELL.
- Attended a school in which at least 40 percent of the school population was Latino, representative of typical educational environments where the majority of Spanish-speaking ELLs are enrolled in (Swanson, 2009).
- Attended a school in which at least 30 percent of the kindergarten population was identified as Limited English Proficient, as designated by district-based, language proficiency testing.
- Enrolled in one of four language of instruction models/programs[1]: Early transitional bilingual; Late transitional bilingual, Dual-language; English Immersion.

---

[1] Language of instruction model/program refers to the language(s) of academic instruction. Early transitional refers to 'early exit' programs where academic instruction in kindergarten is primarily in Spanish, with increasing instruction in English from first grade on with the goal of transitioning to English-only instruction after 3rd grade. Late transitional refers to 'late exit' programs, where Spanish instruction is provided throughout, with increasing instruction in English with the goal of transitioning to English-only instruction after 3rd grade. Dual language refers to language programs where academic instruction is 50% in Spanish and 50% in English across grades; also known as '2-way' instruction. English immersion refers to instruction in English only from the onset (Swanson, 2009).

One component of the large-scale study, referred to as the Bilingual Language Literacy Project (BLLP), focused on the oral narrative skills of the participants. This component of the project collected longitudinal data (the source of the data used for the present study) on the Spanish and English narrative skills of all participants. What follows is a description of the BLLP longitudinal data sampling protocol.

BLLP Narrative Language Sampling Protocol

All participants were asked to retell, in Spanish and English, one of four Frog Stories developed by Mercer Mayer. Mercer Mayer authored a series of wordless picture storybooks that involved the different adventures of a frog along with other characters. The wordless picture storybooks, commonly known as *Frog Stories*, have been used in a number of cross-linguistic studies and have been shown to be non-biased and valid for eliciting detailed narrative language samples from participants of various linguistic and cultural backgrounds (cf. Berman & Slobin, 1994; Özçalişkan & Slobin, 1999; Papafragou, Massey, & Gleitman, 2002; Pavlenko, 2009), including Spanish-speakers (Sebastián & Slobin, 1994; Heilmann, Miller, & Nockerts, 2010; Simon-Cerejido & Gutiérrez-Clellen, 2009). The following *Frog Stories* were used in the BLLP longitudinal dataset: *Frog Where Are You?* (Mayer, 1969); *Frog Goes to Dinner* (Mayer, 1974); *Frog On His Own* (Mayer, 1975b); and *One Frog Too Many* (Mayer, 1975a).

The narrative language elicitation protocol used, similar to the one developed by Strong (1998), required the examiner to sit across from the participant, and provide the child a pre-scripted narrative of the story in the target language (Spanish or English),

while the child viewed the book. After narrating the story, the examiner gave the book to the child and asked her/him to retell the story (e.g., "Ahora, cuéntame lo que pasó en este cuento"/"Now, tell me what happened in the story"). As recommended by Strong (1988), the participants used the pictures in the book and the examiner model in order to facilitate the narrative retell task. Examiner feedback was limited to backchannel responses (e.g., "aha"/"si"/"yes") with rising inflection or the repetition of the participant's preceding last utterance in order to encourage the participant to continue narrating. Examiners were instructed not to provide any additional information or answer any questions from the participants. To promote task familiarity in the native language of the children, all initial sessions were administered in Spanish. The data collection procedure was replicated in English approximately a week after the first session. All examiners were proficient Spanish-English speakers.

All narrative language samples were recorded onto digital media and transcribed orthographically following the conventions for bilingual language samples (Heilmann, et al., 2008; Rojas & Iglesias, 2009; 2006). Modified Communication Units (MC-units), originally proposed by Guitérrez-Clellen and Hofstetter (1994) for Terminable Units in Spanish, were used to segment narratives in Spanish to accommodate for the PRO-Drop nature of Spanish. Utterances containing coordinating conjunctions with co-referential subject deletion were segmented as two utterances, but maintained as one utterance when conjoined with a subordinating conjunction. In Spanish, the omission of subject nouns and pronouns is grammatically acceptable since verbs in Spanish are coded for

number/person and gender (Bedore, 1999, 2001). Although English is not a PRO-Drop language, narratives in English were also segmented using the MC-unit segmentation method to maintain segmentation consistency across Spanish and English measures.

Transcription reliability for protocol accuracy, transcription accuracy, and coding reliability on the original BLLP longitudinal database were conducted on 20 English and 20 Spanish randomly selected samples. Protocol accuracy measures ranged from 98 to 100 percent in English and 94 to 99 percent in Spanish, and transcription accuracy values on orthographic transcriptions ranged from 90 to 98 percent in English and 91 to 99 percent in Spanish (Miller et al., 2006).

BLLP Longitudinal Sampling

Narrative language sample data were collected longitudinally over a 3-year period, involving six discrete waves of data collection (i.e., from the fall of kindergarten to the spring of second grade). Table 1 delineates the data collection points and the corresponding *Frog Story* elicited. The month of each collection period per semester was time-unstructured (October during the fall semester; May during the Spring semester) across the 3 years of data collection and ensured that participants were exposed for at least 1 month to their respective language of instruction program placement.

Oversampling was implemented throughout the longitudinal phase of the project in order to take into account participant attrition. After the first two waves, participant

Table 1. BLLP Longitudinal Data Sampling Procedure

Data Collection Wave	Academic Year	Academic Semester / Month	Mercer Mayer Frog Story
Wave 1: Fall	Kindergarten	Fall / October	FWAY[a]
Wave 2: Spring		Spring / May	FGTD[b]
Wave 3: Fall	First grade	Fall / October	FOHO[c]
Wave 4: Spring		Spring / May	OFTM[d]
Wave 5: Fall	Second grade	Fall /October	FWAY
Wave 6: Spring		Spring /May	FGTD

[a]FWAY: *Frog Where Are You?* (Mayer, 1969). [b]FGTD: *Frog Goes to Dinner* (Mayer, 1974). [c]FOHO: *Frog On His Own* (Mayer, 1975b). [d]OFTM: *One Frog Too Many* (Mayer, 1975a).

replacement was achieved by targeting every child in every classroom rather than a one-for-one replacement. Thus, the number of children in each target classroom determined the number of children added throughout the duration of the study. The attrition and replacement of participants resulted in 48 different patterns of missing data (Swasey Washington, 2009); the most frequently occurring pattern of missing data was that of absent cases across Waves 1 and 2 (fall and spring of kindergarten) of data collection.

It is important to note that several participants were missing gender information; some participants failed one (or more) academic grade(s) during the data collection period; some language samples were as short as 1-to-2 utterances in length; and some language samples were produced primarily in the non-target language as measured by

number of different words (NDWs) (e.g., a "Spanish" language sample composed of 60% NDWs in English). Each one of these characteristics of the BLLP longitudinal dataset can be potentially problematic for research purposes. For example, gender-based differences cannot be appropriately identified if gender information is missing; participants that failed at least one academic grade will result in participants that are atypically older than their grade-matched peers; language samples of extremely limited length can artificially skew the overall distribution of language-based data, assuming that the quantity of these samples does not represent the majority in the dataset; language-specific data analyses and conclusions from ELLs can be contaminated by data in the non-target language. What follows is a description of the narrative language sampling protocol used to collect BLLP longitudinal data.

Secondary Data Used for the Present Study

The complete BLLP longitudinal dataset consisted of 13,812 oral narrative language samples that were produced by 2,051 ELL children across a 3-year period (Francis et al., 2006; Miller et al., 2006). To rule-out the previously discussed problematic characteristics of the BLLP longitudinal dataset, the following inclusionary criteria for participants and narrative language samples were used: (1) gender information had to be available for each participant; (2) participants had not been retained in any academic grade; (3) participants had to be within the expected age range during the fall semester of each grade (4;9 to 7;1 at fall of kindergarten; 5;9 to 8;1 at fall of first grade; and 6;9 to 9;1 at fall of second grade); (4) narrative language samples had to be at least

five or more complete and intelligible utterances in length; and (5) 75 percent or more of the total number of different words (i.e., uninflected word roots) in a transcript had to be produced in the target language. A total of 328 participants were excluded from the study's longitudinal dataset: 226 participants had missing gender information; 94 participants failed one or more academic grades; 2 participants exceeded the age range at fall of kindergarten; 5 participants exceeded the age range at fall of first grade; 1 participant exceeded the age range at fall of second grade. These 328 excluded participants translated into a total of 1,492 excluded narrative language samples. An additional 72 narrative language samples were also excluded from the study's longitudinal dataset because they did not meet the narrative length and target language criteria. Specifically, 37 samples had less than five complete and intelligible utterances, and 35 samples had less than 75% NDWs in the target language.

The final longitudinal dataset used in the present study consisted of 12,248 oral narrative language samples (6,516 Spanish; 5,732 English) that were produced by 1,723 ELLs. The mean age for the participants was 5;7 years (*SD* = 0;4) at initial status (i.e., fall of kindergarten). The total gender distribution across the longitudinal dataset was 50.3 percent female (*n* = 867) and 49.7 percent male (*n* = 856). The study's language sample and chronological age descriptive statistics are provided in Table 2, which indicates the total number of language samples produced and mean chronological age at each data collection wave (i.e., semester).

Table 2. Language Sample and Chronological Age Descriptive Statistics per Semester

	K-Fall	K-Spring	1st-Fall	1st-Spring	2nd-Fall	2nd-Spring
	N	*N*	*N*	*N*	*N*	*N*
Total Samples	1,524	1,824	2,236	2,420	2,182	2,062
Spanish Narrative Retell Samples (Proportion per Semester)	937 (61%)	1,011 (55%)	1,198 (54%)	1,247 (52%)	1,100 (50%)	1,023 (50%)
English Narrative Retell Samples (Proportion per Semester)	587 (39%)	813 (45%)	1,038 (46%)	1,173 (48%)	1,082 (50%)	1,039 (50%)
Children that Retold Narratives in Both Languages (Proportion per Semester)	486 (32%)	712 (39%)	943 (42%)	1,104 (46%)	1,008 (46%)	979 (48%)
Mean Chronological Age (*SD*)	5;7 (0;4)	6;2 (0;4)	6;7 (0;4)	7;2 (0;4)	7;7 (0;4)	8;2 (0;4)

Table 2 demonstrates a number of general patterns. The number of total samples produced increased from the fall of kindergarten (1,524) to the spring of first grade (2,420), with decreases in the fall (2,182) and the spring (2,062) of second grade. Consistent with this pattern, the number of narrative samples in Spanish also increased from the fall of kindergarten (937) to the spring of first grade (1,247), with decreases in the fall (1,100) and the spring (1,023) of second grade. Similarly, the number of narrative samples in English also increased from the fall of kindergarten (587) to the spring of first grade (1,173), with decreases in the fall (1,082) and the spring (1,039) of second grade. These general patterns were consistent with the number of children that retold narratives in Spanish and English each semester, which increased from the fall of kindergarten (486) to the spring of first grade (1,104), and decreased in the fall (1,008) and the spring (979) of second grade. Table 2 also demonstrates that the proportion of narrative samples in Spanish gradually decreased from the fall of kindergarten (61%) to the spring of second grade (50%), while the proportion of narrative samples in English gradually increased from the fall of kindergarten (39%) to the spring of second grade (50%). The proportion of children that retold narratives in both languages each semester explains these opposite trends, where the participants became proportionately more balanced in their ability to produce language samples in Spanish and English from the fall of kindergarten (32%) to the spring of second grade (48%).

Outcome Measures: Expressive Language Skills

This study explored a number of questions regarding the longitudinal profiling of bilingual language growth, which were determined by a combination of dependent and independent variables. Measures of oral language skills in Spanish and English generated from the narrative language sample transcripts using the SALT Research 2010 software (Miller & Iglesias, 2010), were the outcome measures in this study. The following are the dependent variables (i.e., outcome measures) selected to quantify various aspects of expressive language skills: (1) mean length of utterance in words in Spanish and English (MLUw-Spanish; MLUw-English); (2) number of different words in Spanish and English (NDW-Spanish; NDW-English); and (3) words per minute in Spanish and English (WPM-Spanish; WPM-English). MLUw, NDW, and WPM have consistently been shown to be developmentally and language sensitive measures (Heilmann, Miller, & Nockerts, 2010), and they are positively related to bilingual reading achievement (Miller et al., 2006).

Mean length of utterance in words (MLUw), which is preferred over MLUm in cross-linguistic and bilingual research as it controls for varying degrees of inflection across languages (Gutiérrez-Clellen et al., 2000; Iglesias, 2004), was used as a general measure of the morphosyntactic subsystem. MLUw relative to other morphosyntax measures (i.e., Developmental Assessment of Spanish Grammar, Toronto, 1976); MLUm; mean length of terminable units; mean length of response in words) has been found to be particularly useful for assessing Spanish morphosyntax (Bedore & Peña,

2008; Castilla, Restrepo, & Perez-Leroux, 2009; Gutiérrez-Clellen, et al., 2000). In addition, given that this study investigated language growth in typologically distinct languages (i.e., Spanish is a highly inflected language relative to English, Bedore, 1999, 2001; Iglesias, 2004), MLUw was selected for this study to maintain cross-language consistency. MLUw was calculated by summing the total number of intelligible words produced in a Spanish language sample, and dividing that total by the number of complete and intelligible utterances.

Number of different words (NDW), the total number of different uninflected word roots, was used to estimate the diversity of participants' vocabularies (Golberg, Paradis, & Crago, 2008; Paul, 2006) and to represent the lexical subsystem in this study. NDW, which is significantly correlated (r = .71) with age (Miller, 1987), has been shown to be a developmentally sensitive measure of narrative productivity for Spanish-English bilingual children (Uccelli & Páez, 2007), and has also been positively associated with MLUw in bilingual speakers (Bedore et al., 2006; Kohnert et al., 2010; Simon-Cereijido et al., 2009). NDW was calculated by adding the total number of different uninflected word roots in each target language and excluding the number of uninflected word roots in the non-target language.

Words per minute (WPM) was used as a measure of verbal fluency in order to represent the global integration of multiple linguistic subsystems (Miller et al., 2006) in this study. WPM, which is frequently used in research based on language samples (e.g., Heilmann et al., 2008; Price et al., 2010; Tilstra & McMaster, 2007), is positively

correlated with the age of ELL children (Miller & Heilmann, 2004). WPM was calculated by dividing the total number of words by the duration of the language sample in minutes and seconds. Together, MLUw, NDW, and WPM, provided a composite image of specific linguistic subsystems and global subsystem integration in ELLs.

Covariates: Academic Semester, Schooling, and Gender

The three independent variables (covariates) were academic semester, schooling, and gender. It is important to note that the language of instruction program placement was not considered in this study. Adhering to the theoretical DST-framework and the temporal requirements of longitudinal design, the central covariate was time, measured in academic semesters. One of the principal research aims of this study was to identify and model the intra-individual and inter-individual variation in initial status and in growth rates over time in specific oral language skills. Given that time is constant, it is not dependent on any of the other variables to be considered. Consideration of the degree of variance in oral language skills that can be explained by time, then, is paramount to this study. Therefore, growth over time (measured in academic semesters) was modeled by using the academic semester as the primary time-varying covariate.

Another principal research aim of this study was to examine whether schooling or gender optimized model fit and/or predicted growth over time. Schooling, a time-varying covariate, modeled discontinuous growth over time in order to account for the potential effect of having or not having had schooling in the preceding semester. In other words, not having schooling during each summer preceding each fall semester, and having

schooling preceding each spring semester could potentially result in discontinuities of growth. Gender, a time-invariant covariate, was included to identify potential gender-based differences in initial status and growth rates from kindergarten to second grade. What follows is a description of the analytic approach used to profile the language growth of ELLs' oral language skills in Spanish and English.

Analytic Approach

The purpose of this study was to profile the longitudinal growth of ELLs' oral language skills in Spanish and English over time. To achieve this purpose, growth curve model testing was conducted. Model testing involved the construction of a series of growth curve models (GCMs); calculation of the within- and between-person variances accounted for by each model; and the comparison of goodness-of-fit indices across models to determine the best fitting model for each outcome measure in Spanish and English (Luke, 2004; Singer & Willett, 2003). The final GCM for each measure accounted for the most variance (i.e., highest Pseudo-R^2s, representative of the highest proportional variance reductions) and/or achieved the best fit (i.e., lowest goodness-of-fit indices) to explain the respective longitudinal data. The final GCM was used to generate the most representative growth curve trajectory (prototypical growth curve trajectory) for each outcome measure in each language.

The multi-step process of growth curve model testing began with three preliminary steps: establishing a time metric, incorporating a time-centering parameter, and conducting exploratory analyses. These three preliminary steps are discussed first.

Subsequently, the five-step sequence of model testing is discussed in the following order: (1) goodness-of-fit indices, (2) unconditional means model, (3) unconditional growth models, (4) conditional growth models, and (5) the prototypical growth trajectory for the final model (Cromley, 2008; Singer & Willett, 2003). This multi-step model testing process was individually conducted for each of the three outcome measures (MLU, NDW, and WPM) in Spanish and English.

Time Metric and Time-Centering

The goal within GCM is to select the most easily interpretable time metric to carry out and answer specific research questions. The present study included six waves of data collection (fall and spring semesters from kindergarten to second grade) and multiple outcome measures (MLU, NDW, and WPM) in Spanish and English with continuous values that changed over time. Potential time metric candidates for this study were: children's chronological age, months, and academic semester. Children's chronological age was not the most sensible time metric because the longitudinal dataset consisted of >1,000 participants ranging in age from 4;9 to 7;1 (SD = 0;4) at initial status. Therefore, the temporal stability of children's chronological age was neither optimal (i.e., each participant would have their own temporal reference point for each wave of measurement) nor the most easily interpretable. Months was a more sensible metric of time, as most participants were tested around October and May from kindergarten to second grade. This time-unstructured testing schedule provided greater temporal stability than chronological age since it was unaffected by the addition of new participants over

time. However, the utility of interpreting outcome measures relative to "October" and "May" was limited. The most sensible time metric for this study was the academic semester. Semester maintained the temporal relationship across the six waves of measurement (e.g., fall of kindergarten, spring of kindergarten, fall of first grade, etc.) used in this study, it was easily interpretable, it is commonly used in studies of school-age children (as opposed to age, which is more common in studies of infants, toddlers, and preschool-aged children), and it is associated with the curricular and academic demands placed on children throughout elementary school (Alexander & Martin, 2004; Crone & Whitehurst, 1999; Morrison, Griffith, & Alberts,1997; Swasey Washington, 2009). Therefore, academic semester was the default time metric for this study.

The purpose of time-centering, similar to establishing a time metric, is to facilitate interpretation of longitudinal data by anchoring all longitudinal data to a relative point in time. Time can be centered at any point within a series of longitudinal measurement points. For example, if time is centered at a mid-point wave of observation, the interpretation of growth will be prospective from that mid-point forward and retrospective from that mid-point backward. Further, time-centering was also necessary to interpret the first wave of measurement as the *initial status*. The GCM data for this study were centered on the fall of kindergarten. Thus, the starting point of language growth was operationalized as the fall of kindergarten.

Exploratory Analyses

Exploratory analyses consist of visually inspecting data plots to better understand the range and complexity of growth patterns present in the longitudinal data. Exploratory analyses informed and guided the process of model testing by providing visual representation summaries of the different initial statuses and directions of growth (Singer & Willet, 2003). Exploratory analyses conducted consisted of linear regression fitted growth plots, conducted for each outcome measure in each language. The linear regression fitted growth plots for each outcome measure in Spanish and English generated a fitted linear regression line (with a unique intercept and slope) for each participant in the sample, the residual variation for each participant at each wave of measurement, and a visual representation of the overall direction of growth.

Although final GCMs cannot be confirmed with linear regression fitted growth plots, they informed GCM construction by suggesting that GCMs with additional covariates (time-varying and/or time-invariant) might be necessary to account for wide ranges of residual variation. Further, the sample plots suggested modeling of linear and curvilinear growth across its six waves of measurement. Curvilinear growth was modeled using quadratic and cubic polynomial functions, incorporated as time-varying covariates. In addition, discontinuous growth (linear growth with expected discontinuities) was modeled using schooling as a time-varying covariate.

Growth Curve Model Testing

GCM model testing[2] is a multi-step process that involves comparing goodness-of-fit of indices, and the proportion of variance reduction calculated via Pseudo-R^2 statistics. The goodness-of-fit indices are discussed first, followed by a description of the models tested and their corresponding Psuedo- R^2 statistics.

Goodness-of-Fit Indices

Goodness-of-fit indices provided a statistical assessment of the degree of fit that each GCM had to a set of longitudinal data. Model fit was used as a quantitative measure of the degree of correspondence that each GCM had to the sample data (i.e., how well each GCM explained the data). The best fitting GCM (i.e., the final model) had the smallest goodness-of-fit index, relative to all other GCMs. Two different fit indices were estimated and compared across specific sequences of GCMs: the -2 Log-Likelihood (commonly referred to as -2LL or Deviance) was used to compare model fit and the mean change in the deviance test across nested GCMs; Schwarz's Bayesian Criterion (BIC) was used to compare model fit across non-nested GCMs (Nakamoto, Lindsey, & Manis, 2007; Singer & Willet, 2003).

The -2LL deviance statistic compared the difference between the log-likelihood statistics (i.e., maximum likelihood estimates) for each respective GCM and, "a *saturated*

---

[2] For reference purposes, Appendix A contains a glossary of Greek terms used in GCM, with their corresponding English definitions; Appendices B-D outline the formulas for the unconditional means, unconditional growth, and conditional growth models, with descriptions of the fixed effects and variance components.

model, a more general model that fits the sample data perfectly" (Singer & Willett, 2003, p. 117). The smallest -2LL deviance statistic indicated the *best fitting nested* model, taking into account fixed effects (γ's) *and* variance components (σ's). GCMs that were nested within other GCMs shared identical parameters, with simpler GCMs considered to be reduced model versions of the full or final model. -2LL was used in the present study to compare the fit and the mean change in the deviance test of nested models of growth and to also compare the fit and the mean change in the deviance test of nested models of discontinuous growth.

The BIC fit index was used to compare model fit across *non-nested GCMs*. Instead of being nested within one another, non-nested GCMs differ with regard to specific parameters and/or predictors that are used. BIC, like -2LL is also based on log-likelihood statistic differences between non-nested GCMs and a saturated model. However, BIC "penalizes" the log-likelihood statistic of each respective GCM by decreasing it. The BIC penalty is based on the number of parameters (*df*) in each GCM and sample size. BIC was used in the present study to compare the fit of CG-models of continuous growth with CG-models of discontinuous growth.

Unconditional Means (UM) Model

The first in a sequence of GCMs tested was the unconditional means (UM) model, which was a GCM *without time*, and was used primarily for screening the variability to be explained in an outcome measure. The purpose of the UM-model, which was an intercept-only model, was to determine the significance and degree of variation observed

between participants for a particular outcome measure without any predictors, including the proportion of residual variability to be explained. This was achieved by testing the residual variability of the UM-model for significance, and by calculating the intraclass correlation coefficient (*ICC*). The ICC determines the proportion of variation accounted for by differences between participants; the relative magnitude of the within-person (Level-1) and the between-person (Level-2) residual variance are calculated without considering the effect of time (Singer & Willett, 2003).

Unconditional Growth (UG) Models

The next series of GCMs tested were the unconditional growth (UG) models. UG-models were GCMs *with time* as a time-varying covariate (no other predictors), and like the unconditional means (UM) model they were also used primarily for screening purposes. The purpose of testing UG-models was to determine whether additional variance in a particular outcome measure could be explained by adding time as a time-varying covariate. If no additional variance could be accounted for by the addition of time as a covariate, then continuing the model testing process of GCM was not warranted. However, if additional variance could be accounted for by the addition of time, then GCM was warranted as it indicated that the outcome measure changed over time.

Determining whether UG-models were an improvement relative to the UM-model in explaining variance was achieved by calculating the Pseudo-R_{ε}^2 statistic (proportional variance reduction of each UG-model, relative to the UM-model, of Level-1 within-

person residual variance (σ_ε^2) over time), and by determining the goodness-of-fit (smallest -2LL) of each UG-model (Singer & Willett, 2003).

Several UG-models were constructed for each language measure in Spanish and English to attempt modeling distinct shapes of growth trajectories. The most simple UG-model incorporated linear time to model linear growth trajectories. Additional time-varying covariates, including quadratic (γ_{20}) and cubic (γ_{30}) polynomial functions of time, were incorporated across different UG-models to model curvilinear growth trajectories. Linear, quadratic, and cubic UG-models were nested models as the UG-Linear model was nested within the UG-Quadratic model, and the UG-Quadratic model was nested within the UG-Cubic model. The Pseudo-R_ε^2 values of these models were compared and contrasted to determine what UG-model represented the greatest improvement in explaining Level-1 within-person variance (i.e., highest proportional variance reduction) relative to the UM-model (Singer & Willett, 2003).

Apart from potentially modeling time as linear, quadratic, and/or cubic time, UG-models also varied with regard to whether the slopes in each UG-model were set to be fixed or randomly varying. Fixing slopes assumed that although participants could significantly vary with respect to initial status (i.e., intercepts), the entire sample shared the same growth rate. Randomly varying slopes allowed participants to vary in their growth rates over time. In a UG-Linear model, only the linear slope was tested as fixed or randomly varying. In the nested UG-Quadratic model, the linear and/or the quadratic slope were tested as fixed or randomly varying. Similarly, in the nested UG-Cubic

model, the slopes for linear, quadratic, and cubic time were tested in different fixed vs. random combinations to determine the best fitting UG-Cubic model (Cromley, 2008; Singer & Willett, 2003). Randomly varying slopes generated additional variance components, which further partitioned residual variance (i.e., linear slope variance; quadratic slope variance; cubic slope variance). UG-models with randomly varying slopes represented increasingly heterogeneous growth patterns. If any UG-model failed to converge (the sample data was highly unbalanced or there was an overabundance of missing data), the slopes were set as fixed (e.g., Nakamoto, Lindsey, & Manis, 2007; Uchikoshi, 2006b). The present study tested all possible combinations of UG-models with fixed vs. random slopes that were supported by the sample data. The best fitting UG-model (smallest -2LL) served as the baseline conditional growth model for each respective outcome variable in each language.

Conditional Growth (CG) Models

The purpose of testing conditional growth (CG) models was to determine whether additional variation in a particular outcome measure could be accounted for by *adding* time-invariant covariates and/or time-varying covariates. In the present study gender was incorporated as a time-invariant covariate to identify the potential effect of gender-based differences, and schooling was incorporated as a time-varying covariate to model discontinuous growth. Consequently, the present study tested CG-models that added at least one time-invariant covariate at Level-2 (gender) and one time-varying covariate (schooling) at Level-1 that could serve as predictors of intercepts and slopes.

Determining whether CG-models explained more variance, relative to the unconditional means (UM) and the unconditional growth (UG) models, was achieved by calculating a minimum of three Pseudo-R^2 statistics, and by determining the goodness-of-fit (smallest -2LL) of each CG-model.

- **Pseudo-R_{ε}^2** = proportional variance reduction of the CG-model, relative to the UM-model, of Level-1 within-person residual variance (ε_{ij} [σ_{ε}^2]) over time.

- **Pseudo-R_0^2** = proportional variance reduction of the CG-model, relative to its nested (i.e., parallel Level-1 terms) UG-model, of Level-2 between-person residual variance (ζ_{0i} [σ_0^2]) in intercepts.

- **Pseudo-R_1^2** = proportional variance reduction of the CG-model, relative to its nested UG-model, of Level-2 between-person residual variance (ζ_{1i} [σ_1^2]) in linear slopes.

Two sets of CG-models were tested in order to determine whether additional proportional variance reduction, relative to the best fitting UG-models, could be accounted for by adding time-invariant covariates and/or time-varying covariates. The first set was a series of CG-models with Gender added as a time-invariant covariate at Level-2. The second set was a series of discontinuous CG-models with schooling added as a time-varying covariate at Level-1 and Gender added at Level-2.

CG-Models with Gender. The first set of CG-models tested were CG-models with Gender added as a time-invariant covariate. Based on the UG-model testing results,

the baseline CG-model was tested to determine the effect of Gender on initial status (i.e., intercept). Further, Gender was tested with every possible interaction of Gender *x* Slope to determine the effect of Gender on the linear, quadratic, and cubic slopes.

Discontinuous CG-Models with Schooling and Gender. The second set of CG-models tested were discontinuous CG-models with schooling added as a time-varying covariate at Level-1 and Gender added as a time-invariant covariate at Level-2. Discontinuous CG-models model hypothesized discontinuities of growth at expected points in time. Schooling was treated as the effect of having or not having schooling in the preceding semester.

To isolate the effect of Schooling, a series of discontinuous CG-models with schooling *only* were tested with every possible combination of fixed and randomly varying slopes to determine the best fitting discontinuous CG-model first. The effect of Gender on initial status and growth rates was then added to the best fitting Discontinuous CG-model. Testing a series of discontinuous CG-models determined whether additional proportional variance reduction, relative to the UM-model, could be accounted for by adding Linear Time and Schooling as time-varying covariates at Level-1. Only discontinuous linear growth was modeled, since quadratic and cubic models assume continuous growth.

The effect of Gender was then added to the best fitting discontinuous CG-model to determine whether additional proportional variance reduction (relative to the discontinuous CG-model without Gender) could be accounted for by adding Gender as a

time-invariant covariate. Gender was tested with every possible interaction of Gender *x* Slope to determine the effect of Gender on the linear and schooling slopes.

Final CG-Model. The final CG-model determined the additional proportional variance reduction in a particular outcome measure that could be accounted for by *adding* time-invariant covariates and/or time-varying covariates, which was determined by Pseudo-R^2 statistics and goodness-of-fit index comparisons. The fit of the best fitting CG-model and the best fitting Discontinuous CG-model, which were non-nested models, was evaluated using BIC.

Prototypical Growth Trajectory

The best fitting CG-model for each outcome measure was the final model for that particular outcome measure. The final model explained the most variance (i.e., highest overall proportional variance reduction) and/or fits its respective longitudinal data the best as indicated by goodness-of-fit indices (i.e., lowest -2LL for nested models; lowest BIC for non-nested models). A prototypical growth trajectory was generated from the parameter estimates of the final GCM, which allowed for direct interpretation of GCM data. The prototypical growth trajectory predicted a pattern of growth that was representative of the individual growth patterns demonstrated by participants within 1 *SD* of the mean.

Summary

The present study incorporated growth curve modeling to profile the longitudinal growth of ELLs' oral language skills in Spanish and English over time. A series of growth curve models (GCMs) of continuous and discontinuous growth were constructed for each outcome variable (MLUw, NDW, and WPM) in each language. Academic semesters was the default time metric and waves of measurement were centered relative to the fall of kindergarten as initial status. Following exploratory analyses, the subsequent series of GCMs were tested for each outcome variable: (1) UM-model, (2) UG-models (including all potential variations of fixed vs. randomly varying slopes) to determine the baseline UG-model for CG-model analyses, (3) CG-models (including all potential interactions of Gender x Time, Gender x Grade, and Gender x Schooling) to determine the final model, and (4) generation of prototypical growth curve trajectory from the final model parameter estimates. The proportion of variance reduction (Pseudo-R^2 statistics) and goodness-of-fit indices (-2LL for nested models; BIC for non-nested models) determined the baseline UG-model and the final CG-model for each outcome measure.

CHAPTER 3

RESULTS

This study examined the dynamic growth of Spanish and English in ELL children by estimating the initial status and the rates of growth of MLUw, NDW, and WPM in Spanish and English during the first 3 years of formal schooling, and determining the predictive roles of time, gender, and summer vacation on language growth. This focus on language growth identified whether there were individual differences in the language growth rates of ELLs and to what extent initial levels of language predicted growth over time in specific oral language skills. In addition, examining the shapes of growth of each language can begin to answer how the growth in different oral language skills may be associated, and if the shape of growth is different across the languages.

Descriptive analyses, means and standard deviations, were calculated for each of the outcome measures across each wave (academic semester) of measurement. Growth curve modeling (GCM) analyses were then conducted on the outcome measures using maximum likelihood estimation to estimate within- and between-person variation in initial status and growth rates. Statistical analyses were calculated using IBM SPSS Statistics 18.0 for Mac (SPSS Inc., 2010), with the exception of prototypical growth trajectories, which were generated using Microsoft Excel 2008 for Mac, Version 12.2.4 (Microsoft Corporation, 2007).

Descriptive Statistics

Descriptive statistics included means and standard deviations for each outcome measure during each wave of measurement. The means and standard deviations (per semester) for the study's outcome measures are provided in Table 3. As can be seen in Table 3, all of the outcome measures showed an overall increase, in both languages, from the fall of kindergarten to the spring of second grade. In Spanish, MLUw increased from 4.69 to 7.06 (a mean increase of 2.37); NDW increased from 55.26 to 91.78 (a mean increase of 36.52); and WPM increased from 57.60 to 82.48 (a mean increase of 24.88). Similarly, in English, MLUw increased from 5.55 to 7.97 (a mean increase of 2.42); NDW increased from 53.98 to 99.95 (a mean increase of 45.97); and WPM increased from 68.24 to 93.37 (a mean increase of 25.13).

Preliminary Steps

GCM analyses were conducted for each outcome measure in each language to determine within- and between-person variation in initial status and growth rates Preliminary steps included: establishing a time metric, time-centering, and exploratory analyses. As previously discussed, this study included six waves of data collection that occurred every semester from the fall semester of kindergarten to the spring semester of second grade. Therefore the default time metric used for interpretation for all models was the academic semester. All sample data were time-centered on the first wave of data collection, the fall semester of kindergarten. Therefore, the starting point of growth for each outcome measure was interpreted relative to the fall semester of kindergarten.

Table 3. Means and Standard Deviations of Outcome Measures per Semester

	K-Fall	K-Spring	1st-Fall	1st-Spring	2nd-Fall	2nd-Spring
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
MLUw-Spanish	4.69 (0.83)	5.68 (1.00)	5.53 (0.84)	5.96 (0.96)	5.80 (0.84)	7.06 (1.11)
NDW-Spanish	55.26 (18.66)	71.16 (20.06)	80.14 (23.01)	74.87 (18.84)	79.75 (18.35)	91.78 (19.54)
WPM-Spanish	57.60 (19.35)	66.10 (20.97)	66.80 (21.12)	65.66 (20.17)	67.18 (19.46)	82.48 (20.78)
MLUw-English	5.55 (1.18)	6.47 (1.28)	6.30 (1.12)	6.93 (1.05)	6.79 (0.82)	7.97 (1.12)
NDW-English	53.98 (21.83)	65.38 (24.27)	73.03 (27.34)	73.12 (23.33)	81.71 (22.37)	99.95 (23.14)
WPM-English	68.24 (24.06)	69.12 (25.49)	73.83 (25.73)	77.70 (23.83)	88.37 (24.00)	93.37 (22.67)

Figure 2 illustrates the exploratory analyses, linear regression fitted growth plots, conducted for each outcome measure in each language. The linear regression fitted growth plots for each outcome measure in Spanish and English generated: (1) a fitted linear regression line (with a unique intercept and slope) for each participant in the sample, (2) the residual variation for each participant at each wave of measurement, and (3) a visual representation of the overall direction of growth. Based on visual inspection, the three outcome measures for Spanish and English show: variability of intercepts and slopes, a wide range of residual variation during each wave of data collection, and *on average* positive growth over time (negative growth was also observed). Although the final GCMs for each outcome measure cannot be confirmed with the fitted linear regression fitted growth plots, they informed subsequent GCM analyses by suggesting that GCMs, with additional covariates might account for the wide range of residual variation observed for each of the three outcome measures.

Growth Curve Modeling (GCM)

Growth curve model testing of each of the three language measures MLUw, NDW, and WPM for Spanish and English included the following steps: (1) fitting the unconditional means (UM) model, (2) testing a series of unconditional growth (UG) models, (3) testing a series of conditional growth (CG) models of continuous and discontinuous time, (4) determining the final model that achieves the best fit to the sample data and the highest proportional variance reduction, and (5) using the fixed effects parameter estimates from the final model to generate prototypical growth

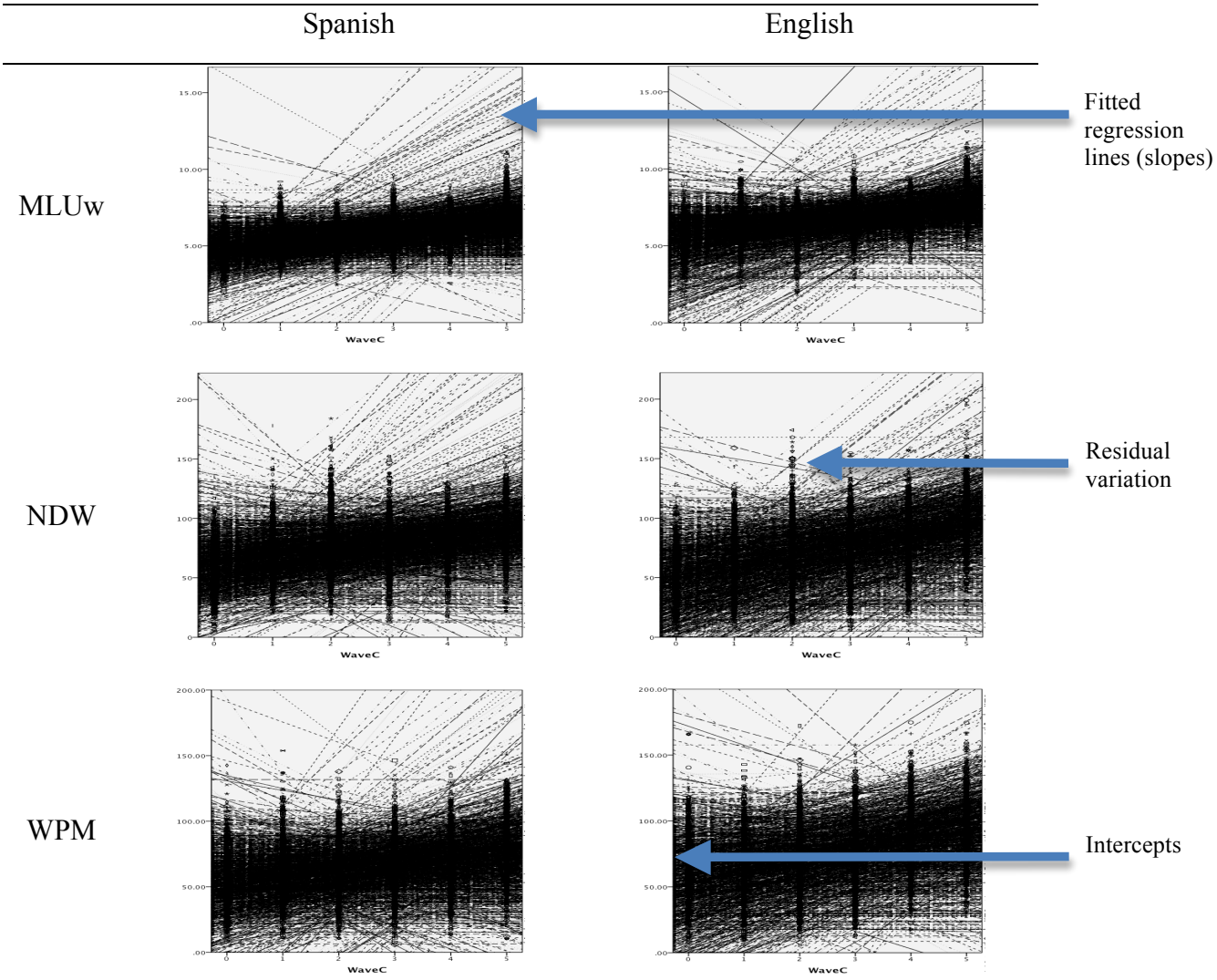


Figure 2. Linear Regression Fitted Growth Plots for Mean Length of Utterance in Words (MLUw), Number of Different Words (NDW), and Words Per Minute (WPM) in Spanish and English

trajectories. What follows are the growth curve modeling results for each of the outcome measures in Spanish, followed by the growth curve modeling results for each of the outcome measures in English.

Mean Length of Utterance in Words (MLUw) – Spanish

MLUw was used as a general measure of the morphosyntactic subsystem in Spanish. GCM analyses were conducted to determine within- and between-person variation in initial status and growth rates of MLUw-Spanish from the fall of kindergarten to the spring of second grade.

Unconditional Means (UM) Model

Fitting the UM-model determined the significance and degree of residual variation that was observed between participants for MLUw-Spanish without taking into consideration time or any predictors. The residual variability of the UM-model was tested for significance, and the intraclass correlation coefficient (*ICC*) was calculated to determine the proportion of variation accounted for by differences between participants. The UM-model for MLUw-Spanish resulted in significant within-person residual variability to be explained at Level-1 ($\sigma_\varepsilon^2 = 1.08$, $p < .05$), significant between-person residual variability to be explained at Level-2 ($\sigma_0^2 = 0.245$, $p < .05$), and an *ICC* of .18. Without considering the effect of time or any other predictors, 18% of the total variation in MLUw-Spanish was accounted for by differences between the participants in the

sample. The UM-model will be referred to as Model A when determining the final growth curve model (GCM) for MLUw-Spanish.

Unconditional Growth (UG) Model Testing

Testing a series of UG-models determined whether additional proportional variance reduction, relative to the UM-model, in MLUw-Spanish could be accounted for by adding time as a time-varying covariate at Level-1. The six waves of measurement in this study permitted modeling of linear and curvilinear continuous growth. Consequentially, apart from linear time, two polynomial functions of time were also tested: quadratic and cubic time. Each of these three nested UG-models (i.e., the UG-Linear model was nested within the UG-Quadratic model; the UG-Quadratic model was nested within the UG-Cubic model) was tested for model fit with every possible combination of fixed and randomly varying slopes. Fixed slopes required all participants to have the same growth rates; randomly varying slopes allowed participants to vary in their growth rates. The -2 Log-Likelihood (-2LL) deviance statistic was used to identify the best fitting (i.e., smallest -2LL) UG-models. Table 4 outlines the comparisons of fixed and randomly slopes for the series of UG-models tested and highlights the best fitting UG-Linear, UG-Quadratic, and UG-Cubic models for MLUw-Spanish. It should be noted that additional covariates adds parameters to be estimated, which in turn results in UG-models with increasing *df*s, as illustrated in Table 4.

Table 4 demonstrates that the best fitting UG-Linear model was the UG-Linear model with fixed linear slopes (-2LL = 17645.7), which will be referred to as Model B

Table 4. Comparisons of Fixed and Randomly Varying Slopes for Unconditional Growth (UG) Models of Mean Length of Utterance in Words-Spanish[1]

	UG-Models: Fixed and Randomly Varying Slopes					
	Linear-Random	Quadratic-Random	Cubic-Random	*df*	BIC	-2LL
UG-Linear (*UG-Ln*)	+			NC	NC	NC
	-			4	17680.8	**17645.7**
UG-Quadratic (*UG-Qd*)	+	+		NC	NC	NC
	-	+		NC	NC	NC
	+	-		NC	NC	NC
	-	**-**		5	17658.4	**17614.4**
UG-Cubic (*UG-Cb*)	+	+	+	NC	NC	NC
	-	**+**	**+**	11	16909.1	**16812.5**
	+	-	+	NC	NC	NC
	+	+	-	NC	NC	NC
	-	-	+	8	16897.3	16827.1
	-	+	-	8	16909.0	16838.8
	+	-	-	8	16924.5	16854.3
	-	-	-	6	16937.3	16884.6

[1]*Note*. **Bold** indicates best fitting models; Shading indicates non-applicable randomly varying slopes; "**-**" indicates fixed slope; "**+**" indicates randomly varying slope; NC: Non-converging model

when determining the final GCM. The UG-Linear model with randomly varying linear slopes did not converge. Similarly, the best fitting UG-Quadratic model was the UG-Quadratic model with fixed linear + quadratic slopes (-2LL = 17614.4), which will be referred to as Model C when determining the final GCM. The other three UG-Quadratic models (i.e., randomly varying linear slopes; randomly varying quadratic slopes; and randomly varying linear + quadratic slopes) did not converge. The best fitting UG-Cubic model was the UG-Cubic model with varying quadratic + cubic slopes (-2LL = 16812.5), which will be referred to as Model D when determining the final GCM.

Comparisons of the best fitting UG-Linear, UG-Quadratic, and UG-Cubic models based on -2LL revealed that the best fitting UG-model was the UG-Cubic model with randomly varying quadratic + cubic slopes (-2LL = 16812.5). This UG-model (referred to as Model D when determining the final GCM) required participants to have the same linear growth rate, but allowed participants to vary in their quadratic and cubic growth rates.

Conditional Growth (CG) Model Testing

Two sets of CG-models were tested in order to determine whether additional proportional variance reduction, relative to the best fitting UG-models, could be accounted for by adding time-invariant covariates and/or time-varying covariates. One set of models was a series of CG-models with Gender added as a time-invariant covariate at Level-2. The second set was a series of discontinuous CG-models with schooling added as a time-varying covariate at Level-1 and Gender added at Level-2.

CG-Models with Gender. The first set of CG-models to be tested were CG-models with Gender added as a time-invariant covariate. Based on the UG-model testing results, the UG-Cubic model with randomly varying quadratic + cubic slopes served as the baseline CG-model for MLUw-Spanish. The CG-Cubic + Gender model was tested to determine the effect of Gender on initial status. In addition, Gender was tested with every possible interaction of Gender *x* Slope to determine the effect of Gender on the linear, quadratic, and cubic slopes. A total of eight CG-Cubic + Gender models were tested to identify the best fitting CG-Cubic + Gender model. Table 5 outlines the comparisons of Gender *x* Slope interactions for the series of CG-Cubic + Gender models tested and highlights the best fitting CG-Cubic + Gender model for MLUw-Spanish. It should be noted that additional covariates adds parameters to be estimated, which in turn results in CG-models with increasing *df*s, as illustrated in Table 5.

Seven out of the eight CG-Cubic + Gender models converged, the exception being the CG-Cubic + Gender model, with Gender *x* Quadratic interaction. Comparisons of these seven CG-Cubic + Gender models based on -2LL revealed that the best fitting CG-model was the CG-Cubic + Gender model, with Gender *x* Linear + Gender *x* Quadratic + Gender *x* Cubic interactions (-2LL = 16757.6), which will be referred to as Model E when determining the final GCM.

Discontinuous CG-Models with Schooling and Gender. The second set of CG-models to be tested were discontinuous CG-models with schooling added as a time-varying covariate at Level-1 and Gender added as a time-invariant covariate at Level-2.

Table 5. Comparisons of Gender *x* Slope Interactions for Conditional Cubic Growth Models with Gender (CG-Cb+G) of Mean Length of Utterance in Words-Spanish[1]

	CG-Cubic + Gender models: Gender *x* Slope Interactions					
	Gender *x* Linear	Gender *x* Quadratic	Gender *x* Cubic	*df*	BIC	-2LL
CG-Cubic with Gender (*CG-Cb+G*)	+	+	+	15	16889.3	**16757.6**
	-	+	+	14	16886.1	16763.1
	+	-	+	14	16889.8	16766.9
	+	+	-	14	16891.7	16768.8
	-	-	+	13	16884.2	16770.1
	-	+	-	NC	NC	NC
	+	-	-	13	16884.7	16770.5
	-	-	-	12	16875.9	16770.5

[1]*Note*. **Bold** indicates best fitting model; "**-**"indicates absent gender by slope interaction; "+" indicates gender by slope interaction; NC: Non-converging model

Discontinuous CG-models model hypothesized discontinuities of growth at expected points in time. Discontinuous CG-models were necessary in this study to account for not having schooling during each summer preceding each fall semester, and for having schooling preceding each spring semester. Accordingly, Schooling (time-varying covariate added at Level-1) was treated as the effect of *having or not having* schooling in the preceding semester.

To isolate the effect of Schooling, a series of discontinuous CG-models with schooling *only* were tested with every possible combination of fixed and randomly varying slopes to determine the best fitting discontinuous CG-model first. The effect of Gender was then added to the best fitting discontinuous CG-model to determine whether additional proportional variance reduction (relative to the discontinuous CG-model without Gender) could be accounted for by adding Gender as a time-invariant covariate. Discontinuous CG-models modeled discontinuous linear growth only, since quadratic and cubic models assume continuous growth. Table 6 outlines the comparisons of fixed and randomly varying slopes for the series of discontinuous CG-models tested and highlights the best fitting discontinuous CG-model for MLUw-Spanish.

Table 6. Comparisons of Fixed and Randomly Varying Slopes for Conditional Discontinuous Growth (Dc-CG) Models of Mean Length of Utterance in Words-Spanish[1]

	Discontinuous CG-Models: Fixed and Randomly Varying Slopes				
	Linear-Random	Schooling-Random	*df*	BIC	-2LL
Discontinuous-CG (*Dc-CG*)	+	+	NC	NC	NC
	-	+	NC	NC	NC
	+	-	7	16995.7	**16934.2**
	-	-	5	16986.5	16942.6

[1]*Note*. **Bold** indicates best fitting model; "**-**" indicates fixed slope; "+" indicates randomly varying slope; NC: Non-converging model

Two of the four discontinuous CG-models converged. The discontinuous CG-model with fixed linear + randomly varying schooling slopes, and the discontinuous CG-model with randomly varying linear + schooling slopes did not converge. Of the two converging discontinuous models (see Table 6), the best fitting discontinuous CG-model was the discontinuous CG-model with randomly varying linear + fixed schooling slopes (-2LL = 16934.2), which will be referred to as Model F when determining the final GCM. This discontinuous CG-model allowed participants to vary in their linear growth rates, but required them to have the same schooling growth rates during the spring semesters.

The effect of Gender was then added to the best fitting discontinuous CG-model to determine whether additional proportional variance reduction (relative to the discontinuous CG-model with randomly varying linear + fixed schooling slopes) in MLUw-Spanish could be accounted for by adding Gender as a time-invariant covariate. Gender (effect of being female) was tested with every possible interaction of Gender *x* Slope to determine the effect of Gender on the linear and schooling slopes. A total of four Discontinuous CG + Gender models were tested to identify the best fitting Discontinuous CG + Gender model. Table 7 outlines the comparisons of Gender *x* Slope interactions for the series of Discontinuous CG + Gender models tested and highlights the best fitting Discontinuous CG + Gender model for MLUw-Spanish.

All Discontinuous CG + Gender models converged. Comparisons of the four Discontinuous CG + Gender models based on -2LL revealed that the best fitting

Table 7. Comparisons of Gender *x* Slope Interactions for Conditional Discontinuous Growth Models with Gender (Dc-CG+G) of Mean Length of Utterance in Words-Spanish[1]

	Discontinuous CG-Models: Gender *x* Slope Interactions				
	Gender *x* Linear	Gender *x* Schooling	*df*	BIC	-2LL
Discontinuous-CG with Gender (*Dc-CG+G*)	+	+	10	16971.6	**16883.8**
	-	+	9	16963.1	16884.1
	+	-	9	16967.9	16888.9
	-	-	8	16959.6	16889.3

[1]*Note*. **Bold** indicates best fitting models; "-"indicates absent gender by slope interaction; "+" indicates gender by slope interaction

discontinuous CG-model was the Discontinuous CG + Gender model, with Gender *x* Linear + Gender *x* Schooling interactions (-2LL = 16883.8), which will be referred to as Model G when determining the final GCM.

MLUw-Spanish: Final Growth Curve Model (GCM)

The parameter estimates from the UM-model (Model A), the best fitting UG-models (Models B, C, and D), and the two sets of best fitting CG-models (Models E, F, and G) were compared to determine the final GCM for MLUw-Spanish (see Table 8). Model A, the UM-model, was the baseline model of within-person residual variance at Level-1 (σ_ε^2 term) for all models. The two sets of best fitting CG-models each had a specific baseline model of between-person intercept (σ_0^2 term) and slope (σ_2^2 term; σ_3^2

Table 8. Comparisons of Growth Curve Model Parameter Estimates for Mean Length of Utterance in Words-Spanish[1,2]

		Model A UM	*Model B UG-Ln*	*Model C UG-Qd*	*Model D UG-Cb*	*Model E CG-Cb+G*	*Model F CG-Dc*	*Model G CG-Dc+G*
Fixed effects								
Intercept	γ_{00}	5.78***	4.86***	4.95***	4.67***	**4.56*** **	4.71***	4.60***
Linear Slope	γ_{10}		0.36***	0.24***	1.41***	**1.30*** **	0.62***	0.63***
Quadratic Slope	γ_{20}			0.02***	-0.61***	**-0.54*** **	0.87*** Schooling	0.82*** Schooling
Cubic Slope	γ_{30}				0.08***	**0.07*** **		
Gender (G)	γ_{01}					**0.21*** **		0.20***
G x Linear Slope	γ_{11}					**0.21***		-0.02
G x Quadratic Slope	γ_{21}					**-0.14** **		0.09* G x Schooling
G x Cubic Slope	γ_{31}					**0.02** **		
Variance components								
L1: Within-person variance	σ_ε^2	1.08*	0.696*	0.692*	0.560*	**0.558***	0.593*	0.592*
L2: B/w-person intercept	σ_0^2	0.245*	0.285*	0.285*	0.320*	**0.301***	0.285*	0.270*
L2: B/w-person quadratic slope	σ_2^2				0.003	**0.003**	0.016 σ_1^2 Linear	0.016 σ_1^2 Linear
L2: B/w-person cubic slope	σ_3^2				0.0002	**0.0001**		

Table 8. (continued)

		Model A UM	*Model B UG-Ln*	*Model C UG-Qd*	*Model D UG-Cb*	*Model E CG-Cb+G*	*Model F CG-Dc*	*Model G CG-Dc+G*
Covariance (σ_0^2, σ_2^2)		σ_{02}			-0.016*	**-0.013**	0.004 σ_{01}	0.005 σ_{01}
Covariance (σ_0^2, σ_3^2)		σ_{03}			0.0036*	**0.0032***		
Covariance (σ_2^2, σ_3^2)		σ_{23}			-0.0007	**-0.0007**		
Proportional variance reduction								
L1: Within-person variance	R_ε^2		36%	36%	48%	**48%**	45%	45%
L2: B/w-person intercept	R_0^2					**6%**		5%
L2: B/w-person quadratic slope	R_2^2					**<1%**		<1% R_1^2 Linear
L2: B/w-person cubic slope	R_3^2					**50%**		
Goodness-of-fit								
-2LL		20008.0	17645.7*	17614.4*	16812.5*	**16757.6***	16934.2	16883.8*
BIC		20034.4	17680.8	17658.4	16909.1	**16889.3**	16995.7	16971.6

*$p < .05$, **$p < .01$, ***$p < .001$

[1]*Note.* **Bold** indicates final model; Shading indicates non-applicable Pseudo-R^2

[2]*Key*. UM: Unconditional means model; UG-Ln: Unconditional linear growth model; UG-Qd: Unconditional quadratic growth model; UG-Cb: Unconditional cubic growth model; CG-Cb+G: Conditional cubic growth model with gender; CG-Dc: Conditional discontinuous growth model; CG-Dc+G: Conditional discontinuous growth model with gender; L1: Level-1 submodel; L2: Level-2 submodel

term) residual variances at Level-2. Model D served as the baseline model for Model E, and Model F served as the baseline model (of Level-2 residual variances) for Model G.

Table 8 outlines the following parameter estimates for the growth curve models tested: (1) fixed effects (i.e., intercepts, slopes, covariates, and interactions); (2) variance components (i.e., Level-1 within-person residual variance, Level-2 between-person residual variances, and covariances); (3) proportional variance reduction (i.e., Level-1 and Level-2 Pseudo-R^2s), and (4) goodness-of-fit indices for nested (-2LL) and non-nested (BIC) models. Models A-D were nested within Model E. Model F was nested within Model G, and Models E and G were non-nested models. If fixed effects were significant, they were included in subsequent nested models. Further, residual variances left to be explained justified subsequent nested models. The final GCM was determined by comparisons of each model's proportional variance reduction (Pseudo-R_ε^2; Pseudo-R_0^2; Pseudo-R_1^2; Pseudo-R_2^2; and Pseudo-R_3^2) and goodness-of-fit, with the final model having the highest proportional variance reduction and the lowest goodness-of-fit indices.

As can be seen in Table 8, Models D and E accounted for the greatest proportional variance reduction (48%) of within-person residual variance at Level-1, relative to Model A. However, Model E also accounted for a 6% proportional variance reduction in between-person intercept (Pseudo-R_0^2) residual variance, and a 50% proportional variance reduction in between-person cubic slope (Pseudo-R_3^2) residual variance relative to Model D. The two discontinuous models (Models F and G) accounted for a 45% proportional variance reduction of within-person residual variance

at Level-1, relative to Model A. The proportional variance reduction of between-person intercept and slope residual variances of Model G was relative to Model F. However, Model G further accounted for a 5% proportional variance reduction in between-person intercept (Pseudo-R_0^2) residual variance, but did not improve the proportional variance reduction in between-person linear slope (Pseudo-R_1^2) residual variance relative to Model F.

Goodness-of-fit indices indicated the degree of fit (correspondence) that the growth curve models had to the sample data. The final GCM had the smallest goodness-of-fit indices, relative to all other growth curve models. As suggested by Singer and Willett (2003), -2LL was used to compare the fit of nested models, and BIC was used to compare the fit of non-nested models. Therefore, -2LL was used to compare the fit of Model B relative to Model A, Model C relative to Model B, Model D relative to Model C, and Model E relative to Model D. The -2LL difference between Model B relative to Model A (2362.3) exceeded the 0.0005 critical level (12.12) for a χ^2 distribution on 1 *df*, indicating that the effect of adding linear time to Model B provided a significantly better fit than Model A to the sample data. The -2LL difference between Model C relative to Model B (31.3) exceeded the 0.0005 critical level (12.12) for a χ^2 distribution on 1 *df*, indicating that the effect of adding quadratic time to Model C provided a significantly better fit than Model B to the sample data. The -2LL difference between Model D relative to Model C (801.9) exceeded the 0.0005 critical level (24.10) for a χ^2 distribution on 6 *df*, indicating that the effect of adding cubic time to Model D provided a

significantly better fit than Model C to the sample data. The -2LL difference between Model E relative to Model D (54.9) exceeded the 0.0005 critical level (20.0) for a χ^2 distribution on 4 *df*, indicating that the effect of adding Gender to Model E provided a significantly better fit than Model D to the sample data. Therefore, Model E provided the best fit to the sample data, relative to its nested models (Models A-D).

Similarly, -2LL was also used to compare the fit of Model G relative to Model F. The -2LL difference between Model G relative to Model F (50.4) exceeded the 0.0005 critical level (17.73) for a χ^2 distribution on 3 *df*, indicating that the effect of adding Gender to Model G provided a significantly better fit than Model F to the sample data. Therefore, Model G provided the best fit to the sample data, relative to its nested model (Model F).

BIC was used to compare the fit of Model E against the fit of Model G, which were the non-nested models. The χ^2 distribution test was not used to compare BICs, only their absolute values (Singer & Willett, 2003). The BIC of Model E (16889.3) was smaller than the BIC of Model G (16971.6) by 82.3, indicating that Model E provided the best fit to the sample data, relative to its non-nested model (Model G).

In summary, the results of these analyses indicated that Model E (CG-Cubic + Gender model with randomly varying quadratic + cubic slopes, and with Gender *x* Linear + Gender *x* Quadratic + Gender *x* Cubic interactions) was the final model for MLU-Spanish, which estimated the effect that Gender (being female) had on the initial status of growth, and on the linear and curvilinear (quadratic + cubic) growth rates. Model E had

the highest overall proportional variance reduction (Pseudo-R_{ε}^2 = 48%; Pseudo-R_0^2 = 6%; Pseudo-R_3^2 = 50%) and the lowest goodness-of-fit indices (-2LL = 16757.6; BIC = 16889.3) for MLUw-Spanish (see Table 8). The GCM formula for Model E-(MLUw-Spanish) was as follows:

Level-1: Within-person variation

$$\text{MLUw-Spanish}_{ij} = \pi_{0i} + \pi_{1i}\,(\text{TIME}_{ij}) + \pi_{2i}\,(\text{TIME}_{ij})^2 + \pi_{3i}\,(\text{TIME}_{ij})^3 + \varepsilon_{ij}$$

Level-2: Between-person variation

$$\pi_{0i} = \gamma_{00} + \gamma_{01}\,(\text{Gender}_i) + \zeta_{0i}$$

$$\pi_{1i} = \gamma_{10} + \gamma_{11}\,(\text{Gender}_i)$$

$$\pi_{2i} = \gamma_{20} + \gamma_{21}\,(\text{Gender}_i) + \zeta_{2i}$$

$$\pi_{3i} = \gamma_{30} + \gamma_{31}\,(\text{Gender}_i) + \zeta_{3i}$$

The fixed effects of this model indicated that boys' expected average initial status of MLUw-Spanish was γ_{00} = 4.56, p <.001, with a positive and significant instantaneous linear rate of change (γ_{10} = 1.30, p <.001) during the fall of kindergarten. The boys' expected rate of deceleration in the linear slope over time was negative and significant (γ_{20} = -0.54, p <.001), but their expected rate of acceleration in the quadratic slope over time was positive and significant (γ_{30} = 0.07, p <.001). There was a positive and significant effect of Gender (being female) on initial status (γ_{01} = 0.21, p <.001) and on instantaneous linear rate of change (γ_{11} = 0.21, p <.05) during the fall of kindergarten. The girls' expected rate of additional deceleration (relative to boys) in the linear slope over time was negative and significant (γ_{21} = -0.14, p <.01), but their expected rate of additional acceleration (relative to boys) in the quadratic slope over time was positive and

significant ($\gamma_{31} = 0.02$, $p < .01$). The variance components of this model indicated that the participants differed significantly from each other over time ($\sigma_\varepsilon^2 = 0.558$, $p < .05$), and that there were significant individual differences in initial status ($\sigma_0^2 = 0.301$, $p < .05$) after controlling for Gender. The covariance components of this model indicated that there were positive and significant individual differences in the initial status, cubic slope covariance ($\sigma_{03} = 0.0032$, $p < .05$) after controlling for Gender.

MLUw-Spanish: Prototypical Growth Trajectory

The fixed effects parameter estimates form the final model (Model E) were used to generate the prototypical growth trajectories by gender for MLUw-Spanish. The prototypical growth trajectory, illustrated in Figure 3, shows that the shapes of growth for boys and girls are curvilinear (cubic), non-monotonic, and continuous according to Elman et al. (1999) growth pattern taxonomy. Figure 3 also demonstrates a number of interesting data observations: (1) initial status differences with girls beginning with significantly higher MLUw-Spanish (4.98) relative to boys (4.56) during the fall of kindergarten; (2) periods of notable MLUw-Spanish growth rate acceleration for both genders during kindergarten (boys = 0.84; girls = 0.93) and second grade (boys = 1.17; girls = 1.36) with second grade acceleration outpacing kindergarten, and girls' overall acceleration outpacing boys; (3) the MLUw-Spanish growth rate decelerated between the spring semester of kindergarten and the fall semester of second grade for boys (0.48) and girls (0.32), with girls' overall deceleration outpacing boys; (4) the boys' decelerating growth rate, however, contributed to limited, positive growth (0.01) during first grade; (5)

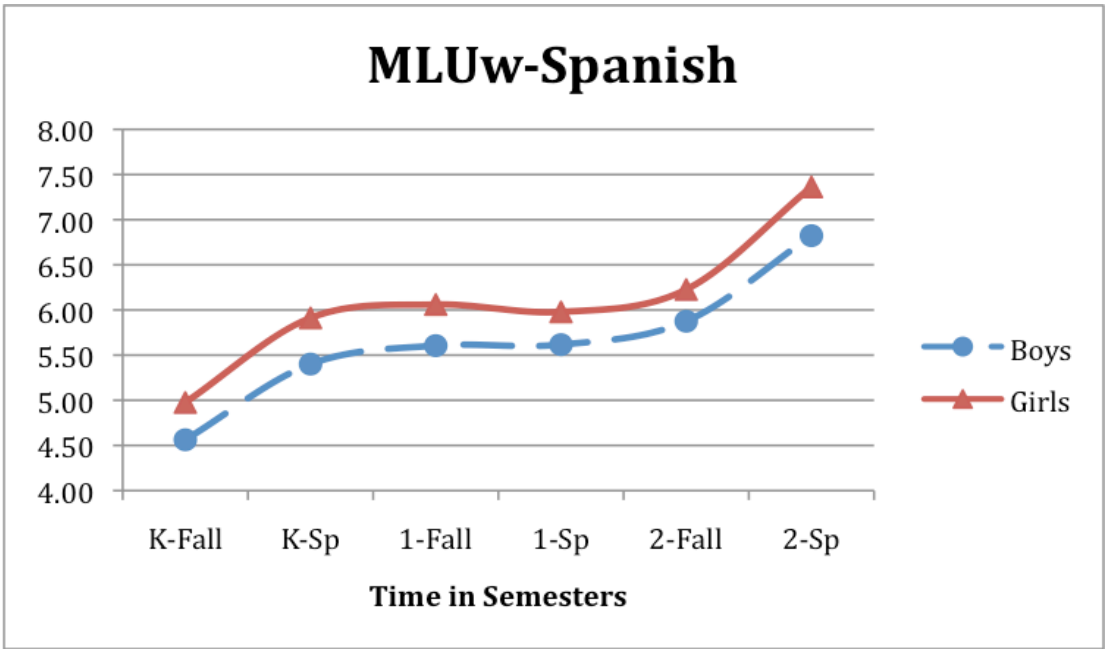


Figure 3. Prototypical Growth Trajectories by Gender for Mean Length of Utterance in Words-Spanish

in contrast, the girls' decelerating growth rate contributed to negative MLUw-Spanish growth (-0.08) during first grade first grade; and (6) girls demonstrated higher MLUw-Spanish (7.23) relative to boys (6.89) during the last wave of data collection, in the spring of second grade.

Number of Different Words (NDW) – Spanish

NDW, the total number of different uninflected word roots, was used as a measure of the lexical subsystem in Spanish. GCM analyses were conducted to

determine within- and between-person variation in initial status and growth rates of NDW-Spanish from the fall of kindergarten to the spring of second grade.

Unconditional Means (UM) Model

The UM-model for NDW-Spanish resulted in significant within-person residual variability to be explained at Level-1 ($\sigma_\varepsilon^2 = 342.34$, $p < .05$), significant between-person residual variability to be explained at Level-2 ($\sigma_0^2 = 169.65$, $p < .05$), and an *ICC* of .33. Without considering the effect of time or any other predictors, 33% of the total variation in NDW-Spanish was accounted for by differences between the participants in the sample. The UM-model will be referred to as Model A when determining the final growth curve model (GCM) for NDW-Spanish.

Unconditional Growth (UG) Model Testing

Testing a series of nested UG-models determined whether additional proportional variance reduction, relative to the UM-model, in NDW-Spanish could be accounted for by adding time as a time-varying covariate at Level-1. Three nested UG-models were tested for model fit with every possible combination of fixed and randomly varying slopes. Table 9 outlines the comparisons of fixed and randomly slopes for the series of UG-models tested and highlights the best fitting UG-Linear, UG-Quadratic, and UG-Cubic models for NDW-Spanish.

As can be seen in Table 9, the best fitting UG-Linear model was the UG-Linear model with fixed linear slopes (-2LL = 56444.3), which will be referred to as Model B

Table 9. Comparisons of Fixed and Randomly Varying Slopes for Unconditional Growth (UG) Models of Number of Different Words-Spanish[1]

	UG-Models: Fixed and Randomly Varying Slopes					
	Linear-Random	Quadratic-Random	Cubic-Random	*df*	BIC	-2LL
UG-Linear (*UG-Ln*)	+			NC	NC	NC
	-			4	56479.4	**56444.3**
UG-Quadratic (*UG-Qd*)	+	+		NC	NC	NC
	-	+		NC	NC	NC
	+	-		NC	NC	NC
	-	-		5	56414.5	**56370.6**
UG-Cubic (*UG-Cb*)	+	+	+	NC	NC	NC
	-	+	+	NC	NC	NC
	+	-	+	NC	NC	NC
	+	+	-	NC	NC	NC
	-	-	+	NC	NC	NC
	-	+	-	8	55894.4	**55824.1**
	+	-	-	8	55897.9	55827.7
	-	-	-	6	55882.8	55830.1

[1]*Note*. **Bold** indicates best fitting models; Shading indicates non-applicable randomly varying slopes; "-" indicates fixed slope; "+" indicates randomly varying slope; NC: Non-converging model

when determining the final GCM. The UG-Linear model with randomly varying linear slopes did not converge. Similarly, the best fitting UG-Quadratic model was the UG-Quadratic model with fixed linear + quadratic slopes (-2LL = 56370.6), which will be referred to as Model C when determining the final GCM. The other three UG-Quadratic models did not converge. The best fitting UG-Cubic model was the UG-Cubic model with varying quadratic slopes (-2LL = 55824.1), which will be referred to as Model D when determining the final GCM.

Comparisons of the best fitting UG-Linear, UG-Quadratic, and UG-Cubic models based on -2LL revealed that the best fitting UG-model was the UG-Cubic model with randomly varying quadratic slopes (-2LL = 55824.1). This UG-model (referred to as Model B when determining the final GCM) required participants to have the same linear and cubic growth rates, but allowed participants to vary in their quadratic growth rates.

Conditional Growth (CG) Model Testing

Two sets of CG-models were tested in order to determine whether additional proportional variance reduction, relative to the best fitting UG-models, could be accounted for by adding time-invariant covariates and/or time-varying covariates. One set of models was a series of CG-models with Gender added as a time-invariant covariate at Level-2. The second set was a series of discontinuous CG-models with schooling added as a time-varying covariate at Level-1 and Gender added at Level-2.

CG-Models with Gender. Based on the UG-model testing results, the UG-Cubic model with randomly varying quadratic slopes served as the baseline CG-model for NDW-Spanish. The CG-Cubic + Gender model was tested to determine the effect of Gender on initial status, and the interaction of Gender with the linear, quadratic, and cubic slopes. Table 10 outlines the comparisons of Gender *x* Slope interactions for the series of CG-Cubic + Gender models tested and highlights the best fitting CG-Cubic + Gender model for NDW-Spanish.

Table 10. Comparisons of Gender *x* Slope Interactions for Conditional Cubic Growth Models with Gender (CG-Cb+G) of Number of Different Words-Spanish[1]

	CG-Cubic + Gender models: Gender *x* Slope Interactions					
	Gender *x* Linear	Gender *x* Quadratic	Gender *x* Cubic	*df*	BIC	-2LL
CG-Cubic with Gender (*CG-Cb+G*)	+	+	+	12	55854.2	**55748.8**
	-	+	+	11	55853.8	55757.2
	+	-	+	11	55856.3	55759.7
	+	+	-	11	55856.6	55760.0
	-	-	+	10	55848.2	55760.3
	-	+	-	10	55847.8	55760.0
	+	-	-	10	55847.8	55760.0
	-	-	-	9	55839.5	55760.5

[1]*Note*. **Bold** indicates best fitting model; "-"indicates absent gender by slope interaction; "+" indicates gender by slope interaction; NC: Non-converging model

All CG-Cubic + Gender models converged. Comparisons of the eight CG-Cubic + Gender models based on -2LL revealed that the best fitting CG-model was the CG-Cubic + Gender model, with Gender *x* Linear + Gender *x* Quadratic + Gender *x* Cubic interactions (-2LL = 55748.8), which will be referred to as Model E when determining the final GCM.

Discontinuous CG-Models with Schooling and Gender. The second set of CG-models to be tested were discontinuous CG-models with schooling added as a time-varying covariate at Level-1 and Gender added as a time-invariant covariate at Level-2. A series of discontinuous CG-models with schooling *only* were tested with every possible combination of fixed and randomly varying slopes to determine the best fitting discontinuous CG-model first. The effect of Gender was then added to the best fitting discontinuous CG-model to determine whether additional proportional variance reduction (relative to the discontinuous CG-model without Gender) could be accounted for by adding Gender as a time-invariant covariate. Table 11 outlines the comparisons of fixed and randomly varying slopes for the series of discontinuous CG-models tested and highlights the best fitting discontinuous CG-model for NDW-Spanish.

The best fitting discontinuous CG-model was the discontinuous CG-model with fixed linear + schooling slopes (-2LL = 56434.9), which was the only discontinuous CG-model that converged. This discontinuous CG-model (referred to as Model F when

Table 11. Comparisons of Fixed and Randomly Varying Slopes for Conditional Discontinuous Growth (Dc-CG) Models of Number of Different Words-Spanish[1]

	Discontinuous CG-Models: Fixed and Randomly Varying Slopes				
	Linear-Random	Schooling-Random	*df*	BIC	-2LL
Discontinuous-CG (*Dc-CG*)	+	+	NC	NC	NC
	-	+	NC	NC	NC
	+	-	NC	NC	NC
	-	**-**	5	56478.8	**56434.9**

[1]*Note*. **Bold** indicates best fitting model; "**-**" indicates fixed slope; "+" indicates randomly varying slope; NC: Non-converging model

determining the final GCM) required participants to have the same linear and schooling growth rates.

The best fitting discontinuous CG-model was the discontinuous CG-model with fixed linear + schooling slopes (-2LL = 56434.9), which was the only discontinuous CG-model that converged. This discontinuous CG-model (referred to as Model F when determining the final GCM) required participants to have the same linear and schooling growth rates.

The effect of Gender was then added to the best fitting discontinuous CG-model to determine whether additional proportional variance reduction (relative to the discontinuous CG-model with fixed linear + schooling slopes) in NDW-Spanish could be accounted for by adding Gender as a time-invariant covariate. Gender (effect of being

female) was tested with every possible interaction of Gender *x* Slope to determine the effect of Gender on the linear and schooling slopes. A total of four Discontinuous CG + Gender models were tested to identify the best fitting Discontinuous CG + Gender model. Table 12 outlines the comparisons of Gender *x* Slope interactions for the series of Discontinuous CG + Gender models tested and highlights the best fitting Discontinuous CG + Gender model for NDW-Spanish.

Table 12. Comparisons of Gender *x* Slope Interactions for Conditional Discontinuous Growth Models with Gender (Dc-CG+G) of Number of Different Words-Spanish[1]

	Discontinuous CG-Models: Gender *x* Slope Interactions				
	Gender *x* Linear	Gender *x* Schooling	*df*	BIC	-2LL
Discontinuous-CG with Gender (*Dc-CG+G*)	+	+	8	56436.4	**56366.2**
	-	+	7	56429.1	56367.6
	+	-	7	56432.0	56370.5
	-	-	6	56424.8	56372.1

[1]*Note*. **Bold** indicates best fitting models; "**-**"indicates absent gender by slope interaction; "+" indicates gender by slope interaction

All Discontinuous CG + Gender models converged. Comparisons of the four Discontinuous CG + Gender models based on -2LL revealed that the best fitting discontinuous CG-model was the Discontinuous CG + Gender model, with Gender *x*

Linear + Gender *x* Schooling interactions (-2LL = 56366.2), which will be referred to as Model G when determining the final GCM.

NDW-Spanish: Final Growth Curve Model (GCM)

The parameter estimates from the UM-model (Model A), the best fitting UG-models (Models B, C, and D), and the two sets of best fitting CG-models (Models E, F, and G) were compared to determine the final GCM for NDW-Spanish (see Table 13). Model A, the UM-model, was the baseline model of within-person residual variance at Level-1 (σ_ε^2 term) for all models. The two sets of best fitting CG-models each had a specific baseline model of between-person intercept (σ_0^2 term) and slope (σ_2^2 term) residual variances at Level-2. Model D served as the baseline model for Model E, and Model F served as the baseline model (of Level-2 residual variances) for Model G. Neither Model F nor Model G had between-person slope (Pseudo-R_1^2) residual variances because both models employed fixed slopes.

Table 13 outlines the following parameter estimates for the growth curve models tested: (1) fixed effects (i.e., intercepts, slopes, covariates, and interactions); (2) variance components (i.e., Level-1 within-person residual variance, Level-2 between-person residual variances, and covariances); (3) proportional variance reduction (i.e., Level-1 and Level-2 Pseudo-R^2s), and (4) goodness-of-fit indices for nested (-2LL) and non-nested (BIC) models. Models A-D were nested within Model E, and Model F was nested within Model G. Models E and G were non-nested models. If fixed effects were

Table 13. Comparisons of Growth Curve Model Parameter Estimates for Number of Different Words-Spanish[1,2]

		Model A UM	*Model B UG-Ln*	*Model C UG-Qd*	*Model D UG-Cb*	*Model E CG-Cb+G*	*Model F CG-Dc*	*Model G CG-Dc+G*
Fixed effects								
Intercept	γ_{00}	74.92***	60.12***	57.51***	52.89***	**50.13*** **	59.77***	56.71***
Linear Slope	γ_{10}		5.81***	9.32***	28.36***	**25.69*** **	11.40***	11.73***
Quadratic Slope	γ_{20}			-0.69***	-10.92***	**-9.44*** **	6.95*** Schooling	6.11*** Schooling
Cubic Slope	γ_{30}				1.36***	**1.16*** **		
Gender (G)	γ_{01}					**5.46*** **		6.03***
G x Linear Slope	γ_{11}					**5.17** **		-0.63
G x Quadratic Slope	γ_{21}					**-2.87** **		1.63* G x Schooling
G x Cubic Slope	γ_{31}					**0.38** **		
Variance components								
L1: Within-person variance	σ_ε^2	342.34*	241.61*	237.58*	212.16*	**211.74***	241.07*	240.86*
L2: B/w-person intercept	σ_0^2	169.65*	185.97*	188.48*	211.85*	**200.64***	186.38*	176.48*
L2: B/w-person quadratic slope	σ_2^2				0.006	**0.006**		
Covariance (σ_0^2, σ_2^2)	σ_{02}				-0.909*	**-0.847**		

Table 13. (continued)

		Model A UM	*Model B UG-Ln*	*Model C UG-Qd*	*Model D UG-Cb*	*Model E CG-Cb+G*	*Model F CG-Dc*	*Model G CG-Dc+G*
Proportional variance reduction								
L1: Within-person variance	R_ε^2		29%	31%	38%	**38%**	30%	30%
L2: B/w-person intercept	R_0^2					**5%**		5%
L2: B/w-person quadratic slope	R_2^2					**<1%**		
Goodness-of-fit								
-2LL		58226.6	56444.3*	56370.6*	55824.1*	**55748.8***	56434.9	56366.2*
BIC		58253.0	56479.4	56414.5	55894.4	**55854.2**	56478.8	56436.4

*$p < .05$, **$p < .01$, ***$p < .001$

[1]*Note.* **Bold** indicates final model; Shading indicates non-applicable Pseudo-R^2

[2]*Key*. UM: Unconditional means model; UG-Ln: Unconditional linear growth model; UG-Qd: Unconditional quadratic growth model; UG-Cb: Unconditional cubic growth model; CG-Cb+G: Conditional cubic growth model with gender; CG-Dc: Conditional discontinuous growth model; CG-Dc+G: Conditional discontinuous growth model with gender; L1: Level-1 submodel; L2: Level-2 submodel

significant, they were included in subsequent nested models. The final GCM was determined by comparisons of each model's proportional variance reduction (Pseudo-R_{ε}^2; Pseudo-R_0^2; and Pseudo-R_2^2) and goodness-of-fit, with the final model having the highest proportional variance reduction and the lowest goodness-of-fit indices.

As can be seen in Table 13, Models D and E accounted for the greatest proportional variance reduction (38%) of within-person residual variance at Level-1, relative to Model A. However, Model E also accounted for a 5% proportional variance reduction in between-person intercept (Pseudo-R_0^2) residual variance relative to Model D. The two discontinuous models (Models F and G) accounted for a 30% proportional variance reduction of within-person residual variance at Level-1, relative to Model A. Model G further accounted for a 5% proportional variance reduction in between-person intercept (Pseudo-R_0^2) residual variance.

Goodness-of-fit indices (-2LL and BIC) indicated the degree of fit (correspondence) that the growth curve models had to the sample data. The final GCM had the smallest goodness-of-fit indices, relative to all other growth curve models. As suggested by Singer and Willett (2003), -2LL was used to compare the fit of nested models, and BIC was used to compare the fit of non-nested models. Therefore, -2LL was used to compare the fit of Model B (Linear) relative to Model A, Model C relative to Model B, Model D relative to Model C, and Model E relative to Model D. The -2LL difference between Model B relative to Model A (1782.3) exceeded the 0.0005 critical level (12.12) for a χ^2 distribution on 1 *df*, indicating that the effect of adding linear time

to Model B provided a significantly better fit than Model A to the sample data. The -2LL difference between Model C relative to Model B (73.7) exceeded the 0.0005 critical level (12.12) for a χ^2 distribution on 1 *df*, indicating that the effect of adding quadratic time to Model C provided a significantly better fit than Model B to the sample data. The -2LL difference between Model D relative to Model C (546.5) exceeded the 0.0005 critical level (17.73) for a χ^2 distribution on 3 *df*, indicating that the effect of adding cubic time to Model D provided a significantly better fit than Model C to the sample data. The -2LL difference between Model E relative to Model D (75.3) exceeded the 0.0005 critical level (20.0) for a χ^2 distribution on 4 *df*, indicating that the effect of adding Gender to Model E provided a significantly better fit than Model D to the sample data. Therefore, Model E provided the best fit to the sample data, relative to its nested models (Models A-D).

Similarly, -2LL was also used to compare the fit of Model G relative to Model F. The -2LL difference between Model G relative to Model F (68.7) exceeded the 0.0005 critical level (17.73) for a χ^2 distribution on 3 *df*, indicating that the effect of adding Gender to Model G provided a significantly better fit than Model F to the sample data. Therefore, Model G provided the best fit to the sample data, relative to its nested model (Model F).

BIC was used to compare the fit of Model E against the fit of Model G, which were the non-nested models. The χ^2 distribution test was not used to compare BICs, only their absolute values (Singer & Willett, 2003). The BIC of Model E (55854.2) was

smaller than the BIC of Model G (56436.4) by 582.2, indicating that Model E provided the best fit to the sample data, relative to its non-nested model (Model G).

In summary, the results of these analyses indicated that Model E (CG-Cubic + Gender model with randomly varying quadratic + cubic slopes, and with Gender *x* Linear + Gender *x* Quadratic + Gender *x* Cubic interactions) was the final model for NDW-Spanish, which estimated the effect that Gender (being female) had on the initial status of growth, and on the linear and curvilinear (quadratic + cubic) growth rates. Model E had the highest overall proportional variance reduction (Pseudo-$R_{\varepsilon}^2 = 38\%$; Pseudo-$R_0^2 = 5\%$) and the lowest goodness-of-fit indices (-2LL = 55748.8; BIC = 55854.2) for NDW-Spanish (see Table 13). The GCM formula for Model E-(NDW-Spanish) was expressed as follows:

Level-1: Within-person variation

$$\text{NDW-Spanish}_{ij} = \pi_{0i} + \pi_{1i}\,(\text{TIME}_{ij}) + \pi_{2i}\,(\text{TIME}_{ij})^2 + \pi_{3i}\,(\text{TIME}_{ij})^3 + \varepsilon_{ij}$$

Level-2: Between-person variation

$$\pi_{0i} = \gamma_{00} + \gamma_{01}\,(\text{Gender}_i) + \zeta_{0i}$$

$$\pi_{1i} = \gamma_{10} + \gamma_{11}\,(\text{Gender}_i)$$

$$\pi_{2i} = \gamma_{20} + \gamma_{21}\,(\text{Gender}_i) + \zeta_{2i}$$

$$\pi_{3i} = \gamma_{30} + \gamma_{31}\,(\text{Gender}_i)$$

The fixed effects of this model indicated that boys' expected average initial status of NDW-Spanish was $\gamma_{00} = 50.13$, $p < .001$, with a positive and significant instantaneous linear rate of change ($\gamma_{10} = 25.69$, $p < .001$) during the fall of kindergarten. The boys' expected rate of deceleration in the linear slope over time was negative and significant

(γ_{20} = -9.44, p <.001), but their expected rate of acceleration in the quadratic slope over time was positive and significant (γ_{30} = 1.16, p <.001). There was a positive and significant effect of Gender (being female) on initial status (γ_{01} = 5.46, p <.001) and on instantaneous linear rate of change (γ_{11} = 5.17, p <.01) during the fall of kindergarten. The girls' expected rate of additional deceleration (relative to boys) in the linear slope over time was negative and significant (γ_{21} = -2.87, p <.01), but their expected rate of additional acceleration (relative to boys) in the quadratic slope over time was positive and significant (γ_{31} = 0.38, p <.01). The variance components of this model indicated that the participants differed significantly from each other over time (σ_{ε}^2 = 211.74, p <.05), and that there were significant individual differences in initial status (σ_0^2 = 200.64, p <.05) after controlling for Gender.

NDW-Spanish: Prototypical Growth Trajectory

The fixed effects parameter estimates form the final model were used to generate the prototypical growth trajectories by gender for NDW-Spanish. The final model for NDW-Spanish was the CG-Cubic + Gender model with randomly varying quadratic slopes, and with Gender *x* Linear + Gender *x* Quadratic + Gender *x* Cubic interactions. The prototypical growth trajectory, illustrated in Figure 4, shows that the shapes of growth for boys and girls are curvilinear (cubic), non-monotonic, and continuous. Figure 4 also demonstrates a number of interesting data observations: (1) initial status differences with girls beginning with significantly higher NDW-Spanish (61.05) relative to boys (50.13) during the fall of kindergarten; (2) periods of notable NDW-Spanish

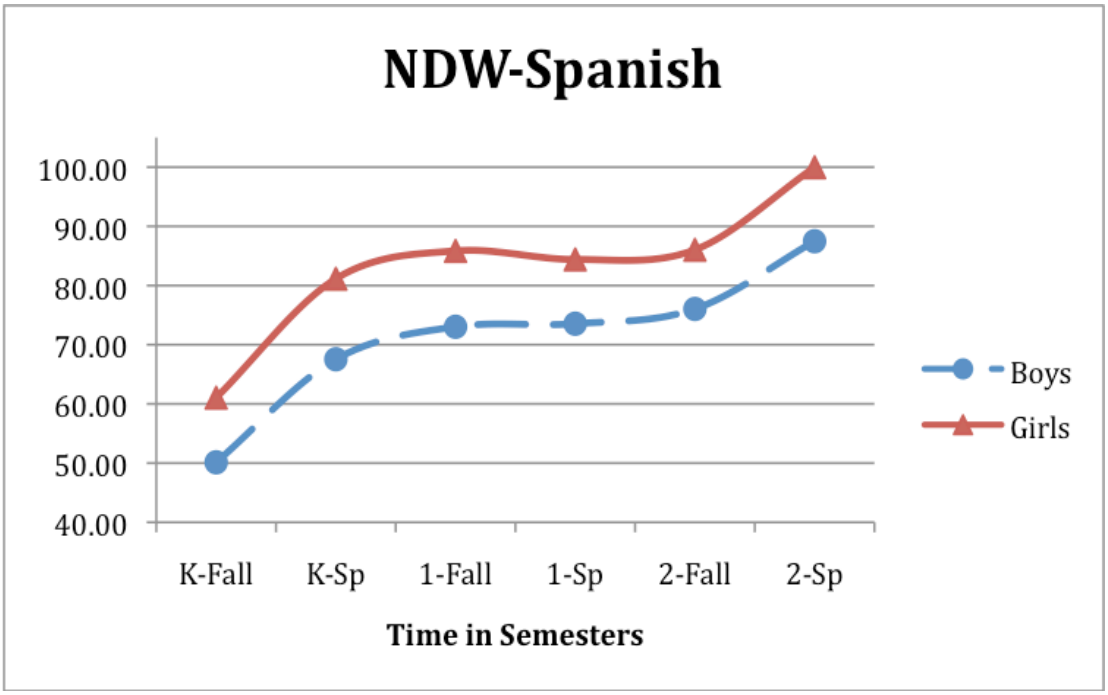


Figure 4. Prototypical Growth Trajectories by Gender for Number of Different Words-Spanish

growth rate acceleration for both genders during kindergarten (boys = 17.41; girls = 20.09) and second grade (boys = 11.44; girls = 13.93) with kindergarten acceleration outpacing second grade, and girls' overall acceleration outpacing boys; (3) the NDW-Spanish growth rate decelerated between the spring semester of kindergarten and the fall semester of second grade for boys (8.5) and girls (4.9), with girls' overall deceleration outpacing boys; (4) the boys decelerating growth rate, however, contributed to limited, positive NDW-Spanish growth (0.51) during first grade; (5) in contrast, the girls' decelerating growth rate contributed to negative NDW-Spanish growth (-1.45) during

first grade; and (6) girls demonstrated higher NDW-Spanish (99.97) relative to boys (87.48) during the last wave of data collection, in the spring of second grade.

Words Per Minute (WPM) – Spanish

WPM was used as a measure of general verbal fluency to represent linguistic subsystem integration in Spanish. GCM analyses were conducted to determine within- and between-person variation in initial status and growth rates of WPM-Spanish from the fall of kindergarten to the spring of second grade.

Unconditional Means (UM) Model

The UM-model for WPM-Spanish resulted in significant within-person residual variability to be explained at Level-1 ($\sigma_\varepsilon^2 = 328.24$, $p < .05$), significant between-person residual variability to be explained at Level-2 ($\sigma_0^2 = 152.78$, $p < .05$), and an *ICC* of .32. Without considering the effect of time or any other predictors, 32% of the total variation in WPM-Spanish was accounted for by differences between the participants in the sample. The UM-model will be referred to as Model A when determining the final growth curve model (GCM) for WPM-Spanish.

Unconditional Growth (UG) Model Testing

Testing a series of nested UG-models determined whether additional proportional variance reduction, relative to the UM-model, in WPM-Spanish could be accounted for by adding time as a time-varying covariate at Level-1. Three nested UG-models were tested for model fit with every possible combination of fixed and randomly varying

slopes. Table 14 outlines the comparisons of fixed and randomly slopes for the series of UG-models tested and highlights the best fitting (i.e., smallest -2LL) UG-Linear, UG-Quadratic, and UG-Cubic models for WPM-Spanish.

Table 14 demonstrates that the best fitting UG-Linear model was the UG-Linear model with randomly varying linear slopes (-2LL = 56936.7), which will be referred to as Model B when determining the final GCM. The best fitting UG-Quadratic model was the UG-Quadratic model with randomly varying linear slopes (-2LL = 56871.8), which will be referred to as Model C when determining the final GCM. The other three UG-Quadratic models did not converge. The best fitting UG-Cubic model was the UG-Cubic model with varying linear + quadratic + cubic slopes (-2LL = 56595.5), which will be referred to as Model D when determining the final GCM.

Comparisons of the best fitting UG-Linear, UG-Quadratic, and UG-Cubic models based on -2LL revealed that the best fitting UG-model was the UG-Cubic model with varying linear + quadratic + cubic slopes (-2LL = 56595.5). This UG-model (referred to as Model D when determining the final GCM) allowed participants to vary in their linear, quadratic, and cubic growth rates.

Conditional Growth (CG) Model Testing

Two sets of CG-models were tested in order to determine whether additional proportional variance reduction, relative to the best fitting UG-models, could be accounted for by adding time-invariant covariates and/or time-varying covariates. One

Table 14. Comparisons of Fixed and Randomly Varying Slopes for Unconditional Growth (UG) Models of Words per Minute-Spanish[1]

	UG-Models: Fixed and Randomly Varying Slopes					
	Linear-Random	Quadratic-Random	Cubic-Random	*df*	BIC	-2LL
UG-Linear (*UG-Ln*)	+			6	56989.3	**56936.7**
	-			4	56989.5	56954.3
UG-Quadratic (*UG-Qd*)	+	+		NC	NC	NC
	-	+		7	56939.1	56877.7
	+	-		7	56933.3	**56871.8**
	-	-		5	56934.5	56890.6
UG-Cubic (*UG-Cb*)	+	+	+	15	56727.2	**56595.5**
	-	+	+	NC	NC	NC
	+	-	+	NC	NC	NC
	+	+	-	NC	NC	NC
	-	-	+	8	56712.0	56641.7
	-	+	-	8	56704.4	56634.1
	+	-	-	8	56696.3	56626.1
	-	-	-	6	56708.9	56656.2

[1]*Note*. **Bold** indicates best fitting models; Shading indicates non-applicable randomly varying slopes; "-" indicates fixed slope; "+" indicates randomly varying slope; NC: Non-converging model

set of models was a series of CG-models with Gender added as a time-invariant covariate at Level-2. The second set was a series of discontinuous CG-models with schooling added as a time-varying covariate at Level-1 and Gender added at Level-2.

CG-Models with Gender. Based on the UG-model testing results, the UG-Cubic model with randomly varying linear + quadratic + cubic slopes served as the baseline CG-model for WPM-Spanish. The CG-Cubic + Gender model was tested to determine the effect of Gender on initial status, and the interaction of Gender with the linear, quadratic, and cubic slopes. Table 15 outlines the comparisons of Gender *x* Slope interactions for the series of CG-Cubic + Gender models tested and highlights the best fitting CG-Cubic + Gender model for WPM-Spanish.

All CG-Cubic + Gender models converged. Comparisons of the eight CG-Cubic + Gender models based on -2LL revealed that the best fitting CG-model was the CG-Cubic + Gender model, with Gender *x* Linear + Gender *x* Quadratic + Gender *x* Cubic interactions (-2LL = 56577.6), which will be referred to as Model E when determining the final GCM.

Discontinuous CG-Models with Schooling and Gender. The second set of CG-models to be tested were discontinuous CG-models with schooling added as a time-varying covariate at Level-1 and Gender added as a time-invariant covariate at Level-2. A series of discontinuous CG-models with schooling *only* were tested with every possible combination of fixed and randomly varying slopes to determine the best fitting discontinuous CG-model first. The effect of Gender was then added to the best fitting

Table 15. Comparisons of Gender *x* Slope Interactions for Conditional Cubic Growth Models with Gender (CG-Cb+G) of Words per Minute-Spanish[1]

	CG-Cubic + Gender models: Gender *x* Slope Interactions					
	Gender *x* Linear	Gender *x* Quadratic	Gender *x* Cubic	*df*	BIC	-2LL
CG-Cubic with Gender (*CG-Cb+G*)	+	+	+	19	56744.5	**56577.6**
	-	+	+	18	56741.2	56583.1
	+	-	+	18	56743.2	56585.1
	+	+	-	18	56742.9	56584.9
	-	-	+	17	56735.3	56586.0
	-	+	-	17	56734.2	56584.9
	+	-	-	17	56734.4	56585.1
	-	-	-	16	56729.7	56589.2

[1]*Note*. **Bold** indicates best fitting model; "**-**"indicates absent gender by slope interaction; "+" indicates gender by slope interaction; NC: Non-converging model

discontinuous CG-model to determine whether additional proportional variance reduction (relative to the discontinuous CG-model without Gender) could be accounted for by adding Gender as a time-invariant covariate. Table 16 outlines the comparisons of fixed and randomly varying slopes for the series of discontinuous CG-models tested and highlights the best fitting discontinuous CG-model for WPM-Spanish.

The best fitting discontinuous CG-model was the discontinuous CG-model with randomly varying linear slopes (-2LL = 56902.2), which will be referred to as Model F

Table 16. Comparisons of Fixed and Randomly Varying Slopes for Conditional Discontinuous Growth (Dc-CG) Models of Words per Minute-Spanish[1]

	Discontinuous CG-Models: Fixed and Randomly Varying Slopes				
	Linear-Random	Schooling-Random	*df*	BIC	-2LL
Discontinuous-CG (*Dc-CG*)	+	+	NC	NC	NC
	-	+	NC	NC	NC
	+	-	7	56963.1	**56902.2**
	-	-	5	56962.1	56918.2

[1]*Note*. **Bold** indicates best fitting model; "**-**" indicates fixed slope; "**+**" indicates randomly varying slope; NC: Non-converging model

when determining the final GCM. This discontinuous CG-model allowed participants to vary in their linear growth rates, but to have the same schooling growth rate.

The effect of Gender was then added to the best fitting discontinuous CG-model to determine whether additional proportional variance reduction (relative to the discontinuous CG-model with fixed linear + schooling slopes) in WPM-Spanish could be accounted for by adding Gender as a time-invariant covariate. Gender (effect of being female) was tested with every possible interaction of Gender *x* Slope to determine the effect of Gender on the linear and schooling slopes. A total of four Discontinuous CG + Gender models were tested to identify the best fitting Discontinuous CG + Gender model. Table 17 outlines the comparisons of Gender *x* Slope interactions for the series of

Table 17. Comparisons of Gender *x* Slope Interactions for Conditional Discontinuous Growth Models with Gender (Dc-CG+G) of Words per Minute-Spanish[1]

	Discontinuous CG-Models: Gender *x* Slope Interactions				
	Gender *x* Linear	Gender *x* Schooling	*df*	BIC	-2LL
Discontinuous-CG with Gender (*Dc-CG+G*)	+	+	10	56969.3	**56881.5**
	-	+	9	56969.3	56890.2
	+	-	9	56964.6	56885.6
	-	-	8	56965.0	56894.8

[1]*Note*. **Bold** indicates best fitting models; "**-**"indicates absent gender by slope interaction; "+" indicates gender by slope interaction

Discontinuous CG + Gender models tested and highlights the best fitting Discontinuous CG + Gender model for WPM-Spanish.

All Discontinuous CG + Gender models converged. Comparisons of the four Discontinuous CG + Gender models based on -2LL revealed that the best fitting discontinuous CG-model was the Discontinuous CG + Gender model, with Gender *x* Linear + Gender *x* Schooling interactions (-2LL = 56881.5), which will be referred to as Model G when determining the final GCM.

WPM-Spanish: Final Growth Curve Model (GCM)

The parameter estimates from the UM-model (Model A), the best fitting UG-models (Models B, C, and D), and the two sets of best fitting CG-models (Models E, F,

and G) were compared to determine the final GCM for WPM-Spanish (see Table 18). Model A, the UM-model, was the baseline model of within-person residual variance at Level-1 (σ_ε^2 term) for all models. The two sets of best fitting CG-models each had a specific baseline model of between-person intercept (σ_0^2 term) and slope (σ_1^2 term; σ_2^2 term; σ_3^2 term) residual variances at Level-2. Model D served as the baseline model for Model E, and Model F served as the baseline model (of Level-2 residual variances) for Model G.

Table 18 outlines the following parameter estimates for the growth curve models tested: (1) fixed effects (i.e., intercepts, slopes, covariates, and interactions); (2) variance components (i.e., Level-1 within-person residual variance, Level-2 between-person residual variances, and covariances); (3) proportional variance reduction (i.e., Level-1 and Level-2 Pseudo-R^2s), and (4) goodness-of-fit indices for nested (-2LL) and non-nested (BIC) models. Models A-D were nested within Model E, and Model F was nested within Model G. Models E and G were non-nested models. If fixed effects were significant, they were included in subsequent nested models. The final GCM was determined by comparisons of each model's proportional variance reduction (Pseudo-R_ε^2; Pseudo-R_0^2; Pseudo-R_1^2; Pseudo-R_2^2; and Pseudo-R_3^2) and goodness-of-fit, with the final model having the highest proportional variance reduction and the lowest goodness-of-fit indices.

As can be seen in Table 18, Models D and E accounted for the greatest proportional variance reduction (31%) of within-person residual variance at Level-1,

Table 18. Comparisons of Growth Curve Model Parameter Estimates for Words per Minute-Spanish[1,2]

		Model A UM	*Model B UG-Ln*	*Model C UG-Qd*	*Model D UG-Cb*	*Model E CG-Cb+G*	*Model F CG-Dc*	*Model G CG-Dc+G*
Fixed effects								
Intercept	γ_{00}	68.86***	55.01***	57.59***	54.63***	**53.54***	54.25***	52.70***
Linear Slope	γ_{10}		4.24***	0.80~	13.67***	**11.31***	8.04***	8.95***
Quadratic Slope	γ_{20}			0.68***	-6.31***	**-4.96***	6.63*** Schooling	5.79*** Schooling
Cubic Slope	γ_{30}				-0.93***	**0.76***		
Gender (G)	γ_{01}					**2.13~**		3.01***
G x Linear Slope	γ_{11}					**4.62***		-1.75**
G x Quadratic Slope	γ_{21}					**-2.64****		1.63* G x Schooling
G x Cubic Slope	γ_{31}					**0.34****		
Variance components								
L1: Within-person variance	σ_ε^2	328.24*	257.72*	255.09*	226.33*	**226.15***	256.28*	255.86*
L2: B/w-person intercept	σ_0^2	152.78*	187.76*	181.75*	168.75*	**167.85***	186.50*	183.39*
L2: B/w-person linear slope	σ_1^2		4.02*	3.94*	152.55*	**150.22***	16.60*	16.12*
L2: B/w-person quadratic slope	σ_2^2				39.26*	**38.51***		

Table 18. (continued)

		Model A UM	*Model B UG-Ln*	*Model C UG-Qd*	*Model D UG-Cb*	*Model E CG-Cb+G*	*Model F CG-Dc*	*Model G CG-Dc+G*
L2: B/w-person cubic slope	σ_3^2				0.65*	**0.64***		
Covariance (σ_0^2, σ_1^2)	σ_{01}		-7.49*	-6.40	22.07	**19.40**	-13.57*	-12.27
Covariance (σ_0^2, σ_2^2)	σ_{02}				-18.10*	**-16.64**		
Covariance (σ_0^2, σ_3^2)	σ_{03}				2.41	**2.23**		
Covariance (σ_1^2, σ_2^2)	σ_{12}				-74.54*	**-73.22***		
Covariance (σ_1^2, σ_3^2)	σ_{13}				9.38*	**9.22***		
Covariance (σ_2^2, σ_3^2)	σ_{23}				-5.03*	**-4.94**		
Proportional variance reduction								
L1: Within-person variance	R_ε^2		21%	22%	31%	**31%**	22%	22%
L2: B/w-person intercept	R_0^2					**1%**		<1%
L2: B/w-person linear slope	R_1^2					**2%**		2%
L2: B/w-person quadratic slope	R_2^2					**2%**		
L2: B/w-person cubic slope	R_3^2					**2%**		

Table 18. (continued)

	Model A UM	*Model B UG-Ln*	*Model C UG-Qd*	*Model D UG-Cb*	*Model E CG-Cb+G*	*Model F CG-Dc*	*Model G CG-Dc+G*
Goodness-of-fit							
-2LL	57888.5	56936.6*	56871.8*	56595.5*	**56577.6***	56902.2	56881.5*
BIC	57914.9	56989.3	56933.3	56727.2	**56744.5**	56963.7	56969.3

~$p < .10$, *$p < .05$, **$p < .01$, ***$p < .001$

[1]*Note.* **Bold** indicates final model; Shading indicates non-applicable Pseudo-R^2

[2]*Key*. UM: Unconditional means model; UG-Ln: Unconditional linear growth model; UG-Qd: Unconditional quadratic growth model; UG-Cb: Unconditional cubic growth model; CG-Cb+G: Conditional cubic growth model with gender; CG-Dc: Conditional discontinuous growth model; CG-Dc+G: Conditional discontinuous growth model with gender; L1: Level-1 submodel; L2: Level-2 submodel

relative to Model A. However, relative to Model D, Model E also accounted for the following proportional variance reductions: 1% in between-person intercept (Pseudo-R_0^2) residual variance, 2% in between-person linear slope (Pseudo-R_1^2) residual variance, 2% in between-person quadratic slope (Pseudo-R_2^2) residual variance, and 2% in between-person cubic slope (Pseudo-R_3^2) residual variance. The two discontinuous models (Models F and G) accounted for a 22% proportional variance reduction of within-person residual variance at Level-1, relative to Model A. Model G further accounted for a 2% proportional variance reduction in between-person linear slope (Pseudo-R_1^2) residual

variance, but did not improve the proportional variance reduction in between-person intercept (Pseudo-R_0^2) residual variance relative to Model F.

Goodness-of-fit indices (-2LL and BIC) indicated the degree of fit (correspondence) that the growth curve models had to the sample data. The final GCM had the smallest goodness-of-fit indices, relative to all other growth curve models. The -2LL difference between Model B relative to Model A (951.9) exceeded the 0.0005 critical level (17.73) for a χ^2 distribution on 3 *df*, indicating that the effect of adding linear time to Model B provided a significantly better fit than Model A to the sample data. The -2LL difference between Model C relative to Model B (64.8) exceeded the 0.0005 critical level (12.12) for a χ^2 distribution on 1 *df*, indicating that the effect of adding quadratic time to Model C provided a significantly better fit than Model B to the sample data. The -2LL difference between Model D relative to Model C (276.3) exceeded the 0.0005 critical level (27.87) for a χ^2 distribution on 8 *df*, indicating that the effect of adding cubic time to Model D provided a significantly better fit than Model C to the sample data. The -2LL difference between Model E relative to Model D (17.9) exceeded the 0.005 critical level (14.86) for a χ^2 distribution on 4 *df*, indicating that the effect of adding Gender to Model E provided a significantly better fit than Model D to the sample data. Therefore, Model E provided the best fit to the sample data, relative to its nested models (Models A-D).

Similarly, -2LL was also used to compare the fit of Model G relative to Model F. The -2LL difference between Model G relative to Model F (20.7) exceeded the 0.0005

critical level (17.73) for a χ^2 distribution on 3 *df*, indicating that the effect of adding Gender to Model G provided a significantly better fit than Model F to the sample data. Therefore, Model G provided the best fit to the sample data, relative to its nested model (Model F).

BIC was used to compare the fit of Model E against the fit of Model G, which were the non-nested models. The χ^2 distribution test was not used to compare BICs, only their absolute values (Singer & Willett, 2003). The BIC of Model E (56744.5) was smaller than the BIC of Model G (56969.3) by 224.8, indicating that Model E provided the best fit to the sample data, relative to its non-nested model (Model G).

In summary, the results of these analyses indicated that Model E (CG-Cubic + Gender model with randomly varying quadratic + cubic slopes, and with Gender *x* Linear + Gender *x* Quadratic + Gender *x* Cubic interactions) was the final model for WPM-Spanish, which estimated the effect that Gender (being female) had on the initial status of growth, and on the linear and curvilinear (quadratic + cubic) growth rates. Model E had the highest overall proportional variance reduction (Pseudo-$R_\varepsilon^2 = 31\%$; Pseudo-$R_0^2 = 1\%$; Pseudo-$R_1^2 = 2\%$; Pseudo-$R_2^2 = 2\%$; Pseudo-$R_3^2 = 2\%$) and the lowest goodness-of-fit indices (-2LL = 56577.6; BIC = 56744.5) for WPM-Spanish (see Table 18). The GCM formula for Model E-(WPM-Spanish) was expressed as follows:

Level-1: Within-person variation

$$\text{WPM-Spanish}_{ij} = \pi_{0i} + \pi_{1i}\,(\text{TIME}_{ij}) + \pi_{2i}\,(\text{TIME}_{ij})^2 + \pi_{3i}\,(\text{TIME}_{ij})^3 + \varepsilon_{ij}$$

Level-2: Between-person variation $\pi_{0i} = \gamma_{00} + \gamma_{01}\,(\text{Gender}_i) + \zeta_{0i}$

$\pi_{1i} = \gamma_{10} + \gamma_{11}\,(\text{Gender}_i) + \zeta_{1i}$

$\pi_{2i} = \gamma_{20} + \gamma_{21}\,(\text{Gender}_i) + \zeta_{2i}$

$\pi_{3i} = \gamma_{30} + \gamma_{31}\,(\text{Gender}_i) + \zeta_{3i}$

The fixed effects of this model indicated that boys' expected average initial status of WPM-Spanish was $\gamma_{00} = 53.54$, $p < .001$, with a positive and significant instantaneous linear rate of change ($\gamma_{10} = 11.31$, $p < .001$) during the fall of kindergarten. The boys' expected rate of deceleration in the linear slope over time was negative and significant ($\gamma_{20} = -4.96$, $p < .001$), but their expected rate of acceleration in the quadratic slope over time was positive and significant ($\gamma_{30} = 0.76$, $p < .001$). There was a positive effect of Gender (being female) on initial status ($\gamma_{01} = 2.13$, $p = .08$) that was non-significant, and a positive and significant effect on instantaneous linear rate of change ($\gamma_{11} = 4.62$, $p < .05$) during the fall of kindergarten. The girls' expected rate of additional deceleration (relative to boys) in the linear slope over time was negative and significant ($\gamma_{21} = -2.64$, $p < .01$), but their expected rate of additional acceleration (relative to boys) in the quadratic slope over time was positive and significant ($\gamma_{31} = 0.34$, $p < .01$). The variance components of this model indicated that the participants differed significantly from each other over time ($\sigma_\varepsilon^2 = 226.15$, $p < .05$), and that there were significant individual differences in initial status ($\sigma_0^2 = 167.85$, $p < .05$), in the linear slope ($\sigma_1^2 = 150.22$, $p < .05$), in the quadratic slope ($\sigma_2^2 = 38.51$, $p < .05$), and in the cubic slope ($\sigma_3^2 = 0.64$, $p < .05$) after controlling for Gender. The covariance components of this model indicated

that there were negative and significant individual differences in the linear slope, quadratic slope covariance ($\sigma_{12} = -73.22$, $p < .05$), and positive and significant individual differences in the linear slope, cubic slope covariance ($\sigma_{13} = 9.22$, $p < .05$) after controlling for Gender.

WPM-Spanish: Prototypical Growth Trajectory

The fixed effects parameter estimates form the final model were used to generate the prototypical growth trajectories by gender for WPM-Spanish. The final model for WPM-Spanish was the CG-Cubic + Gender model with randomly varying linear + quadratic + cubic slopes, and with Gender *x* Linear + Gender *x* Quadratic + Gender *x* Cubic interactions. The prototypical growth trajectory, illustrated in Figure 5, shows that the shapes of growth for boys and girls are curvilinear (cubic), non-monotonic, and continuous. Figure 5 also demonstrates a number of interesting data observations: (1) initial status (non-significant) differences with girls beginning with higher WPM-Spanish (57.79) relative to boys (53.54) during the fall of kindergarten; (2) periods of notable WPM-Spanish growth rate acceleration for both genders during kindergarten (boys = 7.1; girls = 9.42) and second grade (boys = 12.83; girls = 14.26) with second grade acceleration outpacing kindergarten, and girls' overall acceleration outpacing boys; (3) the WPM-Spanish growth rate decelerated between the spring semester of kindergarten and the spring semester of first grade for boys (2.62), and between the spring semester of kindergarten and the fall semester of second grade for girls (2.69), with girls' overall deceleration outpacing boys; (4) the boys' decelerating growth rate, however, contributed

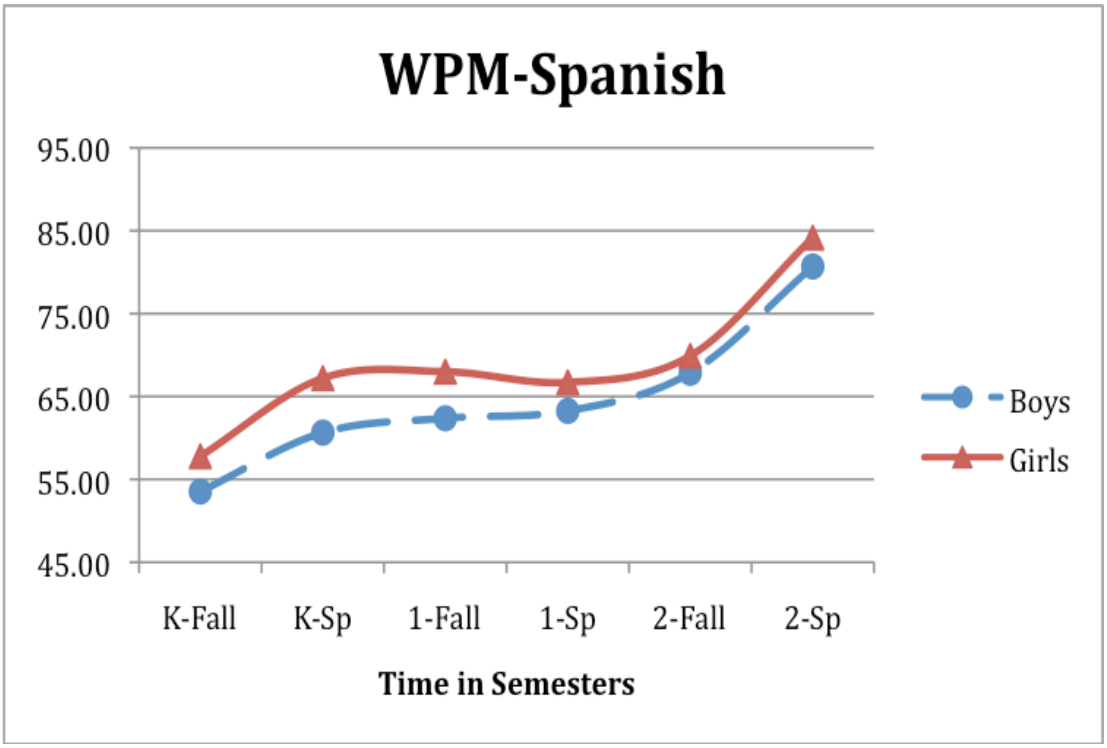


Figure 5. Prototypical Growth Trajectories by Gender for Words per Minute-Spanish

to limited, positive WPM-Spanish growth (0.89) during first grade; (5) in contrast, the girls' decelerating growth rate contributed to negative WPM-Spanish growth (-1.29) during first grade; and (6) girls demonstrated higher WPM-Spanish (84.16) relative to boys (80.68) during the last wave of data collection, in the spring of second grade.

Mean Length of Utterance in Words (MLUw) – English

MLUw was used as a general measure of the morphosyntactic subsystem in English. GCM analyses were conducted to determine within- and between-person

variation in initial status and growth rates of MLUw-English from the fall of kindergarten to the spring of second grade.

Unconditional Means (UM) Model

Fitting the UM-model determined the significance and degree of residual variation that was observed between participants for MLUw-English without taking into consideration time or any predictors. The residual variability of the UM-model was tested for significance, and the intraclass correlation coefficient (*ICC*) was calculated to determine the proportion of variation accounted for by differences between participants. The UM-model for MLUw-English resulted in significant within-person residual variability to be explained at Level-1 ($\sigma_\varepsilon^2 = 1.38$, $p < .05$), significant between-person residual variability to be explained at Level-2 ($\sigma_0^2 = 0.282$, $p < .05$), and an *ICC* of .17. Without considering the effect of time or any other predictors, 17% of the total variation in MLUw-English was accounted for by differences between the participants in the sample. The UM-model will be referred to as Model A when determining the final growth curve model (GCM) for MLUw-English.

Unconditional Growth (UG) Model Testing

Testing a series of UG-models determined whether additional proportional variance reduction, relative to the UM-model, in MLUw-English could be accounted for by adding time as a time-varying covariate at Level-1. The six waves of measurement in this study permitted modeling of linear and curvilinear continuous growth.

Consequentially, apart from linear time, two polynomial functions of time were also tested: quadratic and cubic time. Each of these three nested UG-models (i.e., the UG-Linear model was nested within the UG-Quadratic model; the UG-Quadratic model was nested within the UG-Cubic model) was tested for model fit with every possible combination of fixed and randomly varying slopes. Fixed slopes required all participants to have the same growth rates; randomly varying slopes allowed participants to vary in their growth rates. The -2 Log-Likelihood (-2LL) deviance statistic, which is used to compare model fit across nested GCMs (Singer & Willett, 2003), was used to identify the best fitting (i.e., smallest -2LL) UG-models. Table 19 outlines the comparisons of fixed and randomly slopes for the series of UG-models tested and highlights the best fitting UG-Linear, UG-Quadratic, and UG-Cubic models for MLUw-English. It should be noted that additional covariates added parameters to be estimated.

As can be seen in Table 19, the best fitting UG-Linear model was the UG-Linear model with randomly varying linear slopes (-2LL = 16913.6), which will be referred to as Model B when determining the final GCM. Similarly, the best fitting UG-Quadratic model was the UG-Quadratic model with randomly varying slopes (-2LL = 16887.8), which will be referred to as Model C when determining the final GCM. The best fitting UG-Cubic model was the UG-Cubic model with varying quadratic + cubic slopes (-2LL = 16514.4), which will be referred to as Model D when determining the final GCM.

Comparisons of the best fitting UG-Linear, UG-Quadratic, and UG-Cubic models based on -2LL revealed that the best fitting UG-model was the UG-Cubic model with

Table 19. Comparisons of Fixed and Randomly Varying Slopes for Unconditional Growth (UG) Models of Mean Length of Utterance in Words-English[1]

	UG-Models: Fixed and Randomly Varying Slopes					
	Linear-Random	Quadratic-Random	Cubic-Random	*df*	BIC	-2LL
UG-Linear (*UG-Ln*)	+			6	16965.6	**16913.6**
	-			4	17083.7	17049.1
UG-Quadratic (*UG-Qd*)	+	+		NC	NC	NC
	-	+		7	16958.8	16898.3
	+	**-**		7	16948.4	**16887.8**
	-	-		5	17058.4	17015.2
UG-Cubic (*UG-Cb*)	+	+	+	NC	NC	NC
	-	**+**	**+**	11	16609.6	**16514.4**
	+	-	+	NC	NC	NC
	+	+	-	NC	NC	NC
	-	-	+	8	16681.8	16612.6
	-	+	-	8	16641.6	16572.4
	+	-	-	8	16622.4	16553.2
	-	-	-	6	16789.9	16738.0

[1]*Note*. **Bold** indicates best fitting models; Shading indicates non-applicable randomly varying slopes; "**-**" indicates fixed slope; "**+**" indicates randomly varying slope; NC: Non-converging model

randomly varying quadratic + cubic slopes (-2LL = 16514.4). This UG-model (referred to as Model D when determining the final GCM) required participants to have the same linear growth rate, but allowed participants to vary in their quadratic and cubic growth rates.

Conditional Growth (CG) Model Testing

Two sets of CG-models were tested in order to determine whether additional proportional variance reduction, relative to the best fitting UG-models, could be accounted for by adding time-invariant covariates and/or time-varying covariates. One set of models was a series of CG-models with Gender added as a time-invariant covariate at Level-2. The second set was a series of discontinuous CG-models with schooling added as a time-varying covariate at Level-1 and Gender added at Level-2.

CG-Models with Gender. The first set of CG-models to be tested were CG-models with Gender added as a time-invariant covariate. Based on the UG-model testing results, the UG-Cubic model with randomly varying quadratic + cubic slopes served as the baseline CG-model for MLUw-English. The CG-Cubic + Gender model was tested to determine the effect of Gender on initial status. In addition, Gender was tested with every possible interaction of Gender *x* Slope to determine the effect of Gender on the linear, quadratic, and cubic slopes. A total of eight CG-Cubic + Gender models were tested to identify the best fitting (i.e., smallest -2LL) CG-Cubic + Gender model. Table 20 outlines the comparisons of Gender *x* Slope interactions for the series of CG-Cubic +

Table 20. Comparisons of Gender *x* Slope Interactions for Conditional Cubic Growth Models with Gender (CG-Cb+G) of Mean Length of Utterance in Words-English[1]

	CG-Cubic + Gender models: Gender *x* Slope Interactions					
	Gender *x* Linear	Gender *x* Quadratic	Gender *x* Cubic	*df*	BIC	-2LL
CG-Cubic with Gender (*CG-Cb+G*)	+	+	+	15	16631.1	**16501.3**
	-	+	+	14	16623.7	16502.5
	+	-	+	14	16625.2	16504.0
	+	+	-	14	16626.5	16505.3
	-	-	+	13	16618.5	16506.0
	-	+	-	13	16619.8	16507.2
	+	-	-	13	16621.2	16508.7
	-	-	-	12	16613.3	16509.5

[1]*Note*. **Bold** indicates best fitting model; "**-**"indicates absent gender by slope interaction; "+" indicates gender by slope interaction; NC: Non-converging model

Gender models tested and highlights the best fitting CG-Cubic + Gender model for MLUw-English. It should be noted that additional covariates adds parameters to be estimated, which in turn results in CG-models with increasing *df*s, as illustrated in Table 20.

All CG-Cubic + Gender models converged. Comparisons of the eight CG-Cubic + Gender models based on -2LL revealed that the best fitting CG-model was the CG-Cubic + Gender model, with Gender *x* Linear + Gender *x* Quadratic + Gender *x* Cubic

interactions (-2LL = 16501.3), which will be referred to as Model E when determining the final GCM.

Discontinuous CG-Models with Schooling and Gender. The second set of CG-models to be tested were discontinuous CG-models with schooling added as a time-varying covariate at Level-1 and Gender added as a time-invariant covariate at Level-2. Discontinuous CG-models model hypothesized discontinuities of growth at expected points in time. Discontinuous CG-models were necessary in this study to account for not having schooling during each summer preceding each fall semester, and for having schooling preceding each spring semester.

To isolate the effect of Schooling, a series of discontinuous CG-models with schooling *only* were tested with every possible combination of fixed and randomly varying slopes to determine the best fitting discontinuous CG-model first. The effect of Gender was then added to the best fitting discontinuous CG-model to determine whether additional proportional variance reduction (relative to the discontinuous CG-model without Gender) could be accounted for by adding Gender as a time-invariant covariate. Discontinuous CG-models modeled discontinuous linear growth only, since quadratic and cubic models assume continuous growth. Table 21 outlines the comparisons of fixed and randomly varying slopes for the series of discontinuous CG-models tested and highlights the best fitting discontinuous CG-model for MLUw-English.

Two of the four discontinuous CG-models converged. The best fitting discontinuous CG-model was the discontinuous CG-model with randomly varying linear

Table 21. Comparisons of Fixed and Randomly Varying Slopes for Conditional Discontinuous Growth (Dc-CG) Models of Mean Length of Utterance in Words-English[1]

	Discontinuous CG-Models: Fixed and Randomly Varying Slopes				
	Linear-Random	Schooling-Random	*df*	BIC	-2LL
Discontinuous-CG (*Dc-CG*)	+	+	NC	NC	NC
	-	+	NC	NC	NC
	+	**-**	7	15307.9	**15247.8**
	-	-	5	15430.8	15387.9

[1]*Note*. **Bold** indicates best fitting model; "**-**" indicates fixed slope; "**+**" indicates randomly varying slope; NC: Non-converging model

+ fixed schooling slopes (-2LL = 15247.8), which will be referred to as Model F when determining the final GCM. This discontinuous CG-model allowed participants to vary in their linear growth rates, but required them to have the same schooling growth rates during the spring semesters.

The effect of Gender was then added to the best fitting discontinuous CG-model to determine whether additional proportional variance reduction (relative to the discontinuous CG-model with randomly varying linear + fixed schooling slopes) in MLUw-English could be accounted for by adding Gender as a time-invariant covariate. Gender (effect of being female) was tested with every possible interaction of Gender *x* Slope to determine the effect of Gender on the linear and schooling slopes. A total of four Discontinuous CG + Gender models were tested to identify the best fitting

Discontinuous CG + Gender model. Table 22 outlines the comparisons of Gender *x* Slope interactions for the series of Discontinuous CG + Gender models tested and highlights the best fitting Discontinuous CG + Gender model for MLUw-English.

Table 22. Comparisons of Gender *x* Slope Interactions for Conditional Discontinuous Growth Models with Gender (Dc-CG+G) of Mean Length of Utterance in Words-English[1]

	Discontinuous CG-Models: Gender *x* Slope Interactions				
	Gender *x* Linear	Gender *x* Schooling	*df*	BIC	-2LL
Discontinuous-CG with Gender (*Dc-CG+G*)	+	+	10	15323.6	**15237.8**
	-	+	9	15315.3	15238.1
	+	-	9	15319.6	15242.4
	-	-	8	15311.1	15242.6

[1]*Note*. **Bold** indicates best fitting models; "**-**"indicates absent gender by slope interaction; "+" indicates gender by slope interaction

All Discontinuous CG + Gender models converged. Comparisons of the four Discontinuous CG + Gender models based on -2LL revealed that the best fitting discontinuous CG-model was the Discontinuous CG + Gender model, with Gender *x* Linear + Gender *x* Schooling interactions (-2LL = 15237.8), which will be referred to as Model G when determining the final GCM.

MLUw-English: Final Growth Curve Model (GCM)

The parameter estimates from the UM-model (Model A), the best fitting UG-models (Models B, C, and D), and the two sets of best fitting CG-models (Models E, F, and G) were compared to determine the final GCM for MLUw-English. Model A, the UM-model, was the baseline model of within-person residual variance at Level-1 (σ_ε^2 term) for all models. Therefore, the proportional variance reduction of within-person residual variance (Pseudo-R_ε^2) of each model was relative to Model A. The two sets of best fitting CG-models each had a specific baseline model of between-person intercept (σ_0^2 term) and slope (σ_1^2 term; σ_2^2 term; σ_3^2 term) residual variances at Level-2. Model D served as the baseline model for Model E, and Model F served as the baseline model for Model G.

Table 23 outlines the following parameter estimates for the growth curve models tested: (1) fixed effects (i.e., intercepts, slopes, covariates, and interactions); (2) variance components (i.e., Level-1 within-person residual variance, Level-2 between-person residual variances, and covariances); (3) proportional variance reduction (i.e., Level-1 and Level-2 Pseudo-R^2s), and (4) goodness-of-fit indices for nested (-2LL) and non-nested (BIC) models. Models A-D were nested within Model E, and Model F was nested within Model G. Models E and G were non-nested models. If fixed effects were significant, they were included in subsequent nested models. Further, residual variances left to be explained justified subsequent nested models. The final GCM was determined by comparisons of each model's proportional variance reduction (Pseudo-R_ε^2; Pseudo-

Table 23. Comparisons of Growth Curve Model Parameter Estimates for Mean Length of Utterance in Words-English[1,2]

		Model A UM	*Model B UG-Ln*	*Model C UG-Qd*	*Model D UG-Cb*	*Model E CG-Cb+G*	*Model F CG-Dc*	*Model G CG-Dc+G*
Fixed effects								
Intercept	γ_{00}	6.74***	5.52***	5.66***	5.30***	5.26***	5.34***	**5.33*****
Linear Slope	γ_{10}		0.43***	0.28***	1.32***	1.25***	0.77***	**0.76*****
Schooling Slope	γ_{20}			0.03*** Quad	-0.49*** Quad	-0.45*** Quad	0.93***	**0.88*****
Cubic Slope	γ_{30}				0.07***	0.06***		
Gender (G)	γ_{01}					0.08		**0.01**
G x Linear Slope	γ_{11}					0.13		**0.02**
G x Schooling Slope	γ_{21}					-0.09 G x Quad		**0.10***
G x Cubic Slope	γ_{31}					0.01*		
Variance components								
L1: Within-person variance	σ_ε^2	1.38*	0.780*	0.780*	0.681*	0.681*	0.689*	**0.688***
L2: B/w-person intercept	σ_0^2	0.282*	1.08*	1.04*	1.20*	1.19*	0.995*	**0.995***
L2: B/w-person linear slope	σ_1^2		0.039*	0.038*	0.014* σ_2^2 Quad	0.013* σ_2^2 Quad	0.147*	**0.147***
L2: B/w-person cubic slope	σ_3^2				0.0003	0.0002		

Table 23. (continued)

		Model A UM	*Model B UG-Ln*	*Model C UG-Qd*	*Model D UG-Cb*	*Model E CG-Cb+G*	*Model F CG-Dc*	*Model G CG-Dc+G*
Covariance (σ_0^2, σ_1^2)	σ_{01}		-0.171*	-0.163*	-0.120* σ_{02}	-0.119* σ_{02}	-0.310*	**-0.311***
Covariance (σ_0^2, σ_3^2)	σ_{03}				0.017*	0.017*		
Covariance (σ_2^2, σ_3^2)	σ_{23}				-0.002*	-0.002*		
Proportional variance reduction								
L1: Within-person variance	R_ε^2		43%	43%	51%	51%	50%	**50%**
L2: B/w-person intercept	R_0^2					1%		**<1%**
L2: B/w-person linear slope	R_1^2					7% R_2^2 Quad		**<1%**
L2: B/w-person cubic slope	R_3^2					33%		
Goodness-of-fit								
-2LL		18943.7	16913.6*	16887.8*	16514.4*	16501.3*	15247.8	**15237.8***
BIC		18969.6	16965.6	16948.4	16609.6	16631.1	15307.9	**15323.6**

*$p < .05$, ***$p < .001$

[1]*Note.* **Bold** indicates final model; Shading indicates non-applicable Pseudo-R^2

[2]*Key*. UM: Unconditional means model; UG-Ln: Unconditional linear growth model; UG-Qd: Unconditional quadratic growth model; UG-Cb: Unconditional cubic growth model; CG-Cb+G: Conditional cubic growth model with gender; CG-Dc: Conditional discontinuous growth model; CG-Dc+G: Conditional discontinuous growth model with gender; L1: Level-1 submodel; L2: Level-2 submodel

R_0^2; Pseudo-R_1^2; Pseudo-R_2^2; and Pseudo-R_3^2) and goodness-of-fit, with the final model having the highest proportional variance reduction and the lowest goodness-of-fit indices.

As can be seen in Table 23, Models D and E accounted for the greatest proportional variance reduction (51%) of within-person residual variance at Level-1, relative to Model A. However, Model E also accounted for a 1% proportional variance reduction in between-person intercept (Pseudo-R_0^2) residual variance, a 7% proportional variance reduction in between-person quadratic slope (Pseudo-R_2^2), and a 33% proportional variance reduction in between-person cubic slope (Pseudo-R_3^2) residual variance relative to Model D. The two discontinuous models (Models F and G) accounted for a 50% proportional variance reduction of within-person residual variance at Level-1, relative to Model A. The proportional variance reduction of between-person intercept and slope residual variances of Model G was relative to Model F. Model G did not account for additional proportional variance reduction in between-person intercept (Pseudo-R_0^2) or linear slope (Pseudo-R_1^2) residual variance relative to Model F.

Goodness-of-fit indices indicated the degree of fit (correspondence) that the growth curve models had to the sample data. The final GCM had the smallest goodness-of-fit indices, relative to all other growth curve models. The -2LL difference between Model B relative to Model A (2030.1) exceeded the 0.0005 critical level (17.73) for a χ^2 distribution on 3 *df*, indicating that the effect of adding linear time to Model B provided a significantly better fit than Model A to the sample data. The -2LL difference between Model C relative to Model B (25.8) exceeded the 0.0005 critical level (12.12) for a χ^2

distribution on 1 *df*, indicating that the effect of adding quadratic time to Model C provided a significantly better fit than Model B to the sample data. The -2LL difference between Model D relative to Model C (373.4) exceeded the 0.0005 critical level (20.0) for a χ^2 distribution on 4 *df*, indicating that the effect of adding cubic time to Model D provided a significantly better fit than Model C to the sample data. The -2LL difference between Model E relative to Model D (13.1) exceeded the 0.05 critical level (9.49) for a χ^2 distribution on 4 *df*, indicating that the effect of adding Gender to Model E provided a significantly better fit than Model D to the sample data. Therefore, Model E provided the best fit to the sample data, relative to its nested models (Models A-D).

Similarly, -2LL was also used to compare the fit of Model G relative to Model F. The -2LL difference between Model G relative to Model F (10.0) exceeded the 0.05 critical level (7.81) for a χ^2 distribution on 3 *df*, indicating that the effect of adding Gender to Model G provided a significantly better fit than Model F to the sample data. Therefore, Model G provided the best fit to the sample data, relative to its nested model (Model F).

BIC was used to compare the fit of Model E against the fit of Model G. The χ^2 distribution test was not used to compare BICs, only their absolute values (Singer & Willett, 2003). The BIC of Model G (15323.6) was smaller than the BIC of Model E (16631.1) by 1307.5, indicating that Model G provided the best fit to the sample data, relative to its non-nested model (Model E).

In summary, the results of these analyses indicated that Model G (Discontinuous CG + Gender model, with Gender *x* Linear + Gender *x* Schooling interactions) was the final model for MLU-English, which estimated the effect that Gender had on the initial status of growth and on the linear growth rates – including schooling. Although Model G did not have the highest overall proportional variance reduction (Pseudo-R^2s), it had the lowest goodness-of-fit indices (-2LL = 15237.8; BIC = 15323.6) for MLUw-English (see Table 23), which account for all the fixed effects and variance components. The GCM formula for Model G-(MLUw-English) was as follows:

Level-1: Within-person variation

$$\text{MLUw-English}_{ij} = \pi_{0i} + \pi_{1i}\,(\text{TIME}_{ij}) + \pi_{2i}\,(\text{SCHOOLING}_{ij}) + \varepsilon_{ij}$$

Level-2: Between-person variation

$$\pi_{0i} = \gamma_{00} + \gamma_{01}\,(\text{Gender}_i) + \zeta_{0i}$$

$$\pi_{1i} = \gamma_{10} + \gamma_{11}\,(\text{Gender}_i) + \zeta_{1i}$$

$$\pi_{2i} = \gamma_{20} + \gamma_{21}\,(\text{Gender}_i)$$

The fixed effects of this model indicated that boys' expected average initial status of MLUw-English was $\gamma_{00} = 5.33$, $p < .001$, with a positive and significant linear growth rate ($\gamma_{10} = 0.76$, $p < .001$) during each semester, and an additional positive and significant linear (schooling) growth rate ($\gamma_{20} = 0.88$, $p < .001$) during each spring semester. There was an additional positive and significant effect of Gender (being female) on the linear (schooling) growth rate ($\gamma_{21} = 0.10$, $p < .05$) during each spring semester, relative to boys. The variance components of this model indicated that the participants differed significantly from each other over time ($\sigma_\varepsilon^2 = 0.688$, $p < .05$), and that there were

significant individual differences in initial status ($\sigma_0^2 = 0.995$, $p < .05$) and in the linear slope ($\sigma_1^2 = 0.147$, $p < .05$) after controlling for Gender and Schooling. The covariance components of this model indicated that there were negative and significant individual differences in the initial status, linear slope covariance ($\sigma_{01} = -0.311$, $p < .05$) after controlling for Gender and Schooling.

MLUw-English: Prototypical Growth Trajectory

The fixed effects parameter estimates form the final model (Model G) were used to generate the prototypical growth trajectories by gender for MLUw-English. The prototypical growth trajectory, illustrated in Figure 6, shows that the shapes of growth for boys and girls are linear, non-monotonic, and discontinuous. Figure 6 also demonstrates a number of interesting data observations: (1) initial status (non-significant) differences with girls beginning with higher MLUw-English (5.35) relative to boys (5.33) during the fall of kindergarten; (2) periods of constant, linear MLUw-English growth for both genders (boys = 0.88; girls = 0.98) during each academic year (kindergarten, first grade, and second grade) with girls' linear growth outpacing boys; (3) the MLUw-English linear growth rate declined (negative growth) for both genders between the spring semester of Kindergarten and the fall semester of first grade, and again between the spring semester of first grade and the fall semester of second grade (boys = -0.12; girls = -0.2) with girls' linear growth decline during the summer outpacing boys; and (4) girls demonstrated higher MLUw-English (7.89) relative to boys (7.73) during the last wave of data collection, in the spring of second grade.

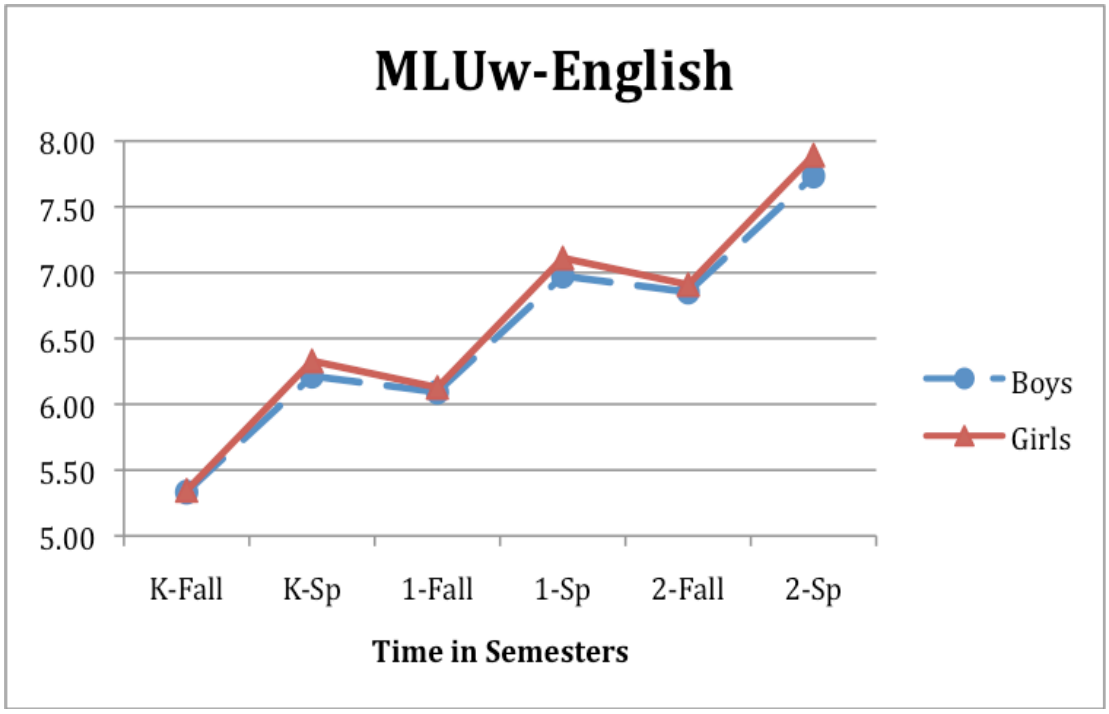


Figure 6. Prototypical Growth Trajectories by Gender for Mean Length of Utterance in Words-English

Number of Different Words (NDW) – English

NDW, the total number of different uninflected word roots, was used as a measure of the lexical subsystem in English. GCM analyses were conducted to determine within- and between-person variation in initial status and growth rates of NDW-English from the fall of kindergarten to the spring of second grade.

Unconditional Means (UM) Model

The UM-model for NDW-English resulted in significant within-person residual variability to be explained at Level-1 ($\sigma_\varepsilon^2 = 482.93$, $p < .05$), significant between-person residual variability to be explained at Level-2 ($\sigma_0^2 = 280.83$, $p < .05$), and an *ICC* of .37. Without considering the effect of time or any other predictors, 37% of the total variation in NDW-English was accounted for by differences between the participants in the sample. The UM-model will be referred to as Model A when determining the final growth curve model (GCM) for NDW-English.

Unconditional Growth (UG) Model Testing

Testing a series of nested UG-models determined whether additional proportional variance reduction, relative to the UM-model, in NDW-English could be accounted for by adding time as a time-varying covariate at Level-1. Three nested UG-models were tested for model fit with every possible combination of fixed and randomly varying slopes. Table 24 outlines the comparisons of fixed and randomly slopes for the series of UG-models tested and highlights the best fitting (i.e., smallest -2LL) UG-Linear, UG-Quadratic, and UG-Cubic models for NDW-English.

As can be seen in Table 24, the best fitting UG-Linear model was the UG-Linear model with randomly varying linear slopes (-2LL = 50227.9), which will be referred to as Model B when determining the final GCM. Similarly, the best fitting UG-Quadratic model was the UG-Quadratic model with randomly varying linear + quadratic slopes

Table 24. Comparisons of Fixed and Randomly Varying Slopes for Unconditional Growth (UG) Models of Number of Different Words-English[1]

	UG-Models: Fixed and Randomly Varying Slopes					
	Linear-Random	Quadratic-Random	Cubic-Random	*df*	BIC	-2LL
UG-Linear (*UG-Ln*)	+			6	50279.8	**50227.9**
	-			4	50326.7	50292.1
UG-Quadratic (*UG-Qd*)	+	+		10	50239.1	**50152.6**
	-	+		7	50257.9	50197.3
	+	-		7	50274.7	50214.1
	-	-		5	50312.7	50269.4
UG-Cubic (*UG-Cb*)	+	+	+	NC	NC	NC
	-	+	+	11	49922.0	49826.9
	+	-	+	11	49918.5	49823.3
	+	+	-	11	49905.9	**49810.7**
	-	-	+	8	49927.7	49858.5
	-	+	-	8	49910.2	49841.0
	+	-	-	8	49913.4	49844.2
	-	-	-	6	50005.6	49953.6

[1]*Note*. **Bold** indicates best fitting models; Shading indicates non-applicable randomly varying slopes; "**-**" indicates fixed slope; "**+**" indicates randomly varying slope; NC: Non-converging model

(-2LL = 50152.6), which will be referred to as Model C when determining the final GCM. The other three UG-Quadratic models did not converge. The best fitting UG-Cubic model was the UG-Cubic model with varying linear + quadratic slopes (-2LL = 49810.7), which will be referred to as Model D when determining the final GCM. The only UG-model that did not converge was the UG-Quadratic model with varying linear + quadratic + cubic slopes.

Comparisons of the best fitting UG-Linear, UG-Quadratic, and UG-Cubic models based on -2LL revealed that the best fitting UG-model was the UG-Cubic model with randomly varying linear + quadratic slopes (-2LL = 49810.7). This UG-model (referred to as Model D when determining the final GCM) required participants to have the same cubic growth rates, but allowed participants to vary in their linear and quadratic growth rates.

Conditional Growth (CG) Model Testing

Two sets of CG-models were tested in order to determine whether additional proportional variance reduction, relative to the best fitting UG-models, could be accounted for by adding time-invariant covariates and/or time-varying covariates. One set of models was a series of CG-models with Gender added as a time-invariant covariate at Level-2. The second set was a series of discontinuous CG-models with schooling added as a time-varying covariate at Level-1 and Gender added at Level-2.

CG-Models with Gender. Based on the UG-model testing results, the UG-Cubic model with randomly varying quadratic slopes served as the baseline CG-model for NDW-English. The CG-Cubic + Gender model was tested to determine the effect of Gender on initial status, and the interaction of Gender with the linear, quadratic, and cubic slopes. Table 25 outlines the comparisons of Gender *x* Slope interactions for the series of CG-Cubic + Gender models tested and highlights the best fitting CG-Cubic + Gender model for NDW-English.

Table 25. Comparisons of Gender *x* Slope Interactions for Conditional Cubic Growth Models with Gender (CG-Cb+G) of Number of Different Words-English[1]

	CG-Cubic + Gender models: Gender *x* Slope Interactions					
	Gender *x* Linear	Gender *x* Quadratic	Gender *x* Cubic	*df*	BIC	-2LL
CG-Cubic with Gender (*CG-Cb+G*)	+	+	+	15	49931.2	**49801.4**
	-	+	+	14	49927.7	49806.6
	+	-	+	14	49929.7	49808.6
	+	+	-	14	49930.7	49809.6
	-	-	+	13	49921.8	49809.3
	-	+	-	13	49922.3	49809.8
	+	-	-	13	49922.6	49810.1
	-	-	-	12	49914.2	49810.3

[1]*Note*. **Bold** indicates best fitting model; "-"indicates absent gender by slope interaction; "+" indicates gender by slope interaction; NC: Non-converging model

All CG-Cubic + Gender models converged. Comparisons of the eight CG-Cubic + Gender models based on -2LL revealed that the best fitting CG-model was the CG-Cubic + Gender model, with Gender *x* Linear + Gender *x* Quadratic + Gender *x* Cubic interactions (-2LL = 49801.4), which will be referred to as Model E when determining the final GCM.

Discontinuous CG-Models with Schooling and Gender. The second set of CG-models to be tested were discontinuous CG-models with schooling added as a time-varying covariate at Level-1 and Gender added as a time-invariant covariate at Level-2. A series of discontinuous CG-models with schooling *only* were tested with every possible combination of fixed and randomly varying slopes to determine the best fitting discontinuous CG-model first. The effect of Gender was then added to the best fitting discontinuous CG-model to determine whether additional proportional variance reduction (relative to the discontinuous CG-model without Gender) could be accounted for by adding Gender as a time-invariant covariate. Table 26 outlines the comparisons of fixed and randomly varying slopes for the series of discontinuous CG-models tested and highlights the best fitting discontinuous CG-model for NDW-English.

The best fitting discontinuous CG-model was the discontinuous CG-model with randomly varying linear slopes (-2LL = 46624.4), which will be referred to as Model F when determining the final GCM. This discontinuous CG-model required participants to have the same schooling growth rate, but allowed participants to vary in their linear growth rates.

Table 26. Comparisons of Fixed and Randomly Varying Slopes for Conditional Discontinuous Growth (Dc-CG) Models of Number of Different Words-English[1]

	Discontinuous CG-Models: Fixed and Randomly Varying Slopes				
	Linear-Random	Schooling-Random	*df*	BIC	-2LL
Discontinuous-CG (*Dc-CG*)	+	+	NC	NC	NC
	-	+	NC	NC	NC
	+	-	7	46684.4	**46624.4**
	-	-	5	46697.1	46654.2

[1]*Note*. **Bold** indicates best fitting model; "**-**" indicates fixed slope; "**+**" indicates randomly varying slope; NC: Non-converging model

The effect of Gender was then added to the best fitting discontinuous CG-model to determine whether additional proportional variance reduction (relative to the discontinuous CG-model with fixed linear + schooling slopes) in NDW-English could be accounted for by adding Gender as a time-invariant covariate. Gender (effect of being female) was tested with every possible interaction of Gender *x* Slope to determine the effect of Gender on the linear and schooling slopes. A total of four Discontinuous CG + Gender models were tested to identify the best fitting Discontinuous CG + Gender model. Table 27 outlines the comparisons of Gender *x* Slope interactions for the series of Discontinuous CG + Gender models tested and highlights the best fitting Discontinuous CG + Gender model for NDW-English.

Table 27. Comparisons of Gender *x* Slope Interactions for Conditional Discontinuous Growth Models with Gender (Dc-CG+G) of Number of Different Words-English[1]

	Discontinuous CG-Models: Gender *x* Slope Interactions				
	Gender *x* Linear	Gender *x* Schooling	*df*	BIC	-2LL
Discontinuous-CG with Gender (*Dc-CG+G*)	+	+	10	46694.4	**46608.65**
	-	+	9	46685.9	46608.68
	+	-	9	46701.2	46623.94
	-	-	8	46692.8	46624.13

[1]*Note*. **Bold** indicates best fitting models; "**-**"indicates absent gender by slope interaction; "+" indicates gender by slope interaction

All Discontinuous CG + Gender models converged. Comparisons of the four Discontinuous CG + Gender models based on -2LL (see Table 27) revealed that the best fitting discontinuous CG-model was the Discontinuous CG + Gender model, with Gender *x* Linear + Gender *x* Schooling interactions (-2LL = 46608.65), which will be referred to as Model G when determining the final GCM.

NDW-English: Final Growth Curve Model (GCM)

The parameter estimates from the UM-model (Model A), the best fitting UG-models (Models B, C, and D), and the two sets of best fitting CG-models (Models E, F, and G) were compared to determine the final GCM for NDW-English (see Table 28). Model A, the UM-model, was the baseline model of within-person residual variance at Level-1 (σ_ε^2 term) for all models. The two sets of best fitting CG-models each had a

Table 28. Comparisons of Growth Curve Model Parameter Estimates for Number of Different Words-English[1,2]

		Model A UM	*Model B UG-Ln*	*Model C UG-Qd*	*Model D UG-Cb*	*Model E CG-Cb+G*	*Model F CG-Dc*	*Model G CG-Dc+G*
Fixed effects								
Intercept	γ_{00}	74.37***	46.0***	49.52***	43.50***	44.01***	45.65***	**46.17*** **
Linear Slope	γ_{10}		9.75***	6.58***	24.01***	21.68***	19.05***	**19.11*** **
Schooling Slope	γ_{20}			0.54*** Quad	-8.13*** Quad	-6.88*** Quad	11.18***	**9.52*** **
Cubic Slope	γ_{30}				1.11***	0.94***		
Gender (G)	γ_{01}					-.98		**-1.02**
G x Linear Slope	γ_{11}					4.54*		**0.12**
G x Schooling Slope	γ_{21}					-2.43** G x Quad		**3.25***
G x Cubic Slope	γ_{31}					0.33**		
Variance components								
L1: Within-person variance	σ_ε^2	482.93*	204.28*	196.28*	173.80*	173.34*	209.06*	**207.97***
L2: B/w-person intercept	σ_0^2	280.83*	575.57*	399.54*	492.72*	493.53*	484.39*	**484.94***
L2: B/w-person linear slope	σ_1^2		6.23*	35.43*	33.29*	33.43*	17.96*	**18.25***
L2: B/w-person quadratic slope	σ_2^2			1.14*	1.01*	1.02*		

Table 28. (continued)

		Model A UM	*Model B UG-Ln*	*Model C UG-Qd*	*Model D UG-Cb*	*Model E CG-Cb+G*	*Model F CG-Dc*	*Model G CG-Dc+G*
Covariance (σ_0^2, σ_1^2)	σ_{01}		-36.10*	31.77*	-1.51	-1.81	-45.95*	**-46.17***
Covariance (σ_0^2, σ_2^2)	σ_{02}			-11.23*	-7.06*	-7.04*		
Covariance (σ_1^2, σ_2^2)	σ_{12}			-5.76*	-4.98*	-5.00*		
Proportional variance reduction								
L1: Within-person variance	R_ε^2		58%	59%	64%	64%	57%	**57%**
L2: B/w-person intercept	R_0^2					<1%		**<1%**
L2: B/w-person linear slope	R_1^2					<1%		**-2%**
L2: B/w-person quadratic slope	R_2^2					-20%		
Goodness-of-fit								
-2LL		53386.4	50227.9*	50152.5*	49810.7*	49801.4~	46624.4	**46608.7***
BIC		53412.3	50279.8	50239.1	49905.9	49931.2	46684.4	**46694.4**

~$p < .10$, *$p < .05$, **$p < .01$, ***$p < .001$

[1]*Note.* **Bold** indicates final model; Shading indicates non-applicable Pseudo-R^2

[2]*Key*. UM: Unconditional means model; UG-Ln: Unconditional linear growth model; UG-Qd: Unconditional quadratic growth model; UG-Cb: Unconditional cubic growth model; CG-Cb+G: Conditional cubic growth model with gender; CG-Dc: Conditional discontinuous growth model; CG-Dc+G: Conditional discontinuous growth model with gender; L1: Level-1 submodel; L2: Level-2 submodel

specific baseline model of between-person intercept (σ_0^2 term) and slope (σ_1^2 term; σ_2^2 term) residual variances at Level-2. Model D served as the baseline model for Model E, and Model F served as the baseline model (of Level-2 residual variances) for Model G.

Table 28 outlines the following parameter estimates for the growth curve models tested: (1) fixed effects (i.e., intercepts, slopes, covariates, and interactions); (2) variance components (i.e., Level-1 within-person residual variance, Level-2 between-person residual variances, and covariances); (3) proportional variance reduction (i.e., Level-1 and Level-2 Pseudo-R^2s), and (4) goodness-of-fit indices for nested (-2LL) and non-nested (BIC) models. Models A-D were nested within Model E, and Model F was nested within Model G. Models E and G were non-nested models. If fixed effects were significant, they were included in subsequent nested models. The final GCM was determined by comparisons of each model's proportional variance reduction (Pseudo-R_ε^2; Pseudo-R_0^2; Pseudo-R_1^2; and Pseudo-R_2^2) and goodness-of-fit, with the final model having the highest proportional variance reduction and the lowest goodness-of-fit indices.

As can be seen in Table 28, Models D and E accounted for the greatest proportional variance reduction (64%) of within-person residual variance at Level-1, relative to Model A. However, Model E did not account for additional proportional variance reduction in between-person intercept (Pseudo-R_0^2) or linear slope (Pseudo-R_1^2) residual variance relative to Model D. In fact, Model E accounted for a 20% proportional variance reduction deterioration in between-person quadratic slope (Pseudo-R_2^2) residual variance relative to Model D. The two discontinuous models (Models F and G)

accounted for a 57% proportional variance reduction of within-person residual variance at Level-1, relative to Model A. Model G was relative to Model F. Model G did not account for additional proportional variance reduction in between-person intercept (Pseudo-R_0^2) residual variance relative to Model F. Further, Model G accounted for a 2% proportional variance reduction deterioration in between-person linear slope (Pseudo-R_1^2) residual variance relative to Model F.

Goodness-of-fit indices (-2LL and BIC) indicated the degree of fit (correspondence) that the growth curve models had to the sample data. The final GCM had the smallest goodness-of-fit indices, relative to all other growth curve models. The -2LL difference between Model B relative to Model A (3158.5) exceeded the 0.0005 critical level (17.73) for a χ^2 distribution on 3 *df*, indicating that the effect of adding linear time to Model B provided a significantly better fit than Model A to the sample data. The -2LL difference between Model C relative to Model B (75.4) exceeded the 0.0005 critical level (20.0) for a χ^2 distribution on 4 *df*, indicating that the effect of adding quadratic time to Model C provided a significantly better fit than Model B to the sample data. The -2LL difference between Model D relative to Model C (341.8) exceeded the 0.0005 critical level (12.12) for a χ^2 distribution on 1*df*, indicating that the effect of adding cubic time to Model D provided a significantly better fit than Model C to the sample data. The -2LL difference between Model E relative to Model D (9.3) did not exceed the 0.05 critical level (9.49) for a χ^2 distribution on 4 *df*, indicating that the effect of adding Gender to Model E did not provide a significantly better fit than Model D to the sample data.

Therefore, Model D provided the best fit to the sample data, relative to its nested models (Models A-E).

Similarly, -2LL was also used to compare the fit of Model G relative to Model F. The -2LL difference between Model G relative to Model F (15.7) exceeded the 0.005 critical level (12.84) for a χ^2 distribution on 3 *df*, indicating that the effect of adding Gender to Model G provided a significantly better fit than Model F to the sample data. Therefore, Model G provided the best fit to the sample data, relative to its nested model (Model F).

BIC was used to compare the fit of Model D against the fit of Model G, which were the non-nested models. The χ^2 distribution test was not used to compare BICs, only their absolute values (Singer & Willett, 2003). The BIC of Model G (46694.4) was smaller than the BIC of Model D (49905.9) by 3211.5, indicating that Model G provided the best fit to the sample data, relative to its non-nested model (Model D).

In summary, the results of these analyses indicated that Model G (Discontinuous CG + Gender model, with Gender *x* Linear + Gender *x* Schooling interactions) was the final model for NDW-English, which estimated the effect that Gender had on the initial status of growth and on the linear growth rates – including schooling. Although Model G did not have the highest overall proportional variance reduction (Pseudo-R^2s), it had the lowest goodness-of-fit indices (-2LL = 46608.7; BIC = 46694.4) for NDW-English (see Table 28), which account for all the fixed effects and variance components. The GCM formula for Model G-(NDW-English) was as follows:

Level-1: Within-person variation

$$\text{NDW-English}_{ij} = \pi_{0i} + \pi_{1i}\,(\text{TIME}_{ij}) + \pi_{2i}\,(\text{SCHOOLING}_{ij}) + \varepsilon_{ij}$$

Level-2: Between-person variation

$$\pi_{0i} = \gamma_{00} + \gamma_{01}\,(\text{Gender}_i) + \zeta_{0i}$$

$$\pi_{1i} = \gamma_{10} + \gamma_{11}\,(\text{Gender}_i) + \zeta_{1i}$$

$$\pi_{2i} = \gamma_{20} + \gamma_{21}\,(\text{Gender}_i)$$

The fixed effects of this model indicated that boys' expected average initial status of NDW-English was $\gamma_{00} = 46.17$, $p < .001$, with a positive and significant linear growth rate ($\gamma_{10} = 19.11$, $p < .001$) during each semester, and an additional positive and significant linear (schooling) growth rate ($\gamma_{20} = 9.52$, $p < .001$) during each spring semester. There was an additional positive and significant effect of Gender (being female) on the linear (schooling) growth rate ($\gamma_{21} = 3.25$, $p < .05$) during each spring semester, relative to boys. The variance components of this model indicated that the participants differed significantly from each other over time ($\sigma_\varepsilon^2 = 207.97$, $p < .05$), and that there were significant individual differences in initial status ($\sigma_0^2 = 484.94$, $p < .05$) and in the linear slope ($\sigma_1^2 = 18.25$, $p < .05$) after controlling for Gender and Schooling. The covariance components of this model indicated that there were negative and significant individual differences in the initial status, linear slope covariance ($\sigma_{01} = -46.17$, $p < .05$) after controlling for Gender and Schooling.

NDW-English: Prototypical Growth Trajectory

The fixed effects parameter estimates form the final model (Model G) were used to generate the prototypical growth trajectories by gender for NDW-English. The

prototypical growth trajectory, illustrated in Figure 7, shows that the shapes of growth for boys and girls are linear, non-monotonic, and discontinuous. Figure 7 also demonstrates a number of interesting data observations: (1) initial status (non-significant) differences with boys beginning with higher NDW-English (46.17) relative to girls (45.15) during the fall of kindergarten; (2) periods of constant, linear NDW-English growth for both genders (boys = 9.52; girls = 12.77) during each academic year (kindergarten, first grade, and second grade) with girls' linear growth outpacing boys; (3) the NDW-English linear growth rate was consistent for both genders between the spring semester of Kindergarten and the fall semester of first grade, and again between the spring semester of first grade and the fall semester of second grade (boys = 9.59; girls = 6.22) with boys' linear growth during the summer outpacing girls; and (4) girls demonstrated higher NDW-English (95.91) relative to boys (93.91) during the last wave of data collection, in the spring of second grade.

Words Per Minute (WPM) – English

WPM was used as a measure of general verbal fluency to represent linguistic subsystem integration in English. GCM analyses were conducted to determine within- and between-person variation in initial status and growth rates of WPM-English from the fall of kindergarten to the spring of second grade.

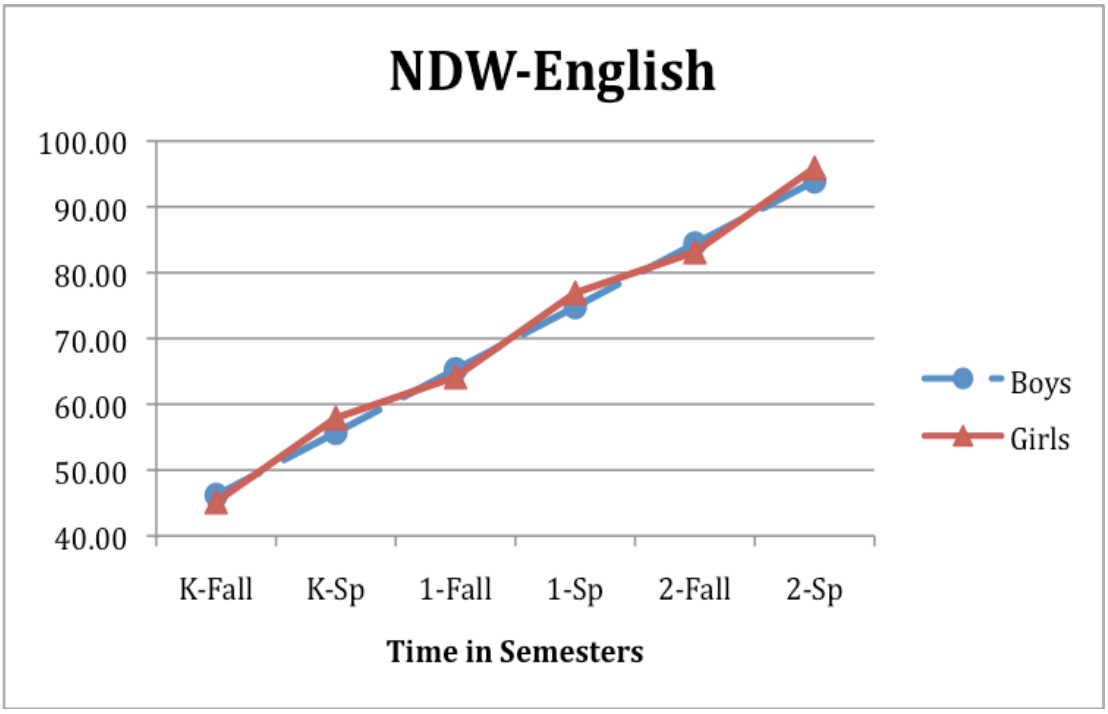


Figure 7. Prototypical Growth Trajectories by Gender for Number of Different Words-English

Unconditional Means (UM) Model

The UM-model for WPM-English resulted in significant within-person residual variability to be explained at Level-1 ($\sigma_\varepsilon^2 = 365.57$, $p < .05$), significant between-person residual variability to be explained at Level-2 ($\sigma_0^2 = 325.11$, $p < .05$), and an *ICC* of .47. Without considering the effect of time or any other predictors, 47% of the total variation in WPM-English was accounted for by differences between the participants in the sample. The UM-model will be referred to as Model A when determining the final growth curve model (GCM) for WPM-English.

Unconditional Growth (UG) Model Testing

Testing a series of nested UG-models determined whether additional proportional variance reduction, relative to the UM-model, in WPM-English could be accounted for by adding time as a time-varying covariate at Level-1. Three nested UG-models were tested for model fit with every possible combination of fixed and randomly varying slopes. Table 29 outlines the comparisons of fixed and randomly slopes for the series of UG-models tested and highlights the best fitting (i.e., smallest -2LL) UG-Linear, UG-Quadratic, and UG-Cubic models for WPM-English.

As can be seen in Table 29, the best fitting UG-Linear model was the UG-Linear model with randomly varying linear slopes (-2LL = 49994.3), which will be referred to as Model B when determining the final GCM. The best fitting UG-Quadratic model was the UG-Quadratic model with randomly varying linear + quadratic slopes (-2LL = 49955.5), which will be referred to as Model C when determining the final GCM. The best fitting UG-Cubic model was the UG-Cubic model with varying linear + quadratic slopes (-2LL = 49944.9), which will be referred to as Model D when determining the final GCM. The UG-Cubic model with varying linear + quadratic + cubic slopes did not converge.

Comparisons of the best fitting UG-Linear, UG-Quadratic, and UG-Cubic models based on -2LL (see Table 29) revealed that the best fitting UG-model was the UG-Cubic model with varying linear + quadratic slopes (-2LL = 49944.9). This UG-model (referred to as Model D when determining the final GCM) required participants to have the same cubic growth rates, but allowed them to vary in their linear and quadratic growth rates.

Table 29. Comparisons of Fixed and Randomly Varying Slopes for Unconditional Growth (UG) Models of Words per Minute-English[1]

	UG-Models: Fixed and Randomly Varying Slopes					
	Linear-Random	Quadratic-Random	Cubic-Random	*df*	BIC	-2LL
UG-Linear (*UG-Ln*)	+			6	50046.2	**49994.3**
	-			4	50175.2	50140.5
UG-Quadratic (*UG-Qd*)	+	+		10	50042.0	**49955.5**
	-	+		7	50088.2	50027.6
	+	-		7	50053.0	49992.4
	-	-		5	50173.2	50129.9
UG-Cubic (*UG-Cb*)	+	+	+	NC	NC	NC
	-	+	+	11	50042.2	49947.0
	+	-	+	11	50043.4	49948.2
	+	+	-	11	50040.1	**49944.9**
	-	-	+	8	50115.9	50046.6
	-	+	-	8	50087.4	50018.1
	+	-	-	8	50054.9	49985.6
	-	-	-	6	50171.9	50120.0

[1]*Note*. **Bold** indicates best fitting models; Shading indicates non-applicable randomly varying slopes; "-" indicates fixed slope; "+" indicates randomly varying slope; NC: Non-converging model

Conditional Growth (CG) Model Testing

Two sets of CG-models were tested in order to determine whether additional proportional variance reduction, relative to the best fitting UG-models, could be accounted for by adding time-invariant covariates and/or time-varying covariates. One set of models was a series of CG-models with Gender added as a time-invariant covariate at Level-2. The second set was a series of discontinuous CG-models with schooling added as a time-varying covariate at Level-1 and Gender added at Level-2.

CG-Models with Gender. Based on the UG-model testing results, the UG-Cubic model with randomly varying linear + quadratic slopes served as the baseline CG-model for WPM-English. The CG-Cubic + Gender model was tested to determine the effect of Gender on initial status, and the interaction of Gender with the linear, quadratic, and cubic slopes. Table 30 outlines the comparisons of Gender *x* Slope interactions for the series of CG-Cubic + Gender models tested and highlights the best fitting CG-Cubic + Gender model for WPM-English.

All CG-Cubic + Gender models converged. Comparisons of the eight CG-Cubic + Gender models based on -2LL revealed that the best fitting CG-model was the CG-Cubic + Gender model, with Gender *x* Linear + Gender *x* Quadratic + Gender *x* Cubic interactions (-2LL = 49924.2), which will be referred to as Model E when determining the final GCM.

Table 30. Comparisons of Gender *x* Slope Interactions for Conditional Cubic Growth Models with Gender (CG-Cb+G) of Words per Minute-English[1]

	CG-Cubic + Gender models: Gender *x* Slope Interactions					
	Gender *x* Linear	Gender *x* Quadratic	Gender *x* Cubic	*df*	BIC	-2LL
CG-Cubic with Gender (*CG-Cb+G*)	+	+	+	15	50054.0	**49924.2**
	-	+	+	14	50046.8	49925.7
	+	-	+	14	50046.7	49925.6
	+	+	-	14	50046.3	49925.2
	-	-	+	13	50038.2	49925.7
	-	+	-	13	50038.2	49925.7
	+	-	-	13	50038.9	49926.4
	-	-	-	12	50031.2	49927.4

[1]*Note*. **Bold** indicates best fitting model; "**-**"indicates absent gender by slope interaction; "+" indicates gender by slope interaction; NC: Non-converging model

Discontinuous CG-Models with Schooling and Gender. The second set of CG-models to be tested were discontinuous CG-models with schooling added as a time-varying covariate at Level-1 and Gender added as a time-invariant covariate at Level-2. A series of discontinuous CG-models with schooling *only* were tested with every possible combination of fixed and randomly varying slopes to determine the best fitting discontinuous CG-model first. The effect of Gender was then added to the best fitting discontinuous CG-model to determine whether additional proportional variance reduction

(relative to the discontinuous CG-model without Gender) could be accounted for by adding Gender as a time-invariant covariate. Table 31 outlines the comparisons of fixed and randomly varying slopes for the series of discontinuous CG-models tested and highlights the best fitting discontinuous CG-model for WPM-English.

Table 31. Comparisons of Fixed and Randomly Varying Slopes for Conditional Discontinuous Growth (Dc-CG) Models of Words per Minute-English[1]

	Discontinuous CG-Models: Fixed and Randomly Varying Slopes				
	Linear-Random	Schooling-Random	*df*	BIC	-2LL
Discontinuous-CG (*Dc-CG*)	+	+	NC	NC	NC
	-	+	NC	NC	NC
	+	-	7	46442.7	**46382.6**
	-	-	5	46531.8	46488.9

[1]*Note*. **Bold** indicates best fitting model; "-" indicates fixed slope; "+" indicates randomly varying slope; NC: Non-converging model

The best fitting discontinuous CG-model was the discontinuous CG-model with randomly varying linear slopes (-2LL = 46382.6), which will be referred to as Model F when determining the final GCM. This discontinuous CG-model required participants to have the same schooling growth rate, but allowed them to vary in their linear growth rates.

The effect of Gender was then added to the best fitting discontinuous CG-model to determine whether additional proportional variance reduction (relative to the discontinuous CG-model with fixed linear + schooling slopes) in WPM-English could be accounted for by adding Gender as a time-invariant covariate. Gender (effect of being female) was tested with every possible interaction of Gender *x* Slope to determine the effect of Gender on the linear and schooling slopes. A total of four Discontinuous CG + Gender models were tested to identify the best fitting Discontinuous CG + Gender model. Table 32 outlines the comparisons of Gender *x* Slope interactions for the series of Discontinuous CG + Gender models tested and highlights the best fitting Discontinuous CG + Gender model for WPM-English.

Table 32. Comparisons of Gender *x* Slope Interactions for Conditional Discontinuous Growth Models with Gender (Dc-CG+G) of Words per Minute-English[1]

	Discontinuous CG-Models: Gender *x* Slope Interactions				
	Gender *x* Linear	Gender *x* Schooling	*df*	BIC	-2LL
Discontinuous-CG with Gender (*Dc-CG+G*)	+	+	10	46445.2	**46359.4**
	-	+	9	46438.4	46361.2
	+	-	9	46439.6	46362.4
	-	-	8	46433.1	46364.5

[1]*Note*. **Bold** indicates best fitting models; "**-**"indicates absent gender by slope interaction; "+" indicates gender by slope interaction

All Discontinuous CG + Gender models converged. Comparisons of the four Discontinuous CG + Gender models based on -2LL revealed that the best fitting discontinuous CG-model was the Discontinuous CG + Gender model, with Gender *x* Linear + Gender *x* Schooling interactions (-2LL = 46359.4), which will be referred to as Model G when determining the final GCM.

WPM-English: Final Growth Curve Model (GCM)

The parameter estimates from the UM-model (Model A), the best fitting UG-models (Models B, C, and D), and the two sets of best fitting CG-models (Models E, F, and G) were compared to determine the final GCM for WPM-English (see Table 33). Model A, the UM-model, was the baseline model of within-person residual variance at Level-1 (σ_ε^2 term) for all models. The two sets of best fitting CG-models each had a specific baseline model of between-person intercept (σ_0^2 term) and slope (σ_1^2 term; σ_2^2 term) residual variances at Level-2. Model D served as the baseline model for Model E, and Model F served as the baseline model (of Level-2 residual variances) for Model G.

Table 33 outlines the following parameter estimates for the growth curve models tested: (1) fixed effects (i.e., intercepts, slopes, covariates, and interactions); (2) variance components (i.e., Level-1 within-person residual variance, Level-2 between-person residual variances, and covariances); (3) proportional variance reduction (i.e., Level-1 and Level-2 Pseudo-R^2s), and (4) goodness-of-fit indices for nested (-2LL) and non-nested (BIC) models. Models A-D were nested within Model E, and Model F was nested within Model G. Models E and G were non-nested models. If fixed effects were

Table 33. Comparisons of Growth Curve Model Parameter Estimates for Words per Minute-English[1,2]

		Model A UM	*Model B UG-Ln*	*Model C UG-Qd*	*Model D UG-Cb*	*Model E CG-Cb+G*	*Model F CG-Dc*	*Model G CG-Dc+G*
Fixed effects								
Intercept	γ_{00}	76.96***	55.02***	56.69***	57.82***	60.44***	56.35***	**58.53*****
Linear Slope	γ_{10}		7.52***	6.07***	3.01**	1.67	15.32***	**15.81*****
Schooling Slope	γ_{20}			0.24** Quad	1.74*** Quad	2.31*** Quad	5.05***	**4.36*****
Cubic Slope	γ_{30}				-0.19**	-0.25**		
Gender (G)	γ_{01}					-5.11**		**-4.37****
G x Linear Slope	γ_{11}					2.50		**0.91**
G x Schooling Slope	γ_{21}					-1.06 G x Quad		**1.35~**
G x Cubic Slope	γ_{31}					0.11		
Variance components								
L1: Within-person variance	σ_ε^2	365.57*	181.92*	173.81*	173.34*	173.34*	179.74*	**179.62***
L2: B/w-person intercept	σ_0^2	325.11*	637.62*	594.51*	577.54*	572.93*	578.06*	**575.01***
L2: B/w-person linear slope	σ_1^2		11.12*	58.84*	60.15*	59.47*	41.49*	**41.52***
L2: B/w-person quadratic slope	σ_2^2			1.22*	1.30*	1.27*		

Table 33. (continued)

		Model A UM	*Model B UG-Ln*	*Model C UG-Qd*	*Model D UG-Cb*	*Model E CG-Cb+G*	*Model F CG-Dc*	*Model G CG-Dc+G*
Covariance (σ_0^2, σ_1^2)	σ_{01}		-51.46*	-59.62*	-53.18*	-52.10*	-83.54*	**-85.00***
Covariance (σ_0^2, σ_2^2)	σ_{02}			1.21	0.34	0.01		
Covariance (σ_1^2, σ_2^2)	σ_{12}			-7.71*	-8.06*	-7.93*		
Proportional variance reduction								
L1: Within-person variance	R_ε^2		50%	52%	53%	53%	51%	**51%**
L2: B/w-person intercept	R_0^2					1%		**1%**
L2: B/w-person linear slope	R_1^2					1%		**<1%**
L2: B/w-person quadratic slope	R_2^2					2%		
Goodness-of-fit								
-2LL		52244.6	49994.3*	49955.5*	49944.9*	49924.2*	46382.6	**46359.4***
BIC		52270.6	50046.2	50042.0	50040.1	50054.0	46442.7	**46445.2**

~$p < .10$, *$p < .05$, **$p < .01$, ***$p < .001$

[1]*Note.* **Bold** indicates final model; Shading indicates non-applicable Pseudo-R^2

[2]*Key*. UM: Unconditional means model; UG-Ln: Unconditional linear growth model; UG-Qd: Unconditional quadratic growth model; UG-Cb: Unconditional cubic growth model; CG-Cb+G: Conditional cubic growth model with gender; CG-Dc: Conditional discontinuous growth model; CG-Dc+G: Conditional discontinuous growth model with gender; L1: Level-1 submodel; L2: Level-2 submodel

significant, they were included in subsequent nested models. The final GCM was determined by comparisons of each model's proportional variance reduction (Pseudo-R_ε^2; Pseudo-R_0^2; Pseudo-R_1^2; and Pseudo-R_2^2) and goodness-of-fit, with the final model having the highest proportional variance reduction and the lowest goodness-of-fit indices.

As can be seen in Table 33, Models D and E accounted for the greatest proportional variance reduction (53%) of within-person residual variance at Level-1, relative to Model A. However, relative to Model D, Model E also accounted for the following proportional variance reductions: 1% in between-person intercept (Pseudo-R_0^2) residual variance, 1% in between-person linear slope (Pseudo-R_1^2) residual variance, and 2% in between-person quadratic slope (Pseudo-R_2^2) residual variance. The two discontinuous models (Models F and G) accounted for a 51% proportional variance reduction of within-person residual variance at Level-1, relative to Model A. Model G further accounted for a 1% proportional variance reduction in between-person intercept (Pseudo-R_0^2) residual variance, but did not improve the proportional variance reduction in between-person linear slope (Pseudo-R_1^2) residual variance relative to Model F.

Goodness-of-fit indices (-2LL and BIC) indicated the degree of fit (correspondence) that the growth curve models had to the sample data. The final GCM had the smallest goodness-of-fit indices, relative to all other growth curve models. The -2LL difference between Model B relative to Model A (2250.3) exceeded the 0.0005 critical level (17.73) for a χ^2 distribution on 3 *df*, indicating that the effect of adding linear time to Model B provided a significantly better fit than Model A to the sample data. The

-2LL difference between Model C relative to Model B (38.8) exceeded the 0.0005 critical level (20.0) for a χ^2 distribution on 4 *df*, indicating that the effect of adding quadratic time to Model C provided a significantly better fit than Model B to the sample data. The -2LL difference between Model D relative to Model C (10.6) exceeded the 0.005 critical level (7.88) for a χ^2 distribution on 1 *df*, indicating that the effect of adding cubic time to Model D provided a significantly better fit than Model C to the sample data. The -2LL difference between Model E relative to Model D (20.7) exceeded the 0.0005 critical level (20.0) for a χ^2 distribution on 4 *df*, indicating that the effect of adding Gender to Model E provided a significantly better fit than Model D to the sample data. Therefore, Model E provided the best fit to the sample data, relative to its nested models (Models A-D).

Similarly, -2LL was also used to compare the fit of Model G relative to Model F. The -2LL difference between Model G relative to Model F (23.2) exceeded the 0.0005 critical level (17.73) for a χ^2 distribution on 3 *df*, indicating that the effect of adding Gender to Model G provided a significantly better fit than Model F to the sample data. Therefore, Model G provided the best fit to the sample data, relative to its nested model (Model F).

BIC was used to compare the fit of Model E against the fit of Model G, which were the non-nested models. The χ^2 distribution test was not used to compare BICs, only their absolute values (Singer & Willett, 2003). The BIC of Model G (46445.2) was smaller than the BIC of Model E (50054.0) by 3608.8, indicating that Model G provided the best fit to the sample data, relative to its non-nested model (Model E).

In summary, the results of these analyses indicated that Model G (Discontinuous CG + Gender model, with Gender *x* Linear + Gender *x* Schooling interactions) was the final model for WPM-English, which estimated the effect that Gender had on the initial status of growth and on the linear growth rates – including schooling. Although Model G did not have the highest overall proportional variance reduction (Pseudo-R^2s), it had the lowest goodness-of-fit indices (-2LL = 46359.4; BIC = 46445.2) for WPM-English, which account for all the fixed effects and variance components. The GCM formula for Model G-(WPM-English) was as follows:

Level-1: Within-person variation

$$\text{WPM-English}_{ij} = \pi_{0i} + \pi_{1i}\,(\text{TIME}_{ij}) + \pi_{2i}\,(\text{SCHOOLING}_{ij}) + \varepsilon_{ij}$$

Level-2: Between-person variation

$$\pi_{0i} = \gamma_{00} + \gamma_{01}\,(\text{Gender}_i) + \zeta_{0i}$$

$$\pi_{1i} = \gamma_{10} + \gamma_{11}\,(\text{Gender}_i) + \zeta_{1i}$$

$$\pi_{2i} = \gamma_{20} + \gamma_{21}\,(\text{Gender}_i)$$

The fixed effects of this model indicated that boys' expected average initial status of WPM-English was $\gamma_{00} = 58.53$, $p < .001$, with a positive and significant linear growth rate ($\gamma_{10} = 15.81$, $p < .001$) during each semester, and an additional positive and significant linear (schooling) growth rate ($\gamma_{20} = 4.36$, $p < .001$) during each spring semester. There was a negative and significant effect of Gender (being female) on initial status ($\gamma_{01} = -4.37$, $p < .001$), and an additional positive effect of Gender (that approached significance) on the linear (schooling) growth rate ($\gamma_{21} = 1.35$, $p = .08$) during each spring semester, relative to boys. The variance components of this model indicated that the participants

differed significantly from each other over time ($\sigma_\varepsilon^2 = 179.62$, $p < .05$), and that there were significant individual differences in initial status ($\sigma_0^2 = 575.01$, $p < .05$) and in the linear slope ($\sigma_1^2 = 41.52$, $p < .05$) after controlling for Gender and Schooling. The covariance components of this model indicated that there were negative and significant individual differences in the initial status, linear slope covariance ($\sigma_{01} = -85.00$, $p < .05$) after controlling for Gender and Schooling.

WPM-English: Prototypical Growth Trajectory

The fixed effects parameter estimates form the final model (Model G) were used to generate the prototypical growth trajectories by gender for WPM-English. The prototypical growth trajectory, illustrated in Figure 8, shows that the shapes of growth for boys and girls are linear, non-monotonic, and discontinuous. Figure 8 also demonstrates a number of interesting data observations: (1) initial status differences with boys beginning with significantly higher WPM-English (58.53) relative to girls (54.16) during the fall of kindergarten; (2) periods of constant, linear WPM-English growth for both genders (boys = 4.35; girls = 5.71) during each academic year (kindergarten, first grade, and second grade) with girls' linear growth outpacing boys; (3) both genders demonstrated an increase in the WPM-English linear growth rate between the spring semester of Kindergarten and the fall semester of first grade, and again between the spring semester of first grade and the fall semester of second grade (boys = 11.46; girls = 9.19) with boys' linear growth during the summer outpacing girls; and (4) boys

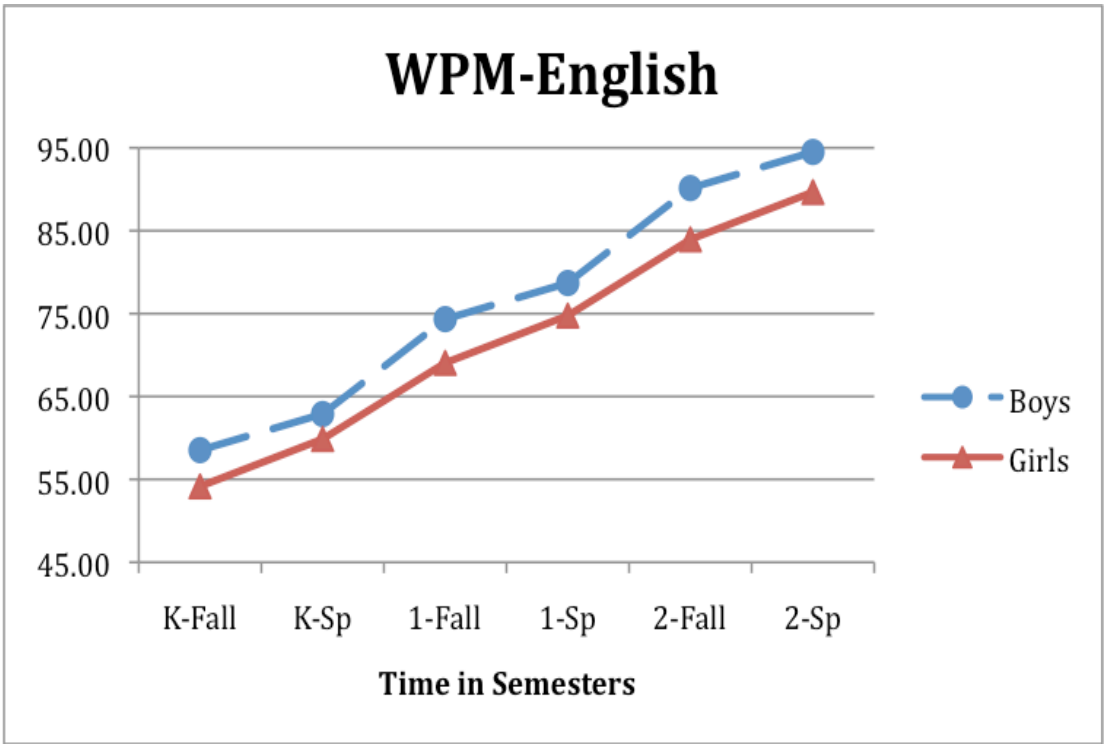


Figure 8. Prototypical Growth Trajectories by Gender for Words per Minute-English

demonstrated higher WPM-English (94.50) relative to girls (89.66) during the last wave of data collection, in the spring of second grade.

Summary

The present study profiled the longitudinal growth of ELLs' oral language skills in Spanish and English over time by testing a series of growth curve models (GCMs) for each outcome variable (MLUw, NDW, and WPM). GCM testing resulted in: examining whether the covariates academic semester, gender, and schooling optimized model fit and predicted growth over time; identification of significant within-person (intra-individual;

Level-1) and between-person (inter-individual; Level-2) variation in initial status (i.e., intercepts, the starting point of participants' growth) and participants' growth rates (i.e., slopes); and determining complex, final growth curve models that provided the most representative growth curve trajectories for each outcome measure. The results of GCM testing for MLUw, NDW, and WPM in Spanish and English are summarized in Table 34. Specifically, the direction and significance of the parameter estimates for each final GCM are outlined, as well as the non-significant parameter estimates.

Growth curve model testing of MLUw, NDW, and WPM in Spanish resulted the following commonalities: (1) the CG-Cubic + Gender model, with Gender *x* Linear + Gender *x* Quadratic + Gender *x* Cubic interactions (Model E) was the best fitting model for all outcome measures, indicating curvilinear growth over time of Spanish; (2) Model E had the highest overall proportional variance reduction of within-person (Pseudo-R_{ε}^2) residual variance and of between-person intercept (Pseudo-R_0^2) residual variance and the lowest goodness-of-fit indices (-2LL for nested models; BIC for non-nested models) for all outcome measures; (3) the fixed effects from the final GCM (Model E) for each outcome measure indicated that boys and girls had a positive and significant instantaneous linear rate of change (γ_{10}; γ_{11}) during the fall of kindergarten, that their expected rate of deceleration in the linear slope over time (γ_{20}; γ_{21}) was negative and significant, and their expected rate of acceleration in the quadratic slope over time (γ_{30}; γ_{31}) was positive and significant; (4) for all outcome measures, there was a positive and significant effect of Gender (being female) on initial status (γ_{01}); and (5) the variance

Table 34. Final Growth Curve Model (GCM) Parameter Estimates for MLUw, NDW, and WPM in Spanish and English[1,2]

	Spanish GCM: CG-Cb+G				*English GCM: CG-Dc+G*			
	MLUw	*NDW*	*WPM*		*MLUw*	*NDW*	*WPM*	
Fixed effects								*Fixed effects*
Intercept	+✓***	+✓***	+✓***	γ_{00}	+✓***	+✓***	+✓***	Intercept
Linear Slope	+✓***	+✓***	+✓***	γ_{10}	+✓***	+✓***	+✓***	Linear Slope
Quadratic Slope	-✓***	-✓***	-✓***	γ_{20}	+✓***	+✓***	+✓***	Schooling Slope
Cubic Slope	+✓***	+✓***	+✓***	γ_{30}				
Gender (G)	+✓***	+✓***	+✓~	γ_{01}	×	×	-✓**	Gender (G)
G x Linear Slope	+✓*	+✓**	+✓*	γ_{11}	×	×	×	G x Linear Slope
G x Quadratic Slope	-✓**	-✓**	-✓**	γ_{21}	+✓*	+✓*	+✓~	G x Schooling Slope
G x Cubic Slope	+✓**	+✓**	+✓**	γ_{31}				
Variance components								*Variance components*
L1: Within-person variance	+✓*	+✓*	+✓*	σ_ε^2	+✓*	+✓*	+✓*	L1: Within-person variance
L2: B/w-person intercept	+✓*	+✓*	+✓*	σ_0^2	+✓*	+✓*	+✓*	L2: B/w-person intercept
L2: B/w-person linear slope			+✓*	σ_1^2	+✓*	+✓*	+✓*	L2: B/w-person linear slope
L2: B/w-person quadratic slope	×	×	+✓*	σ_2^2				
L2: B/w-person cubic slope	×		+✓*	σ_3^2				

Table 34. (continued)

	Spanish GCM: CG-Cb+G				*English GCM: CG-Dc+G*			
	MLUw	*NDW*	*WPM*		*MLUw*	*NDW*	*WPM*	
Covariance (σ_0^2, σ_1^2)			×	σ_{01}	-✓*	-✓*	-✓*	Covariance (σ_0^2, σ_1^2)
Covariance (σ_0^2, σ_2^2)	×	×	×	σ_{02}				
Covariance (σ_0^2, σ_3^2)	+✓*		×	σ_{03}				
Covariance (σ_1^2, σ_2^2)			-✓*	σ_{12}				
Covariance (σ_1^2, σ_3^2)			+✓*	σ_{13}				
Covariance (σ_2^2, σ_3^2)	×		×	σ_{23}				
Proportional variance reduction								*Proportional variance reduction*
L1: Within-person variance	48%	38%	31%	R_ε^2	50%	57%	51%	L1: Within-person variance
L2: B/w-person intercept	6%	5%	1%	R_0^2	0%	0%	1%	L2: B/w-person intercept
L2: B/w-person linear slope			2%	R_1^2	0%	-2%	0%	L2: B/w-person linear slope
L2: B/w-person quadratic slope	0%	0%	2%	R_2^2				
L2: B/w-person cubic slope	50%		2%	R_3^2				

Table 34. (continued)

	Spanish GCM: CG-Cb+G			*English GCM: CG-Dc+G*			
	MLUw	*NDW*	*WPM*	*MLUw*	*NDW*	*WPM*	
Goodness-of-fit							*Goodness-of-fit*
-2LL	+✓*	+✓*	+✓*	+✓*	+✓*	+✓*	-2LL

~$p < .10$, *$p < .05$, **$p < .01$, ***$p < .001$

[1]*Note.* +✓ indicates a positive and significant parameter estimate; -✓ indicates a negative and significant parameter estimate; × indicates a non-significant parameter estimate; Shading indicates non-applicable Pseudo-R^2

[2]*Key*. CG-Cb+G: Conditional cubic growth model with gender; CG-Dc+G: Conditional discontinuous growth model with gender; MLUw: Mean length of utterance in words; NDW: Number of different words; WPM: Words per minute; L1: Level-1 submodel; L2: Level-2 submodel

components from the final model (Model E) for each outcome measure indicated that the components from the final model (Model E) for each outcome measure indicated that the participants differed significantly from each other over time (σ_ε^2), and that there were significant individual differences in initial status (σ_0^2) after controlling for Gender.

Growth curve model testing of MLUw, NDW, and WPM in English resulted the following commonalities: (1) the Discontinuous CG + Gender model, with Gender *x* Linear + Gender *x* Schooling interactions (Model G) was the best fitting model for all outcome measures, indicating discontinuous growth over time of English; (2) although Model G did not have the highest overall proportional variance reduction (Pseudo-R^2s), it

had the lowest goodness-of-fit indices (-2LL = 15237.8; BIC = 15323.6), which accounted for the fixed effects and variance components for all outcome measures; (3) the fixed effects from the final GCM (Model G) for each outcome measure indicated that boys had a positive and significant linear growth rate (γ_{10}) during each semester, and an additional positive and significant linear (schooling) growth rate (γ_{20}) during each spring semester; (4) for MLUw and NDW, there was an additional positive and significant effect of Gender (being female) on the linear (schooling) growth rate (γ_{21}) during each spring semester, relative to boys (this effect was non-significant for WPM); (5) for WPM only, there was a positive and significant effect of Gender (being female) on initial status (γ_{01}); (5) the variance components from the final model (Model G) for each outcome measure indicated that the participants differed significantly from each other over time (σ_ε^2), that there were significant individual differences in initial status (σ_0^2) and in the linear slope (σ_1^2) after controlling for Gender and Schooling; and (6) the covariance components from the final model (Model G) for each outcome measure indicated that there were negative and significant individual differences in the initial status, linear slope covariance (σ_{01}) after controlling for Gender and Schooling.

Figures 9-10 provide distinct perspectives of the prototypical growth trajectories for the outcome measures in both languages. Figure 9 summarizes the prototypical growth trajectories for MLUw, NDW, and WPM in Spanish and English. Figure 10 collapses the prototypical growth trajectories across measures (top row; due to scale,

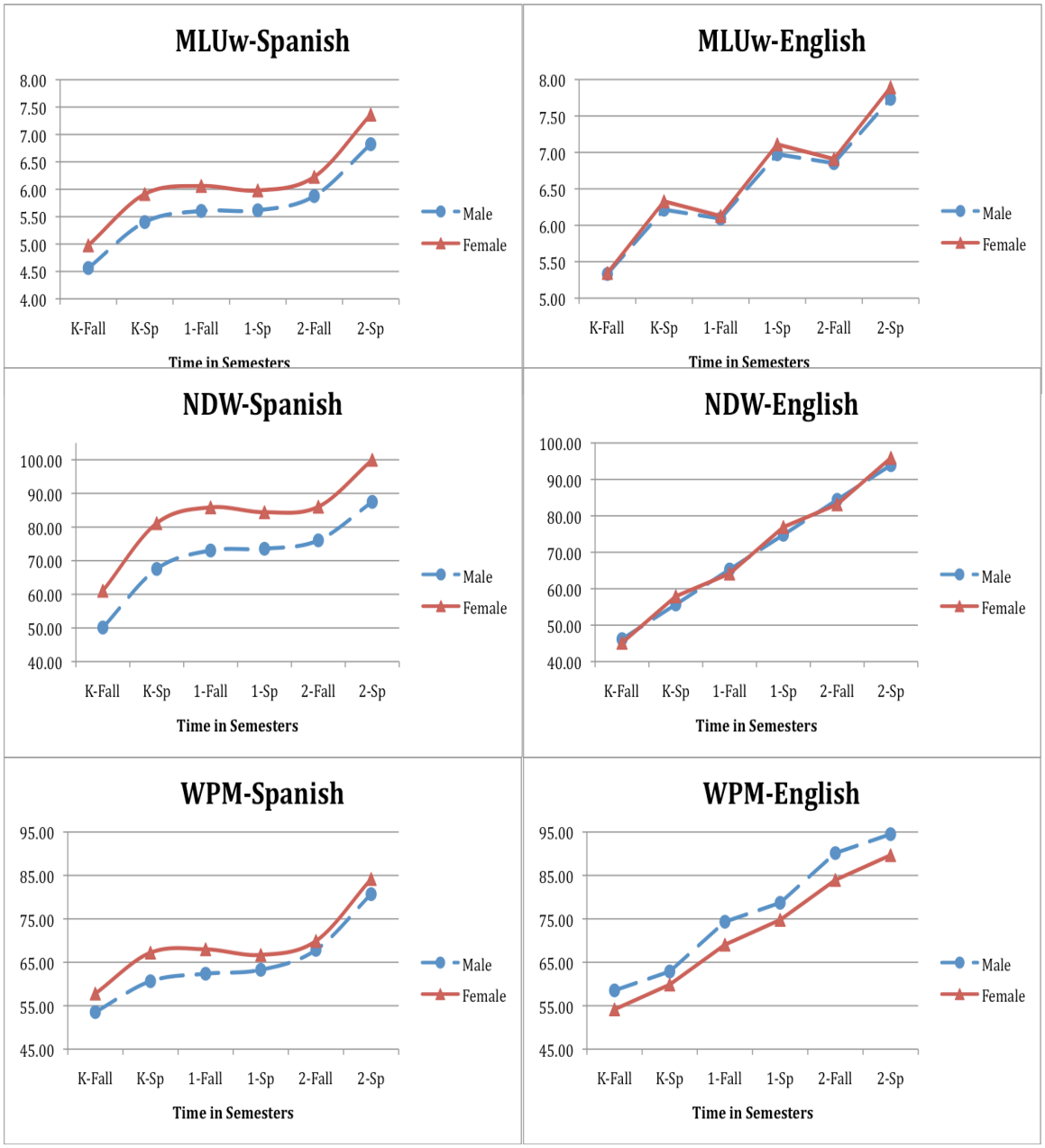


Figure 9. Prototypical Growth Trajectories by Gender for Mean Length of Utterance in Words (MLUw), Number of Different Words (NDW), and Words Per Minute (WPM) in Spanish and English

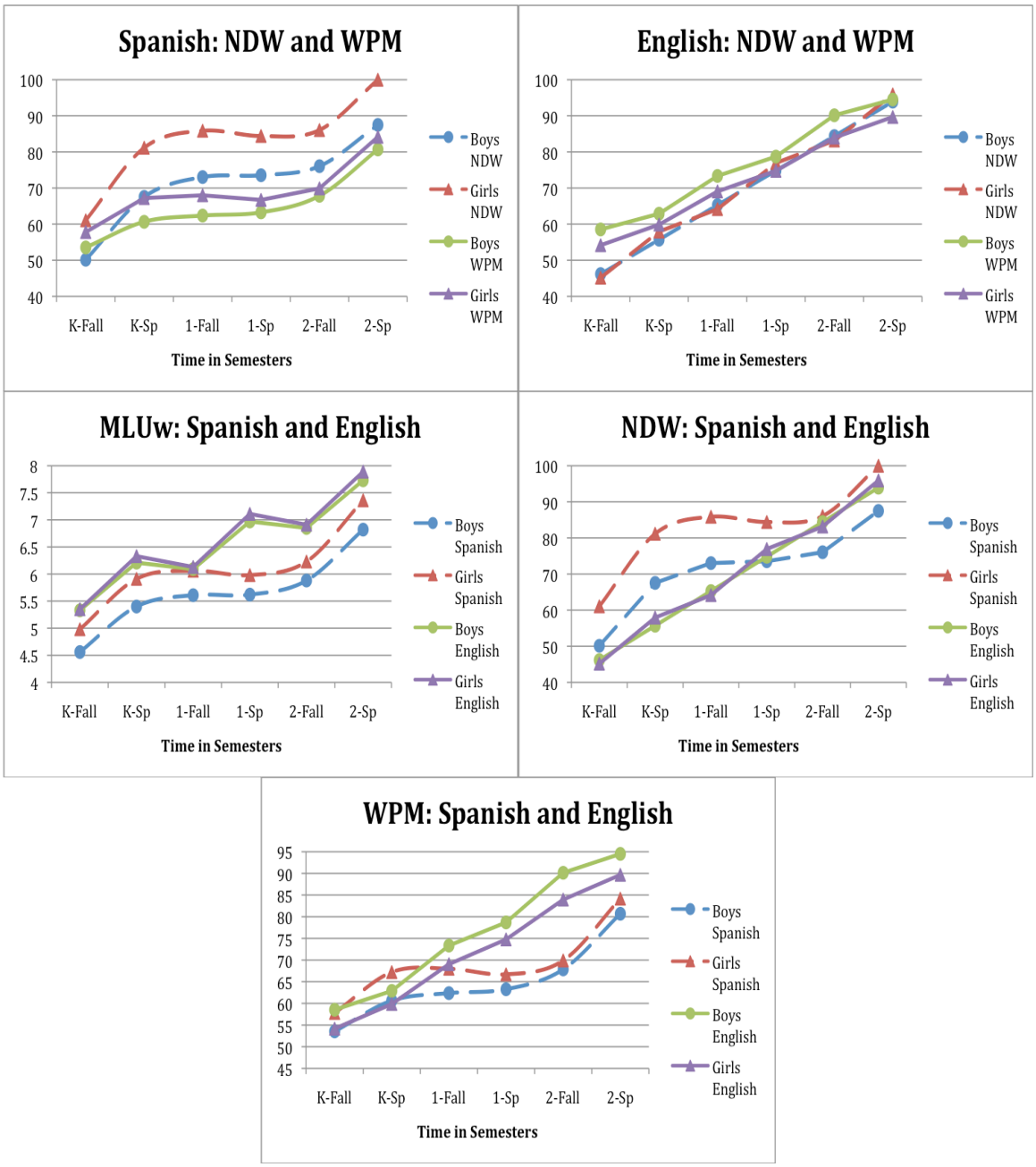


Figure 10. Prototypical Growth Trajectories by Gender for Mean Length of Utterance in Words (MLUw), Number of Different Words (NDW), and Words Per Minute (WPM) Collapsed Across Measures and Across Languages

NDW and WPM only) and across languages (middle and bottom row). The prototypical growth trajectories for the outcome measures in Spanish were curvilinear (cubic), non-monotonic, and continuous, while they were linear, non-monotonic, and discontinuous for English. The prototypical growth trajectories indicated that girls began (fall of kindergarten) and ended (spring of second grade) with significantly higher outcome measures in Spanish than boys, while the English prototypical growth trajectories indicated that boys began (fall of kindergarten) with higher outcomes measures than girls; MLUw-English was the exception, where girls began (fall of kindergarten) and ended (spring of second grade) with moderately higher outcome measures in English than boys. In addition, although boys began with higher NDW-English, girls ended with higher NDW-English, and boys began and ended with higher WPM-English than girls. It is important to note that although the gender-based differences in the prototypical growth trajectories in English appear less consistent than those for Spanish, the effect of Gender on initial status (γ_{01}) was non-significant for MLUw and NDW in English, while it was positive and significant for all outcomes measures in Spanish.

The prototypical growth trajectories indicated common within-language findings. The growth rates of Spanish outcome measures accelerated during kindergarten and second grade, with girls' acceleration outpacing boys, and second grade acceleration outpacing kindergarten (NDW-Spanish was an exception, with kindergarten acceleration outpacing second grade). The growth rate of MLUw and NDW in Spanish decelerated between the spring semester of kindergarten and the fall semester of second grade, with

girls' overall deceleration outpacing boys. The WPM-Spanish growth rate of girls also decelerated between the spring semester of kindergarten and the fall semester of second grade, but the boys' WPM-Spanish growth rate decelerated between the spring semester of kindergarten and the spring semester of second grade (girls' overall deceleration outpaced boys). The fluctuating accelerations and decelerations from kindergarten to second grade contributed to the deceleration in Spanish language skills from the fall of kindergarten to the spring of second grade.

Boys and girls demonstrated constant, linear growth in their English outcome measures during kindergarten, first grade, and second grade, with girls' linear growth outpacing boys. During the summers, boys and girls demonstrated reduced growth in MLUw- (negative growth, with girls negatively outpacing boys) and NDW-English (decrease in growth rate, but overall positive growth, with boys outpacing girls). In contrast, WPM-English demonstrated increased growth during the summers, with boys' growth outpacing girls.

Overall, GCM testing of the outcome measures in Spanish and English (1) determined that in addition to Academic Semester, distinct covariates predicted growth over time and optimized model fit in Spanish (Gender) and English (Schooling + Gender); (2) identified significant intra-individual (within-person; Level-1) and inter-individual (between-person; Level-2) variation in initial status (i.e., intercepts, the starting point of participants' growth) and participants' growth rates (i.e., slopes) for each

outcome measure; and (3) resulted in complex, final growth curve models that provided the most representative growth curve trajectories for each outcome measure.

CHAPTER 4

DISCUSSION

The purpose of this study was to examine the language growth trajectories in children who are in the process of acquiring a second language (Spanish-speaking children classified as ELLs upon entering school) by estimating the initial status and the growth rates of specific oral language skills (MLUw, NDW, and WPM) in each language during the first 3 years of formal schooling and by determining the predictive roles of time, gender, and summer vacation on language growth. The findings were contextualized from the perspective of language as a dynamic system, considering the four principal elements of DST.

The present study utilized secondary data from a larger project, the Bilingual Language Literacy Project (BLLP). The BLLP, a longitudinal project, collected narrative retell language samples produced in Spanish and English from ELL children over a 3-year period. The current study specifically focused on the shape of ELLs' language growth that occurs *during* the first 3 years of formal schooling. Factors that may have impacted language growth *prior* to formal schooling (e.g., age of acquisition of Spanish and English; the amount and types of early language exposure at home) were not considered. The only factors considered during the school years were the effect of gender and the effect of schooling preceding each spring semester.

The final longitudinal dataset used in this study consisted of 12,248 oral narrative language samples (6,516 Spanish; 5,732 English) that were produced by 1,723 ELLs. Growth curve model (GCM) testing was used to profile the longitudinal growth of the ELLs' oral language skills in Spanish and English over time. The theoretical framework is revisited first, followed by a discussion of the findings. The limitations of this study and future work are also considered.

Dynamic Systems Theory (DST): Elements

In this study language was conceptualized as a complex dynamic system composed of linguistic subsystems. The oral language measures MLUw, NDW, and WPM were used as indicators of the morphosyntactic subsystem, the lexical subsystem, and the global integration of linguistic subsystems, respectively. The findings of this study are discussed with respect to the four principal elements of Dynamic Systems Theory (DST): (1) an emphasis on change over time; (2) intra- and inter-individual variability; (3) impact of initial status on growth; and (4) interconnection of subsystems.

Change (Growth) Over Time

The first research question focused on change over time. Specifically, did measures associated with different subsystems (MLUw and NDW) and the global integration of subsystems (WPM) demonstrate change (growth) over time? If so, what was the shape of this growth in each language, and what predictive roles (if any) were

played by academic semester, gender, and/or schooling on language growth in Spanish and English?

Growth in Spanish

Based on the existing literature on the language growth of ELLs (Conboy et al., 2006; Hammer et al., 2008a; Kohnert et al., 2010; Miller et al., 2006; Simon-Cereijido et al., 2009), the Spanish oral language measures were hypothesized to demonstrate positive growth over time. This hypothesis was confirmed. The prototypical growth trajectories for each of the oral language measures in Spanish indicated continuous, curvilinear growth. The data indicated that academic semester (metric of time) optimized model fit and predicted growth over time as indicated by the following significant fixed effects: positive linear slope (γ_{10}), negative quadratic slope (γ_{20}), and positive cubic slope (γ_{30}). These significant fixed effects emphasized the importance of adding academic semester to account for growth over time, and of modeling nested linear and curvilinear growth rates over time.

Perhaps the most interesting aspect of the growth trajectory is the shape of this growth. Bates et al. (1995) suggested that determining the shape of growth trajectories would ultimately be necessary to explain some of the well-known aspects of language development, which are decidedly non-linear. For example, the vocabulary explosion seen in many children during early vocabulary acquisition demonstrates a trajectory of exponential growth, which has been shown (Huttenlocher, 1991; McMurray, 2007) to eventually decelerate. Past tense learning in English (Bates et al., 1993) demonstrates a

trajectory of alternating growth and decline. Recent work by Rice and colleagues (Rice et al., 2006, 2009) has revealed quadratic trajectories that capture the growth of morphosyntactic (tense marking; MLUm level) skills of monolingual English typically developing (TD) children and children with specific language impairment (SLI).

Although the shape and direction of ELLs' language growth trajectories are largely unknown (Hammer et al., 2008b), based on prior findings (Bedore et al., 2006; Conboy et al., 2006; Kohnert et al., 2010; Simon-Cereijido et al., 2009), the shape of each measure in Spanish was hypothesized to be similar. The results of this study confirmed this hypothesis. The overall shape of the prototypical growth trajectories of each of the measures in Spanish was cubic—more specifically, it was curvilinear, non-monotonic, and continuous according to Elman et al. (1999) taxonomy. The shape of the prototypical growth trajectories estimated Spanish outcome measures that accelerated during kindergarten, decelerated to an almost-flat trajectory between the spring semester of kindergarten and the fall semester of second grade, and then accelerated again during second grade.

The deceleration in Spanish language skills from the fall of kindergarten to the spring of second grade was unexpected. A number of factors could account for the period of acceleration during kindergarten, deceleration during first grade, and renewed acceleration in second grade. One possibility is that the temporary deceleration in Spanish found in this study is real. It is possible that this study captured a period of time during which the Spanish language system was temporarily restructuring or reorganizing

itself (Perkins et al., 1996), as a function of competing growth from linguistic subsystems not considered in this study (e.g., phonology), or from the linguistic subsystem development of the second language system. Phonological development, which was not considered in this study, has been shown (Smit et al., 1990) to plateau and to demonstrate periods of growth and decline well into middle childhood. The internal restructuring or reorganization of dynamic subsystems, in response to the growth of other subsystems, is recognized within DST (De Bot et al., 2007a; Larsen-Freeman, 2006) as evidence of co-development. This possibility is later addressed in the discussion regarding the co-development of interconnected subsystems.

Another explanation for this temporary acceleration, deceleration, and acceleration, and this is purely speculative, is the possibility of distinct subgroups of ELLs (Hammer et al., 2008b; Iglesias, Miller, Jones, Hanlon, & Francis, in preparation) collectively producing an 'average' effect that is not typical of the children. The prototypical growth trajectory captures a pattern of growth that is representative of the individual growth patterns demonstrated by participants within 1 *SD* of the mean. Therefore, if there were subgroups of children with unique trajectories, the growth trajectories of these subgroups could be masked within this prototypical trajectory. The significant variability in initial status found in this study suggests the possibility of subgroups (other than gender), each with their own starting points of growth. It is possible that the Spanish language skills of most ELL subgroups were increasing during kindergarten, except for a few children who were systematically exposed to English in

the classroom (e.g., Hammer et al., 2009), and may have begun to show negative growth (decline) in Spanish. During first grade, it is possible that additional children would also begin to demonstrate native language attrition (Anderson, 1994; Kohnert, 2008) as a function of increasingly using English to interact with their peers (Rojas et al., in preparation). However, other ELL subgroups (especially ELLs enrolled in dual language or late transitional programs) may continue to demonstrate positive growth in Spanish language skills during first grade. The interaction of the subgroups' distinct growth trajectories during first grade (many decelerating and some accelerating) may have resulted in the overall period of deceleration found in this study. Then in second grade, it is possible that participants whose Spanish language skills had been declining, either begin to level off (plateau) or minimally decline. It is also possible that the Spanish language skills of ELL subgroups that had not previously shown negative growth will continue to show positive growth during second grade. This combination of subgroup growth trajectories (few decelerating and many accelerating or leveling off) may have resulted in renewed acceleration during second grade. The interaction of these distinct growth trajectories over time may explain the S-shaped, cubic growth trajectories found in this study for each of the Spanish outcome measures. To verify these possibilities future studies must first identify ELL subgroups, and then model the corresponding subgroup growth trajectories in Spanish. What we know from this study's findings, however, is that ELLs' Spanish prototypical growth trajectories indicate overall positive growth over time, with a period of declaration between the fall of kindergarten and the spring of second grade.

Based on prior findings (Hammer et al., 2008b, 2009; Uchikoshi, 2006b), gender and schooling were expected to play a predictive role in the shape of Spanish growth. Although this hypothesis was confirmed for gender, it was not for schooling. Boys and girls demonstrated near parallel shapes in their growth trajectories, with girls' growth significantly outpacing boys throughout the duration of the study. Gender (effect of being female) predicted group-based differences of growth over time for the outcome measures in Spanish as indicated by the following significant fixed effects: gender (positive effect on the initial status of MLUw and NDW) (γ_{01}), positive interaction of gender on linear slope (γ_{11}), negative interaction of gender on quadratic slope (γ_{21}), and positive interaction of gender on cubic slope (γ_{31}). These significant fixed effects emphasized the importance of adding gender to account for group-based differences, and of modeling the effect of gender on nested (linear and curvilinear) growth rates over time.

Girls began with significantly higher initial levels of MLUw, NDW, and WPM in Spanish. Further, as a result of the positive and significant interaction of gender (effect of being female) on the growth rates, girls maintained an advantage over boys in the growth of their oral language measures. The female growth-advantage found in this study was consistent with prior gender-based findings of monolingual *early* language development (e.g., Bauer et al., 2002; Bornstein et al., 2004; Coates, 1993; Fenson et al., 1994, 2000; Huttenlocher et al., 1991). One key difference between the results of this study and monolingual findings was that while the female growth-advantage disappears

around the age of 3;0 in monolinguals, in ELLs this advantage continues until at least the end of second grade (mean age = 8;2) in Spanish.

The female growth-advantage found in this study, however, was *not* consistent with prior findings with ELL children (Hammer et al., 2009; Uchikoshi, 2006b). There are several factors that could account for the discrepancy. First, the differences could be due to the different aspects of language that were studied. Hammer et al. (2009) and Uchikoshi (2006b) failed to find gender-based differences in the growth of ELLs' Spanish *receptive* language, while the female growth-advantage found in this study was based on measures of *expressive* language. Given that receptive language development typically outpaces expressive language development (Gleason & Ratner, 2009; Hoff, 2005; Paul, 2006), it is possible that this study captured a period of ongoing expressive Spanish language development in ELLs, but that Hammer et al. (2009) and Uchikoshi (2006b) captured periods of time during which receptive language skills had largely been established in the native language of ELLs, and thus gender-based differences had disappeared.

Second, the tasks used to estimate the growth of language skills were different. Hammer et al. (2009) and Uchikoshi (2006b) administered formal assessment tools that were not standardized on bilingual children (a fact noted by Hammer et al., 2009). In contrast, the present study's oral language measures (MLUw, NDW, and WPM) were elicited from narrative retell language samples in Spanish. It is possible that this study's narrative language sampling data were more representative of ELLs' true linguistic skills

(Gutiérrez-Clellen et al., 2000; Heilmann et al., 2008, 2010; Miller et al., 2006; Rojas & Iglesias, 2009, 2010), and that the standardized tools used by Hammer et al. (2009) and Uchikoshi (2006b) were too global to detect gender-based differences.

Third, the time period of language growth captured by each study and the frequency of measurement within that time period differed. Therefore, each study captured a different temporal component of the overall shape of ELLs' language growth. Hammer et al. (2009) measured ELL language growth during 2 years of Head Start and 1 year in kindergarten using 6 waves of measurement. Uchikoshi (2006b) tracked the language growth of ELLs during 1 year, kindergarten, using 3 waves of measurement. The present study captured language growth from kindergarten to second grade using 6 waves of measurement. It is possible that the Spanish female-growth advantage may be active during early Spanish language development (this would be consistent with monolingual findings; Bauer et al., 2002; Bornstein et al., 2004), and that gender-based differences temporarily become inactive during the preschool years (Head Start to kindergarten), only to be reactivated from kindergarten to the end of second grade.

It is also possible that each study captured gender-based patterns of growth that were affected by the degree of exposure the participants had to the native language. Along with capturing different time periods of language growth, the systematic exposure of ELLs to the native and the second language also differed in each study. Hammer et al. (2009) participants received the earliest and most intensive systematic exposure to English (2 years in English-only Head Start; 1 year in English-only kindergarten).

Uchikoshi's (2006b) participants received an undefined, but presumably near-balanced systematic exposure to both languages (1 year; Spanish-English bilingual kindergarten). The present study's participants received a range of systematic exposure to English and Spanish (3 years; multiple programs of language instruction, kindergarten to second grade). It is possible that early, prolonged, and systematic exposure to English as a second language may attenuate gender-based differences in Spanish as a function of native language attrition (Anderson, 2004; Kohnert, 2008).

Fourth, the language(s) of academic instruction received by the ELLs and the number of participants sampled in these studies differed. Hammer et al. (2009) attributed their non-finding to a total of 72 ELLs being enrolled solely in English-immersion programs, while Uchikoshi (2006b) who did not speculate on her non-finding, included 150 participants enrolled in "Spanish-English bilingual kindergarten classrooms" (p. 35). Although English-immersion (all children, regardless of gender, receive the same type of input) may explain the lack of gender-based differences in English found by Hammer et al. (2009), it does not fully explain the lack of gender-based differences in Spanish. The fact that the present study sampled a much larger group of ELLs (1,723) enrolled in a wide range of programs of language instruction (English-immersion, early-transitional bilingual, late-transitional bilingual, and dual language) may have contributed to a wider continuum of gender-based differences not captured by prior studies (Hammer et al., 2009; Uchikoshi, 2006b).

As noted previously, the hypothesis that schooling would play a predictive role in the shape of Spanish growth was not confirmed. The fit of growth curve models with schooling was inferior to the fit of growth curve models without schooling for the Spanish outcome measures. Thus, the effect of schooling (effect of having or not having had schooling in the preceding semester) on Spanish language growth was not directly tested. However, not directly testing the schooling effect did not prevent the examination of Spanish language growth between academic semesters. Although the overall rates of growth of the Spanish language outcome measures decelerated (slowed down) between the spring of kindergarten and the fall of second grade, relative to their growth rates during kindergarten and second grade, they nevertheless continued to positively grow throughout this period of deceleration. This finding was at odds with the negative growth (decline) found by Hammer et al. (2008b) during the summer vacation in the standard scores for receptive language in Spanish for ELLs. As with Hammer et al. (2009), the study by Hammer et al. (2008b) also utilized formal assessment tools normed on monolingual children to assess the receptive language skills in Spanish of 83 ELLs who were enrolled in English-immersion Head Start programs. Thus, the same factors (receptive versus expressive language measures; standardized assessment versus language sampling; English-immersion classrooms versus a range of academic programs of language instruction; restricted sample size versus broad sample size) that may account for the discrepancy of gender-based findings, may also account for the difference in finding decelerating (yet positive growth) versus true negative language growth in Spanish.

Growth in English

Based on the existing literature on the language growth of ELLs (Conboy et al., 2006; Hammer et al., 2008a; Kohnert et al., 2010; Miller et al., 2006; Simon-Cereijido et al., 2009), the English oral language measures were hypothesized to demonstrate positive growth over time. This hypothesis was confirmed. The prototypical growth trajectories for each of the oral language measures in English indicated discontinuous growth over time. The data indicated that academic semester (metric of time) and the effect of schooling (time-varying covariate) optimized model fit and predicted growth over time as indicated by the following significant fixed effects: positive linear slope (γ_{10}) and a positive schooling slope (γ_{20}). These significant fixed effects emphasized the importance of adding academic semester and schooling to account for growth over time.

The shape and direction of ELLs' language growth trajectories are largely unknown (Hammer, Lawrence, & Miccio, 2008b). Based on prior findings (Bedore et al., 2006; Conboy et al., 2006; Kohnert et al., 2010; Simon-Cereijido et al., 2009) the shape of English language growth was hypothesized to be similar across all three measures. The results of this study partially confirmed this hypothesis. The overall shape of the prototypical growth trajectories of each of the measures in English was discontinuous—more specifically, it was linear, non-monotonic, and discontinuous according to Elman et al. (1999) taxonomy. The shape of the prototypical growth trajectories estimated English outcome measures that demonstrated constant, linear growth during the academic year. The difference among the measures is the growth

during the summer. During the summers, the shape of the prototypical growth trajectories for MLUw- and NDW-English demonstrated reduced growth. Although NDW-English continued to demonstrate positive growth (decrease in growth rate, but overall positive growth), MLUw-English demonstrated negative growth (decline) during the summers. WPM-English demonstrated increased growth during the summers.

Although the reduced growth of MLUw- and NDW-English during the summers (relative to their linear growth during academic years) was expected based on prior findings (Alexander, et al., 2001; Cooper et al., 1996; Hammer et al., 2008b; McCoach et al., 2006), the increased growth of WPM-English was unexpected. Prior findings on the negative impact of summer vacations on reading skills (Alexander et al., 2001; Cooper et al., 1996; McCoach et al., 2006) and the varying impact on ELLs' receptive language skills (Hammer et al., 2008b) allow one to speculate on the differential growth of the English language measures during the summer, including the reduced growth of MLUw and NDW, and the increased growth of WPM. If the constant, linear growth of English morphosyntactic and lexical skills during academic years is associated with systematic support and exposure to English in the classroom, it is possible that the periods of reduced growth of English are associated with the lack of this systematic support and exposure during summer vacations. It is also possible that the negative growth (decline) isolated to MLUw-English may be due to the morphosyntactic subsystem being more susceptible to regress than other linguistic subsystems. Interestingly, the increased growth of WPM during the summers found in this study might be related to the negative

growth of MLUw. As noticed in ELLs by Bedore et al. (2006) and Carias and Ingram (2006), an increase in MLUw tends to result in a decrease in verbal fluency (more pauses, repetitions, and revisions). Thus, the negative growth (decline) in MLUw during the summers may have caused an increased growth in WPM. It is possible that the shorter utterances used by ELL children during the summer (decline in MLUw) may have resulted in decreased pause time, repetitions, and revisions, which would in turn increase overall verbal fluency (increased WPM). Although this has not been previously shown in the language literature (Conboy et al., 2006; Kohnert et al., 2010; Simon-Cereijido et al., 2009), the differential effects of summer vacations across domains have been seen in the educational research literature. Cooper et al., (1996), for example, found that summer vacation had a greater negative impact on math than on reading skills; with math computation and spelling being the most negatively impacted academic skills. More fine-grained analyses of the changes in the other measures, in both languages, during summer might provide some insight as to why certain linguistic subsystems are more susceptible to periods of no instruction.

Based on prior findings (Hammer et al., 2008b, 2009; Uchikoshi, 2006b) gender and schooling were expected to play a predictive role in the shape of English growth. This hypothesis was confirmed for gender and schooling. Boys and girls demonstrated near parallel shapes in their growth trajectories. Gender (effect of being female) and schooling (effect of having or not having had schooling in the preceding semester) predicted differences of growth over time for the outcome measures in English as

indicated by the following significant fixed effects: gender (negative effect on the initial status of WPM only) (γ_{01}), an additional positive linear (schooling) slope (γ_{20}) during each spring semester for all measures, and a positive interaction of gender on the additional linear (schooling) slope (γ_{21}) for MLUw and NDW. These significant fixed effects emphasized the importance of adding gender to account for group-based differences, adding schooling to account for the effect of having had schooling preceding each spring semester, and of modeling the effect of gender on schooling over time.

The prototypical growth trajectories of English demonstrated gender-based differences that would not have been detected if the effect of gender had been modeled solely on initial status rather than on initial status and the interaction of gender with growth rates. Although significant gender-based differences in initial status were not found for MLUw- and NDW-English, gender (effect of being female) had a positive and significant interaction with the schooling growth rate of these two oral language measures. As a result of the positive gender-growth rate interaction, girls maintained a moderate advantage over boys in the growth of their oral language measures, particularly during each spring semester. In contrast, boys began with significantly higher initial status of WPM-English than girls, and they maintained this WPM-English advantage over time even though a gender-growth interaction was not found for this oral language measure. If a positive and significant gender-schooling growth rate interaction had been found, it is possible that girls' WPM-English growth may have outpaced boys, erasing the boys' initial status advantage.

The female growth-advantage found for MLUw- and NDW-English (like the Spanish female growth-advantage) was consistent with the same advantage found in monolingual early language development (e.g., Bauer et al., 2002; Bornstein et al., 2004; Coates, 1993; Fenson et al., 1994, 2000; Huttenlocher et al., 1991), but in ELLs the female-growth advantage continues until at least the end of second grade (mean age = 8;2) in English. However, the female growth-advantage found for MLUw- and NDW-English was not consistent with prior findings with ELL children (Hammer et al., 2009; Uchikoshi, 2006b). Many of the factors that can account for the discrepancy of gender-based findings between this and prior work on ELLs' Spanish language skills, are the same factors that account for the female growth-advantage found in the growth of MLUw- and NDW-English: receptive versus expressive language; formal assessment versus narrative language sampling, and differences in time period and frequency of measurement.

Although this study found that girls moderately outpaced boys in their growth of MLUw- and NDW-English, it also found that boys began with significantly higher initial levels of WPM-English than girls, and they maintained this advantage over time. The male-growth advantage found for WPM-English differs from all other English and Spanish oral language measures in this study, and from previous monolingual findings (e.g., Bauer et al., 2002; Bornstein et al., 2004; Coates, 1993; Fenson et al., 1994, 2000; Huttenlocher et al., 1991) of a female-growth advantage. However, it is consistent with the work of Uchikoshi (2006b), who also found a male-growth advantage in the English

expressive language skills of ELLs. Given that Uchikoshi's (2006b) male-growth advantage was found during kindergarten, this study extends Uchikoshi's (2006b) findings by estimating an ELL male-growth advantage in WPM-English from kindergarten to second grade.

As noted previously, the hypothesis that schooling would play a predictive role in the shape of English growth was confirmed. The fit of growth curve models with schooling was superior to the fit of growth curve models without schooling for the English outcome measures. Thus, the effect of schooling (effect of having or not having had schooling in the preceding semester) on English language growth was directly tested. Although the English outcome measures demonstrated constant, linear growth during each academic grade, NDW- and WPM-English demonstrated positive, yet reduced growth during the summers. MLUw-English was the only measure that demonstrated negative, linear growth during the summers. The finding of reduced growth of ELLs' language skills during the summer was consistent with the negative growth (decline) found during the summer in reading skills (Alexander et al., 2001; Cooper et al., 1996; McCoach et al., 2006), and in the English receptive language skills of ELLs (Hammer et al., 2008b). As previously discussed, the reduction in the rate of growth shown in ELLs' English language skills during the summers may be associated with the lack of ongoing systematic support and exposure to the second language.

Summary of Growth in Spanish and English

The first research question focused on the change (growth) of ELLs' language skills over time. The results from the present study indicated the following: all oral language measures demonstrated growth; the shapes of growth were similar within-language; language growth in Spanish was predicted by academic semester and gender; and language growth in English was predicted by academic semester, gender, and schooling.

Intra- and Inter-Individual Variability

The second research question focused on the variability of individuals' performance. Specifically, were there significant individual differences in the initial status of growth and in the growth rates for MLUw, NDW, and WPM in Spanish and English? If so, were these individual differences detected at the intra- and inter-individual levels?

Although the number of published studies on the individual differences of ELLs' language growth are small (Hammer et al., 2008a,b), varying degrees of inter- and intra-individual variance across measures and across languages were hypothesized. This hypothesis was confirmed. The final growth curve models (GCMs) for each of the oral language measures in Spanish and English indicated significant individual differences at the intra- and inter-individual levels that were not accounted for by gender-based differences and/or the effect of schooling. All of the outcome measures in both

languages demonstrated significant intra-individual variance, and significant inter-individual variance in initial status. Significant inter-individual variance in the growth rates was found at varying degrees for each of the outcome measures in Spanish and English. Individual differences were indicated by the following significant variance components: within-person variance (σ_ε^2), between-person intercept variance (σ_0^2), between-person linear slope variance (WPM-Spanish; all measures in English) (σ_1^2), between-person quadratic slope variance (WPM-Spanish only) (σ_2^2), and between-person cubic slope variance (WPM-Spanish only) (σ_3^2). These significant variance components emphasized the importance of accounting for individual differences at the inter- and intra-individual levels in each language.

Notwithstanding that the study of variability and individual differences in child language development has traditionally been underemphasized, there is an emerging body of work (Bates et al., 1995; Fenson et al., 1994; Hadley and Holt, 2005; Hammer et al., 2008a,b; Rice et al., 1998, 2006, 2009) that has begun to consider the role of variability and individual differences in accounting for the heterogeneity of language learners, which extends our understanding of change over time (growth) beyond mean initial status and mean growth rates (Seltzer et al., 2003; Siegler, 2002). The present study found significant individual differences in ELLs at the intra-individual (within-person) level across all six waves of measurement for all the outcome measures in Spanish and English. This finding suggested that the rate of growth of each participant, irrespective of his or her gender or the effect of schooling, significantly differed from one

academic semester to the next. This significant intra-individual variability indicated a considerable degree of heterogeneity of ELLs in their Spanish and English language growth over time. This finding has important implications for purposes of language assessment. Although prototypical growth trajectories summarize mean language growth, this growth is highly variable over time within each child, in each of their languages. This intra-individual variability is particularly relevant when considering alternative approaches to assessment, such as dynamic assessment (Peña, Iglesias, & Lidz, 2001), which measures the child's ability to learn over a short period of time. While it is a step in the right direction, a limitation of dynamic assessment is that it may not detect the intra-individual variability found in this study, unless if used to measure language growth across academic semesters. Such alternative approaches may capture the intra-individual variability of ELLs found in this study, which would otherwise be undetected by traditional assessment methods that are typically administered in one testing session and which may yield inconsistent findings depending on exactly when the measures are collected. This finding of intra-individual variability advances our knowledge of language learners in general, as the language development literature is largely silent with regard to either reporting (e.g., Hammer et al., 2009; Pan et al., 2005; Kan, 2010) or discussing (e.g., Uchikoshi, 2006b) intra-individual differences. The work of Rice and colleagues (Rice et al., 2006, 2009), which examines the growth of morphosyntactic skills in monolingual children with SLI, is a notable exception to this trend.

In addition to finding significant variability at the intra-individual level, this study found significant variability at the inter-individual (between-person) level in both languages. The results indicted that, regardless of gender, participants differed significantly in their starting points of growth (fall of kindergarten) for all outcome measures across both languages. An implication of this finding is that researchers and professional service practitioners should expect to encounter ELLs with a significantly broad, and highly variable range of language skills in Spanish and English at formal school entry. Particularly important is that the language skills of these children are not homogeneous despite their categorization as ELLs. These significant inter-individual differences in initial status (intercept) need to be considered in both research and practice.

Different degrees of inter-individual variance in the growth rates (slopes) were found across the oral language measures in Spanish and English. The data indicated that participants differed significantly in their rates of linear, quadratic, and cubic growth over time for WPM in Spanish only, but not for MLUw- or NDW-Spanish[3]. In English, the data indicated that participants differed significantly in their rates of linear growth over time across all oral language measures. These significant inter-individual differences in growth rates further suggest the possibility of ELL subgroups (perhaps as a function of program of language instruction) that may each be contributing unique growth trajectories in English. In concert, these findings indicated three general patterns of inter-

---

[3] Given that fixed rates of growth do not allow for variance to be estimated, it should be noted that the final GCM for MLUw-Spanish did not allow participants to vary in their linear rates of growth, and that the final GCM for NDW-Spanish did not allow participants to vary in their linear or their cubic rates of growth.

individual differences in ELLs' language growth. First, regardless of gender, initial status in English and in Spanish varied significantly across participants. Second, regardless of gender, ELLs demonstrated significantly different rates of growth across all oral language measures in English and in WPM-Spanish. Third, the differences in the growth rates of MLUw- and NDW-Spanish were largely accounted for by gender.

These findings of inter-individual differences in initial status and differential rates of growth were consistent with prior findings with slowly developing monolingual language learners (Hadley & Holt, 2005), children with SLI (Rice et al., 1998, 2006, 2009) and ELLs in Head Start (Hammer et al., 2008a,b). Hadley and Holt (2005) found significant inter-individual differences in the initial status and the growth rates of tense marking in monolinguals. Rice et al. (1998, 2006, 2009) found a range of significant inter-individual differences in the initial status and the growth rates of morphosyntactic skills (verb tense marking; MLUm level) with monolingual children with SLI. The group membership (positive growth versus negative growth of language skills in the first year of Head Start) of Hammer et al. (2008b) was determined by patterns of inter-individual variation unexpectedly found in a previous study (Hammer et al., 2008a). The range of findings concerning inter-individual differences in initial status and rates of growth suggest that there may be distinct subgroups of language learners, including ELLs. These subgroups begin with significantly different levels of language skills at initial status and also show significantly different rates and perhaps patterns of growth over time.

In summary, significant intra-individual differences in the growth of all the oral language measures, across each wave of measurement, were found for both languages; significant intra-individual differences in the initial status of participants for all the oral language measures were found for both languages; significant inter-individual differences in the growth rates were found for WPM-Spanish; and significant inter-individual differences in the growth rates were found for all the oral language measures in English.

Impact of Initial Status on Growth

The third research question focused on the extent to which initial status impacts growth trajectories. Specifically, did the initial status of growth have an impact on the growth trajectories of MLUw, NDW, and WPM in Spanish and English? If so, what was the direction and significance of the impact of initial status on growth?

Determining the impact of initial status on growth trajectories allows for the following question to be answered: do participants who are lower performing at initial status 'catch up' to higher performers? The intercept-slope covariance estimates the strength and direction of the relationship between the initial status of growth (intercept) and the rate of growth (slope), after controlling for group or program membership (Singer & Willett, 2003). The extent to which initial status affects the growth trajectory, information obtained by examining the intercept-slope covariance, is commonplace in the education literature (e.g., Aarnoutse & van Leeuwe, 2000; Ding & Davison, 2005). However, intercept-slope covariance data are not often found in contemporary work on language development (but see Rice et al., 2006, 2009 for an exception). Based on prior

findings that the slope of verb marking skills and MLUm were not affected by initial status in monolingual English TD and SLI children (Rice et al., 2006, 2009), MLUw in Spanish and English were hypothesized to demonstrate non-significant intercept-slope covariances. By extension, NDW and WPM were hypothesized to also demonstrate non-significant intercept-slope covariances in both languages. These hypotheses were partially confirmed.

The final growth curve models (GCMs) of MLUw in Spanish demonstrated positive and significant relationship between initial status and growth rate, as indicated by the significant intercept-cubic slope covariance (σ_{03}), after controlling for the effect of gender on initial status. The relationships between initial status of NDW and initial status of WPM and their respective growth rates were non-significant. The growth of MLUw-Spanish was systematically related to initial status, as indicated by a *positive* and significant intercept-cubic slope covariance. This finding suggested that participants with higher levels of MLUw-Spanish at initial status (fall of kindergarten) tended to have *faster* rates of growth, on average, than those with lower levels at initial status. This finding also suggested that the lower performing participants at initial status may not catch up to higher performers at initial status, as the initial differences in MLUw-Spanish become more pronounced over time. Although this significant covariance finding in TD ELL children was not consistent with the intercept-slope covariance data reported in Rice et al. (2006, 2009), it should be noted that the work of Rice et al. (2006, 2009) involved monolingual English TD and SLI children.

The present study found that the acceleration in growth of NDW-Spanish was unrelated to initial status, as indicated by a non-significant intercept-quadratic slope covariance. The study also found that the growth and acceleration in growth of WPM-Spanish were also unrelated to initial status, as indicated by a non-significant intercept-linear slope covariance, a non-significant intercept-quadratic slope covariance, and a non-significant intercept-cubic slope covariance. These non-significant covariances were consistent with prior findings (Rice et al., 2006, 2009) and suggested that the growth rates of NDW- and WPM-Spanish might be similar among participants (irrespective of their gender), regardless of their initial status of growth. These non-significant findings also suggested that the initial differences in NDW- and WPM-Spanish should be expected to remain consistent over time. Therefore, ELLs with initially lower levels of NDW- and WPM-Spanish may, or may not catch up to ELLs with higher levels of NDW- and WPM-Spanish at initial status.

The final growth curve models (GCMs) demonstrated systematic relationships between initial status and growth rates (after controlling for gender) as indicated by negative and significant intercept-linear slope covariances (σ_{01}) for all of the oral language measures in English. These findings suggested that participants with higher levels of MLUw-, NDW-, and WPM-English at initial status (fall of kindergarten) tended to have *slower* rates of growth, on average, than those with lower levels at initial status, suggesting a "topping out" phenomenon. The findings also suggested that the lower performing participants at initial status may eventually catch up to higher performers at

initial status, as the initial differences in MLUw-, NDW-, and WPM-English become less pronounced over time. The discrepancy between this study's negative and significant covariances in English, and prior findings (Rice et al., 2006, 2009) of non-significant covariances, may be associated with differences in the participant populations studied. The Rice et al. (2006, 2009) group studied monolingual children with SLI, while the present study studied TD ELL children.

These findings have theoretical implications for Spanish and English as dynamic language systems. The sensitivity to initial conditions posited by DST specifies that differences at the beginning of growth may result in major changes over time. For Spanish, the positive and significant covariance findings suggested that the morphosyntactic subsystem (measured by MLUw-Spanish) demonstrated sensitivity to initial conditions. One should expect that, on the average, children who come in to school with a low level of MLUw-Spanish will grow slower than those with higher levels, and they probably will not catch up to them. However, the non-significant covariances of NDW- and WPM-Spanish suggested that the lexical subsystem and the global integration of the subsystems did not demonstrate this sensitivity, and one should expect that some initially lower performing kids in NDW- and WPM-Spanish will catch up and some will not. For English, the negative and significant covariance findings suggest that the morphosyntactic (MLUw-English) subsystem, the lexical subsystem (NDW-English), and the global integration of subsystems (WPM-English) demonstrated sensitivity to initial conditions. One should expect that, on the average, children who

come in to school with low levels of MLUw-, NDW-, and WPM-English will grow faster than those with initially higher levels, and that they will probably catch up to them.

In summary, the growth of MLUw-Spanish was systematically related to initial status (lower performers at initial status may not catch up to higher performers); the growth of NDW- and WPM-Spanish were unrelated to its initial status (lower performers at initial status may, or may not catch up to higher performers); and the growth of MLUw-, NDW-, and WPM-English was systematically related to initial status (lower performers at initial status may catch up to higher performers).

Co-Development of Interconnected Subsystems

The fourth research question focused on the co-development of interconnected linguistic subsystems. Specifically, did the growth of the different linguistic subsystems and their integration demonstrate dynamic co-development over time? If so, was this co-development restricted across-subsystems and within-language, or was it also evident within-subsystems and across-language?

Based on prior findings (Bedore et al., 2006; Carias & Ingram, 2006; Castilla, et al., 2009; Conboy et al., 2006; Kohnert et al., 2010; Simon-Cereijido et al., 2009), the interconnected linguistic subsystems and their integration were hypothesized to demonstrate within-language and across-subsystem co-development. This hypothesis was not empirically confirmed, given that the present study did not directly test the within-language interactions between outcome measures of different linguistic

subsystems (e.g., the interaction of MLUw and NDW in Spanish) or across-language interactions (e.g., the interaction of WPM-Spanish and WPM-English). However, qualitative observations based on visual inspection of the available prototypical growth trajectories, and growth curve modeling (GCM) estimates, suggested the possibility of across-subsystem and within-language co-development, and also of and within-subsystem and across-language co-development.

One of the principal elements of a dynamic system is the complexity of having multiple interconnected subsystems that co-develop over time, where the change of one subsystem can have an impact on another subsystem(s). The co-development of interconnected subsystems can be explained by two dynamic phenomena that occur over time: connected growers and phase transitions. Connected growers refer to interconnected subsystems that bootstrap each other's growth over time (De Bot et al., 2007a,b; Larsen-Freeman, 2007; Marin & Peltzer-Karpf, 2009; Thelen & Bates, 2003), resulting in states of parallel, dynamic growth. This would be indicated by two positive rates of growth, or by two negative rates of growth. Phase transitions refer to the competing growth of interconnected subsystems (Hohenberger & Peltzer-Karpf, 2009; Marin & Peltzer-Karpf, 2009; van Geert, 2007, 2008; Thelen & Bates, 2003), which result in an ongoing fluctuation between steady states and states of dynamic change. This would be indicated by steady state and a state of positive or negative growth, or by opposing states of positive and negative growth.

Across-Subsystem Co-Development Within a Language

Qualitative observations, based on visual inspection of the available prototypical growth trajectories and GCM estimates suggested that although the overall cubic shapes of growth for MLUw, NDW, and WPM in Spanish were not identical, they generally reflected each other's general periods of acceleration (during kindergarten), deceleration (between spring of kindergarten and fall of second grade), and renewed acceleration (during second grade). Similarly, the overall linear shapes of growth for NDW and WPM in English, though not identical, generally reflected each other's consistent periods of positive, linear growth during each academic year. The within-language similarities in the overall shape of growth of the oral language measures suggested the possibility that morphosyntactic and lexical subsystems were connected growers, also reflected in their global integration. Previous findings of the positive within-language relationship between the development of morphological and lexical skills in ELLs (Bedore et al., 2006; Carias & Ingram, 2006; Castilla, et al., 2009; Conboy et al., 2006; Kohnert et al., 2010; Simon-Cereijido et al., 2009) support this qualitative observation.

Although MLUw, NDW, and WPM in Spanish demonstrated periods of acceleration during kindergarten and second grade, the overall acceleration of NDW-Spanish during kindergarten outpaced the kindergarten acceleration of MLUw- and WPM-Spanish. In contrast, the overall acceleration of MLUw- and WPM-Spanish during second grade outpaced the second grade acceleration of NDW-Spanish. Although no true steady states were found, there were periods of increased acceleration versus

periods of reduced acceleration. The fluctuating periods of acceleration within the Spanish oral language measures suggested the possibility of phase transitions across linguistic subsystems over time. Visual inspection of the English prototypical growth trajectories suggested that although MLUw, NDW, and WPM demonstrated consistent periods of positive, linear growth during each academic year, their growth during the summers differed. MLUw- and NDW-English demonstrated reduced growth during the summers, with MLUw-English showing considerable negative growth. In contrast, WPM-English demonstrated increased growth during the summers, relative to its growth during each academic year. The fluctuating periods of growth and decline within the English oral language measures suggested the possibility of phase transitions across linguistic subsystems over time. It is possible that this study captured differential periods of accelerated growth during which each language system was reorganizing itself (Perkins et al., 1996) as a function of competing growth from its own linguistic subsystems (e.g., lexicon versus morphosyntax in Spanish; lexicon versus morphosyntax in English).

Within-Subsystem Co-Development Across a Language

Visual inspection of the prototypical growth trajectories suggested that while NDW and WPM in Spanish demonstrated cubic growth trajectories (acceleration during kindergarten, deceleration during first grade, and renewed acceleration during second grade), NDW and WPM in English demonstrated positive, linear growth during each academic year and during the summers. The similarities between the periods of positive

growth of NDW and WPM across languages during kindergarten and second grade suggested the possibility that the lexical subsystem may have been a connected grower within the global integration of subsystems. Although previous findings of positive across-language relationships are not common (Castilla, et al., 2009; Conboy et al., 2006), they nevertheless support this qualitative observation. However, the differences between decelerating growth during first grade for NDW and WPM in Spanish, relative to the positive, linear growth during the summers for NDW and WPM in English suggested the possibility of phase transitions over time within linguistic subsystems.

Visual inspection suggested that although MLUw demonstrated positive growth in Spanish (cubic acceleration) and English (positive, linear growth) during kindergarten and second grade, MLUw-Spanish demonstrated deceleration between spring of kindergarten and fall of second grade. In contrast, MLUw-English demonstrated negative growth between the spring and fall semesters (summers), but positive growth during first grade. This qualitative observation suggested the possibility of phase transitions within the linguistic subsystems of morphosyntax across Spanish and English. In similar fashion to the reorganization of a language system in response to competing growth from its language subsystems, it is also possible that the differential periods of growth and decline found in this study were representative of the morphosyntactic subsystems (Spanish morphosyntax versus English morphosyntax) reorganizing themselves (Hohenberger & Peltzer-Karpf, 2009) as a function of competing growth from one another over time.

In summary, although the co-development of interconnected subsystems was not empirically confirmed, the qualitative observations of the available prototypical growth trajectories provided initial insight into the possible co-development occurring within- and across-languages.

Limitations

Despite the fact that the present study broke new ground by specifying the shape of growth for MLUw, NDW, and WPM in Spanish and English, it had a number of limitations. The BLLP longitudinal database, which provided this study with its final longitudinal dataset (6,516 Spanish and 5,732 English oral narrative language samples), was ironically its major limitation. Although the BLLP longitudinal database was arguably the most extensive longitudinal database (to date) of ELL children's narrative language in Spanish and English, compromises were required to carry out the present study. More than 1,500 narrative language samples were excluded from the study's longitudinal dataset because they did not meet the inclusionary criteria.

Although this study's inclusionary criteria resulted in trajectories of growth of typically developing, age-appropriate ELLs, the exclusion of 11% of the BLLP longitudinal database cannot be ignored. The data from excluded participants may have affected the results from this study, and warrant future consideration (e.g., subgroup of ELLs that are retained in the same grade). Another limitation is that a number of potentially important covariates (children's classroom placement with regard to language of instruction; children's language use with interlocutors; maternal education; age of first

and second language acquisition; amount and types of early language exposure at home) to include in GCM testing were not integrated in this study as they were either available for a limited subset of participants, or were not collected.

One methodological limitation is that, even though the growth of MLUw, NDW, and WPM was profiled in Spanish and English, this growth was modeled using the final longitudinal dataset as one population. Given the significant intra- and inter-individual differences identified in initial status and growth rates, and the size of the longitudinal dataset, the likelihood of multiple population subgroups (other than gender) contributing distinct patterns of growth to the prototypical growth trajectories modeled in this study was not considered. The limitations serve to guide and motivate future research that can augment the precision and the applicability of the current findings.

Future Directions

The present study's findings suggest that future work examining the Spanish and English language growth of ELLs needs to consider: the possibility of ELL subgroups and their unique growth trajectories, the time period to be studied and the frequency of measurement within, the systematic exposure to the native and second language throughout the duration of the study, and the number of ELLs to be followed over time.

The results from this study demonstrated that academic semester, gender, and schooling determined final growth curve models and prototypical growth trajectories for MLUw, NDW, and WPM in Spanish. Although the present study used a nascent

analytical approach (growth curve modeling; GCM), even more advanced analytical approaches for modeling growth are available. Future research on language growth will use growth mixture modeling (GMM), which integrates GCM and cluster analysis to conduct group-based growth modeling (Mueller & Hancock, 2010; Nagin, 2005; Wang & Bodner, 2007). GMM extends the capabilities of GCM, by modeling the growth of multiple subgroups, which are likely considering the significant heterogeneity of ELL children's oral language skills modeled in the present study. With GMM in place as the analytical approach to model growth in the future, the incorporation of additional covariates that can optimize model fit and predict growth is the next step.

The BLLP longitudinal database has information on children's classroom placement with regard to language of instruction; children's language use with interlocutors; and maternal education. These three time-invariant covariates, which are often identified as important factors in monolingual and bilingual language development (e.g., Beithcman et al., 2008; Dollaghan et al., 1999; Genesee et al., 2005; Hadle & Holt, 2006; Hammer et al., 2008a,b; Miller et al., 2006; Rice et al., 2009; Rojas et al., in preparation; Uchikoshi, 2006b), should be considered in future studies of language growth. Apart from the incorporation of additional covariates to model language growth, recent advances in computational linguistics in the near future should permit automated tagging of syntactic, lexical, and semantic features within transcribed language samples from ELL children (Gabani et al., 2009; Solorio & Liu, 2008). Automated tagging will facilitate the use of more specific outcome measures of morphosyntax (verbs, adverbs,

adjectives, nouns, pronouns, and articles), lexicon (content words, function words, words by lexical category), and oral fluency (syllables per minute) within transcribed language samples. Therefore, the growth of highly specific outcome measures (via automated tagging) can also be modeled in the future. Fine-grained measures may provide further insight, and perhaps clarify some of the nuances of growth trajectories found in this study.

From a clinical perspective, the findings from the present study can serve as an initial, grade-matched normative reference for MLUw, NDW, and WPM growth rates in the Spanish and English of typically developing ELL children from kindergarten to second grade. Such a normative reference could be used to inform assessment and intervention with ELL children with language impairments or at risk for language impairment (Rojas & Iglesias, 2010), by considering whether children's oral language skills are growing at comparable rates as typically developing children who are in the process of acquiring a second language. However, in order to increase the utility of this potential normative reference, future enhancements should include determining the true growth rates for children developing at 1 *SD* above and below the mean, and incorporating maternal education information. In addition, including a comparable normative reference for MLUw, NDW, and WPM growth rates from typically developing, English monolingual children would be useful. Researchers and clinicians would then have two anchor reference points for evaluating the growth rates of ELL children: (1) the Spanish and English growth rates from other ELL children, and (2) the

growth rates in English from English monolingual children. The availability of both normative references would be of great clinical utility for speech-language pathologists and for teachers of English as a second language, and should translate into better-informed clinical and educational decisions for ELLs.

REFERENCES CITED

Aarnoutse, C., & van Leeuwe, J. (2000). Development of poor and better readers during the elementary school. *Educational Research and Evaluation*, *6*(3), 251-278.

Abedi, J. (2006). Psychometric issues in the ELL assessment and special education eligibility. *Teachers College Record*, *108*(11), 2282-2303.

Alexander, K., Entwisle, D., & Olson, L. (2001). Schools, achievement, and inequality: A seasonal perspective. *Educational Evaluation and Policy Analysis*, *23*, 171-191.

Alexander, J.R.M., & Martin, F. (2004). The end of the reading age: Grade and age effects in early schooling. *Journal of School Psychology*, *42*(5), 403-416.

Anderson, R.T. (2004). First language loss in Spanish-speaking children: Patterns of language loss and implications for clinical practice. In B. Goldstein (Ed.), *Bilingual language development and disorders in Spanish-English speakers* (pp. 187-211). Baltimore: Paul H. Brookes.

Baraldi, A.N., & Enders, C.K. (2010). An introduction to modern missing data analyses. *Journal of School Psychology*, *48*, 5-37.

Bates, E., & Carnevale, G.F. (1993). New directions in research on language development. *Developmental Review*, *13*, 436-470.

Bates, E., Dales, P.S., & Thal, D. (1995). Individual differences and their implications for theories of language development. In P. Fletcher & B. MacWhinney (Eds.), *Handbook of child language* (pp. 96-151). Oxford, UK: Basil Blackwell.

Bauer, D.J., Goldfield, B.A., & Reznick, J.S. (2002). Alternative approaches to analyzing individual differences in the rate of early vocabulary development. *Applied Psycholinguistics*, *23*, 313-335.

Bedore, L.M. (2001). Assessing morphosyntax in Spanish-speaking children. In A. Iglesias (Ed.), *Communicative Assessment of the Hispanic Child*. *Seminars in Speech and Language*, *22*(1), 65-77.

Bedore, L.M. (1999). The acquisition of Spanish. In O. Taylor & L. Leonard (Eds.), *Language acquisition across North America: Cross-cultural and cross-linguistic perspectives* (pp. 157-208). San Diego: Singular.

Bedore, L.M., Fiestas, C., Peña, E.D., & Nagy, V. (2006). Cross-language comparisons of maze use in Spanish and English in functionally monolingual and bilingual children. *Bilingualism: Language and Cognition*, *9*(3), 233-247.

Bedore, L.M. & Leonard, L.B. (1998). Specific language impairment and grammatical morphology: A discriminant function analysis. *Journal of Speech, Language, and Hearing Research*, *41*, 1185-1192.

Bedore, L.M., Peña, E.D. (2008). Assessment of bilingual children for identification of language impairment: Current findings and implications for practice. *International Journal of Bilingual Education and Bilingualism*, *11*(1), 1-29.

Beitchman, J.H., Jiang, H., Koyama, E., Johnson, C.C., Escobar, M., & Atkinson, L., et al. (2008). Models and determinants of vocabulary growth from kindergarten to adulthood. *Journal of Childhood Psychology and Psychiatry*, *49*(6), 626-634.

Berman, R.A., & Slobin, D.I. (1994). *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale: Lawrence Erlbaum Associates.

Bleise, P.D., Chan, D., & Ployhart, R.E. (2007). Multilevel methods: Future directions in measurement, longitudinal analyses, and nonnormal outcomes. *Organizational Research Methods*, *10*(4), 551-563.

Bornstein, M.H., Hahn, C.S., & Haynes, O.M. (2004). Specific and general language performance across early childhood: Stability and gender considerations. *First Language*, *24*(3), 267-304.

Bouchard, C., Trudeau, N., Sutton, A., Boudreault, M.C., & Deneault, J. (2009). Gender differences in language development in French Canadian children between 8 and 30 months of age. *Applied Psycholinguistics*, *30*, 685-707.

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

Burchinal, M., Nelson, L., & Poe, M. (2006). Growth curve analysis: An introduction to various methods for analyzing longitudinal data. In K.L. Bub, M. Burchinal, & K.

McCarthey (Eds.), *Best practices in quantitative methods for developmentalists*, *71*(3) (pp. 65-87). Indianapolis: Wiley-Blackwell.

Carias, S., & Ingram, D. (2006). Language and disfluency: Four case studies on Spanish-English bilingual children. *Journal of Multilingual Communication Disorders*, *4*(2), 149-157.

Castilla, A.P., Restrepo, A., & Perez-Leroux, A.T. (2009). Individual differences and language independence: A study of sequential bilingual development in Spanish-English preschool children. *International Journal of Bilingual Education and Bilingualism*, *12*(5), 565-580.

Cheung, M.W.L. (2007). Comparison of methods of handling missing time-invariant covariates in latent growth models under the assumption of missing completely at random. *Organizational Research Methods*, *10*(4), 609-634.

Coates, J. (1993). The acquisition of gender-differentiated language. In J. Coates (Ed.), *Women, men, and language: A sociolinguistic account of gender differences in language* (2nd ed., pp. 143-167). London, UK: Longman.

Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research, 66*, 227-268.

Crone, D.A., & Whitehurst, G.J. (1999). Age and schooling effects on emergent literacy and reading skills. *Journal of Educational Psychology, 91*(4), 604-614.

Cromley, J. (2008, Fall). ED PSY 8960: *Problems in Educational Psychology: Growth curve analysis for longitudinal data*. Philadelphia: Temple University.

De Bot, K., Lowie, W., & Verspoor, M. (2007b). A dynamic view as a complementary perspective. *Bilingualism: Language and Cognition*, *10*(1), 51-55.

De Bot, K., Lowie, W., & Verspoor, M. (2007a). A dynamic systems theory approach to second language acquisition. *Bilingualism: Language and Cognition*, *10*(1), 7-21.

de Villiers, J.G. & de Villiers, P.A. (1973). A cross-sectional study of the acquisition of grammatical morphemes in child speech. *Journal of Psycholinguistic Research*, *2*(3), 267-278.

De Von Figueroa-Moseley, C., Ramey, C., Keltner, B., & Lanzi, R. (2006). Variations in Latino practices and their effects on child cognitive developmental outcomes. *Hispanic Journal of Behavioral Sciences*, *28*, 102-114.

Ding, C.S., & Davison, M.L. (2005). A longitudinal study of math achievement gains for initially low achieving students. *Contemporary Educational Psychology*, *30*(1), 81-95.

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.

Eisenberg, S.L., Fersko, T.M., Lundgren, C. (2001). The use of MLU for identifying language impairment in preschool children. *American Journal of Speech-Language Pathology*, *10*, 323-342.

Elman, J.L., Bates, E.A., Johnson, M.H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1999). *Rethinking innateness: A connectionist perspective on development*. Cambridge: MIT Press.

Evans, J. (2001). An emergent account of language impairments in children with SLI: Implications for assessment and intervention. *Journal of Communication Disorders*, *34*, 39-54.

Farkas, G., & Beron, K. (2004). The detailed age trajectory of oral vocabulary knowledge: Differences by class and race. *Social Science Research*, *33*, 464-497.

Fenson, L., Bates, E., Dale, P., Goodman, J., Reznick, J.S., & Thal, D.J. (2000). Measuring variability in early child language: Don't shoot the messenger. *Child Development*, *71*(2), 323-328.

Fenson, L., Dale, P.S., Reznick, J.S., Bates, E., Thal, D.J., & Pethick, S.J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, *59*(5), 40-185.

Fenson, L., Dale, P.S., Reznick, J.S., Thal, D., Bates, E., Hartung, J.P., et al. (1993). *The MacArthur Communicative Development Inventories: User's guide and technical manual*. San Diego: Singular.

Fichman, M., & Cummings, J.N. (2003). Multiple imputation for missing data: Making the most of what you know. *Organizational Research Methods*, *6*(3), 282-308.

Francis, D.J., Carlson, C.D., Fletcher, J.M., Foorman, B.R., Goldenberg, C.R., & Vaughn, S., et al. (2005). Oracy/Literacy Development of Spanish-speaking Children: A multi-level program of research on language minority children and the instruction, school and community contexts, and interventions that influence their academic outcomes. *Perspectives: The International Dyslexia Association*, *31*(2), 8-12.

Francis, D.J., Snow, C.E., August, D., Carlson, C.D., Miller, J., & Iglesias, A. (2006). Measures of reading comprehension: A latent variable analysis of the diagnostic assessment of reading comprehension. *Scientific Studies of Reading, 10*(3)*,* 301-322.

Gabani, K., Sherman, M., Solorio, T., Liu, Y., Bedore, L., & Peña, E. (2009). A corpus-based approach for the prediction of language impairment in monolingual English and Spanish-English bilingual children. In: *North American Chapter of the Association for Computational Linguistics – Human Language Technologies* (pp. 46-55). NAACL HLT 2009: Association for Computational Linguistics. Boulder, CO.

Genesee, F., Lindholm-Leary, K., Saunders, W., & Christian, D. (2005). English language learners in US schools: An overview of research findings. *Journal of Education for Students Placed at Risk, 10*(4)*,* 363-385.

Gleason, J.B., & Ratner, N.B. (2009). *The development of language* (7th ed.). New York: Pearson.

Golberg, H., Paradis, J., & Crago, M. (2008). Lexical acquisition over time in minority first language children learning English as a second language. *Applied Psycholinguistics*, *29*, 41-65.

Graham, J.W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*, 549-576.

Gutiérrez-Clellen, V.F., & Hofstetter, R. (1994). Syntactic complexity in Spanish narratives: A developmental study. *Journal of Speech and Hearing Research, 37*, 645-654.

Gutiérrez-Clellen, V.F., Restrepo, M.A., Bedore, L., Peña, E., & Anderson, R. (2000). Language sample analysis in Spanish-speaking children: Methodological considerations. *Language, Speech, and Hearing Services in Schools, 31*, 88-98.

Hammer, C.S., Farkas, G., Maczuga, S. (2010). The language and literacy development of Head Start children: A study using the family and child experiences database. *Language, Speech, and Hearing Services in Schools*, *41*, 70-83.

Hammer, C.S., Lawrence, F.R., & Miccio, A.W. (2008b). The effect of summer vacation on bilingual preschoolers' language development. *Clinical Linguistics & Phonetics*, *22*(9), 686-702.

Hammer, C.S., Lawrence, F.R., & Miccio, A.W. (2008a). Exposure to English before and after entry into Head Start: Bilingual children's receptive language growth in Spanish and English. *International Journal of Bilingual Education and Bilingualism*, *11*(1), 30-56.

Hammer, C.S., Lawrence, F.R., & Miccio, A.W. (2007). Bilingual children's language abilities in early reading outcomes in head start and kindergarten. *Language, Speech, and Hearing Services in Schools*, *38*, 237-248.

Heilmann, J., Miller, J.F., Iglesias, A., Fabiano-Smith, L., Nockerts, A., & Andriacchi, K.D. (2008). Narrative transcription accuracy and reliability in two languages. *Topics in Language Disorders*, *28*(2), 178-187.

Heilmann, J., Miller, J.F., & Nockerts, A. (2010). Using language sample databases. *Language, Speech, and Hearing Services in Schools*, *41*, 81-75.

Herdina, P., & Jessner, U. (2002). *A dynamic model of multilingualism*. Clevedon, UK: Multilingual Matters.

Hoff, E. (2005). *Language development* (3rd ed.). Belmont: Wadsworth.

Hohenberger, A., & Peltzer-Karpf, A. (2009). Language learning from the perspective of nonlinear dynamic systems. *Linguistics*, *47*(2), 481-511.

Hopstock, P.J., & Stephenson, T.G. (2003). *Descriptive study of services to LEP students and LEP students with disabilities* [Special Topic Report #1]. U.S. Department of Education: Office of English Language Acquisition, Language Enhancement and Academic Achievement of LEP Students.

Howe, M.L., & Lewis, M.D. (2005). The importance of dynamic systems approaches for understanding development. *Developmental Review*, *25*, 247-251.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology*, *27*(2), 236-248.

Iglesias, A. (2004, June). *Narrative language skills and reading achievement in bilingual children*. Paper presented at Symposium on Research in Child Language Disorders, Madison, WI.

Iglesias, A., Miller, J., Jones, P., Hanlon, A., & Francis, D. (in preparation). An empirically based typology of bilingual English language learners.

Justice, L.M. (2009). Increasing your statistical edge: Dealing with nested data structures. *American Journal of Speech-Language Pathology*, *18*, 210-211.

Kan, P.F. (2010). Measuring word learning ability in sequential bilingual children. *Perspectives on Communication Disorders and Sciences in Culturally and Linguistically Diverse Populations*, *17*(1), 25-32.

Kelso, S.J.A. (1994). *Dynamic patterns: The self-organization of brain and behavior*. Cambridge: MIT Press.

Kohnert, K. (2008). *Language disorders in bilingual children and adults*. San Diego: Plural Publishing.

Kohnert, K., Kan, P.F., & Conboy, B.T. (2010). Lexical and grammatical associations in sequential bilingual preschoolers. *Journal of Speech, Language, and Hearing Research*, *53*, 684-698.

Laija-Rodriguez, W., Ochoa, S.H., & Parker, R. (2006). The crosslinguistic role of cognitive academic language proficiency on reading growth in Spanish and English. *Bilingual Research Journal*, *30*(1), 87-106.

Larsen-Freeman, D. (2007). On the complementarity of Chaos/Complexity Theory and Dynamic Systems Theory in understanding the second language acquisition process. *Bilingualism: Language and Cognition*, *10*(1), 35-37.

Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics*, *27*(4), 590-619.

Law, J., Tomblin, B.J., & Zhang, X. (2008). Characterizing the growth trajectories of language-impaired children between 7 and 11 years of age. *Journal of Speech, Language, and Hearing Research*, *51,* 739-749.

Leonard, L.B. (2000). *Children with specific language impairment*. Cambridge: The MIT Press.

Luke, D.A. (2004). *Multilevel modeling*. Thousand Oaks: Sage Publications.

Marin, M.M., & Peltzer-Karpf, A. (2009). Towards a dynamic systems approach to the development of language and music: Theoretical foundations and methodological issues. In: J.Louhivuori, T. Eerola, S. Saarikallio, T. Himberg, & P.S. Eerola (Eds.) *Conference of European Society for the Cognitive Sciences of Music* (pp. 284-292). ESCOM 2009: Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music; Jyväskylä, Finland, August 12-16, 2009. Jyväskylä: University of Jyväskylä.

Mayer, M. (1974). *Frog Goes To Dinner.* New York: Dial Press.

Mayer, M. (1975b). *Frog On His Own.* New York: Dial Press.

Mayer, M. (1975a). *One Frog Too Many.* New York: Dial Press.

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

McCarthy, D. A. (1930). The language development of the pre-school child. *University of Minnesota Institute of Child Welfare Monograph Series IV*. Minneapolis:

University of Minnesota Press.

McCoach, D.B., O'Connell, A., Reis, S.M., & Levitt, H. (2006). Growing readers: A hierarchical linear model of children's reading growth during the first 2 years of school. *Journal of Educational Psychology*, *98*, 14-28.

McHale, S., Updegraff, K., Shanahan, L., Crouter, A., & Killoren, S. (2005). Siblings' differential treatment in Mexican American families. *Journal of Marriage and Family*, *67*, 1259-1274.

McMurray, B. (2007). Defusing the childhood vocabulary explosion. *Science*, *317*, 631.

Microsoft Corporation. (2007). Microsoft Excel 2008 for Mac, Version 12.2.4.

Miller, J.F. (1987). A grammatical characterization of a language disorder. *Proceedings of the First International Symposium on Specific Speech and Language Disorders in Children*, London, England.

Miller, J.F., & Chapman, R.S. (1981). The relation between age and mean length of utterance in morphemes. *Journal of Speech and Hearing Research*, *42*, 154-161.

Miller, J.F., & Heilmann, J. (2004, February). *Bilingual language project update*. Paper presented at the Department of Communicative Disorders Colloquium, Madison, WI.

Miller, J., Heilmann, J., Nockerts, A., Iglesias, A, Fabiano, L., & Francis, D. (2006). Oral language and reading in bilingual children. *Learning Disabilities Research and Practice, 21*(1)*,* 30-43.

Miller, J.F., & Iglesias, A. (2010). *Systematic analysis of language transcripts (SALT), Research Version 2010* [Computer software]. SALT Software, LLC.

Miller, J., & Klee, T. (1995). Computational approaches to the analysis of language impairment. In P. Fletcher & B. MacWhinney (Eds.), *Handbook of child language* (pp. 545-572). Oxford, UK: Basil Blackwell.

Morrison, F.J., Griffith, E.M., & Alberts, D.M. (1997). Nature-nurture in the classroom: Entrance age, school readiness, and learning in children. *Developmental Psychology*, *33*, 254-262.

Mueller, R.O., & Hancock, G.R. (2010). Structural equation modeling. In G.R. Hancock & R.O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (pp. 371-383). New York: Routledge.

Nagin, D.S. (2005). *Group-based modeling of development*. Cambridge: Harvard University Press.

Nakamoto, J., Lindsey, K.A., & Manis, F.R. (2007). A longitudinal analysis of English language learners' word decoding and reading comprehension. *Reading and Writing*, *20*, 691-719.

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

Özçalişkan, S., & Slobin, D.I. (1999). Learning how to search for the frog: Expression of manner of motion in English, Spanish, and Turkish. In A. Greenhill, H. Littlefield, & C. Tano (Eds.), *Proceedings of the 23rd Annual Boston University Conference on Language Development: Vol. 2* (pp. 541-552). Somerville: Cascadilla Press.

Pan, B.A., Rowe, M.L., Singer, J.D., & Snow, C.E. (2005). Maternal correlates of growth in toddler vocabulary production in low-income families. *Child Development*, *76*(4), 763-782.

Papafragou, A., Massey, C., Gleitman, L. (2002). Shake, rattle, 'n' roll: The representation of motion in language and cognition. *Cognition*, *84*, 189-219.

Paul, R. (2006). *Language disorders from infancy through adolescence: Assessment and intervention* (3rd ed.). St. Louis: Mosby.

Pavlenko, A. (2009). Verbs of motion in L1 Russian of Russian-English bilinguals. *Bilingualism: Language and Cognition*, *13*(1), 49-62.

Peña, E., Iglesias, A., & Lidz, C.S. (2001). Reducing test bias through dynamic assessment of children's word learning ability. *American Journal of Speech-Language Pathology*, *10*, 138-154.

Perkins, K., Brutten, S.R., & Grass, S.M. (1996). An investigation of patterns of discontinuous learning: Implications for ESL measurement. *Language Testing*, *13*(1), 63-82.

Port, R., & van Gelder, T. (1995). *Mind as motion: Dynamical perspectives on behavior and cognition*. Cambridge: MIT Press.

Price, L.H., Hendricks, S., & Cook, C. (2010). Incorporating computer-aided language sample analysis into clinical practice. *Language, Speech, and Hearing Services in Schools*, *41*, 206-222.

Ram, N., & Grimm, K. (2007). Using simple and complex growth models to articulate developmental change: Matching theory to method. *International Journal of Behavioral Development*, *31*(4), 303-316.

Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks: Sage Publications.

Rice, M.L., Hoffman, L., & Wexler, K. (2009). Judgments of omitted BE and DO in questions as extended finiteness clinical markers of specific language impairment (SLI) to 15 years: A study of growth and asymptote. *Journal of Speech, Language, and Hearing Research*, *52*, 1417-1433.

Rice, M.L., Redmond, S.M., & Hoffman, L. (2006). Mean length of utterance in children with specific language impairment and in younger control children shows

concurrent validity and stable and parallel growth trajectories. *Journal of Speech, Language, and Hearing Research*, *49*, 793-808.

Rice, M.L., Wexler, K., & Hershberger, S. (1998). Tense over time: The longitudinal course of tense acquisition in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *41*, 1412-1431.

Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse Processes*, *14*, 423-441.

Rogosa, D.R., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin*, *90*, 726-748.

Rojas R., & Iglesias, A. (2010). Using language sampling to measure language growth. *Perspectives on Language Learning and Education*, *17*(1), 24-31.

Rojas, R., & Iglesias, A. (2009, March 3). Making a case for language sampling. *The ASHA Leader, 14*(3), 10–13.

Rojas R., & Iglesias, A. (2006). Bilingual (Spanish-English) narrative language analyses: Why and how? *Perspectives on Communication Disorders and Sciences in Culturally and Linguistically Diverse Populations*, *13*(1), 3-8.

Rojas R., Iglesias, A., Bunta, F., Goldstein, B., Goldenberg, C., & Reese, L. (in preparation). Interlocutor differential effects on bilingual children's expressive language skills.

Sebastián, E., & Slobin, D.I. (1994). Development of linguistic forms: Spanish. In R. A. Berman & D.I. Slobin (Eds.), *Relating events in narrative: A crosslinguistic developmental study* (pp. 239-284). Hillsdale: Lawrence Erlbaum Associates.

Schafer, J.L., & Graham, J.W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147-177.

Shriner, T. H. (1969). A review of mean length of response as a measure of expressive language development in children. *Journal of Speech and Hearing Disorders*, *34*, 61–68.

Siegler, R.S. (2002). Variability and infant development. *Infant Behavior and Development*, *25*, 550-557.

Silliman, E.R., Bahr, R.H., Brea, M.R., Hnath-Chisolm, T., & Mahecha, N.R. (2002). Spanish and English proficiency in the linguistic encoding of mental states in narrative retellings. *Linguistics and Education*, *13*(2), 199-234.

Simon-Cereijido, G., & Gutiérrez-Clellen, V.F. (2009). A cross-linguistic and bilingual evaluation of the interdependence between lexical and grammatical domains. *Applied Psycholinguistics*, *30*, 315-337.

Singer, J.D., & Willett, J.B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.

Smit, A.B., Hand, L., Freilinger, J.J., Bernthal, J.E., & Bird, A. (1990). The Iowa articulation norms project and its Nebraska replication. *Journal of Speech and Hearing Disorders*, *55*, 779-798.

Smith, L.B., & Thelen, E. (2003). Development as a dynamic system. *Trends in Cognitive Sciences*, *7*(8), 343-348.

Solorio, T., & Liu, Y. (2008). Using language models to identify language impairment in Spanish-English bilingual children. *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing* (pp. 116-117). In: Association for Computational Linguistics. Columbus, OH.

SPSS Inc. (2010). IBM SPSS Statistics 18.0 for Mac.

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

Swasey Washington, P. (2009). *Factors affecting norming: A developmental study of oral language measures in Spanish-speaking English language learners*. PhD dissertation, Department of Communication Sciences and Disorders, Temple University, Philadelphia, PA.

Swanson, C.B. (2009). *Perspectives on a population: English-language learners in American schools*. Bethesda: Editorial Projects in Education.

Tabachnick,B.G., & Fidell, L.S. (2007). *Using multivariate statistics* (5th ed.). Boston: Pearson Educational.

Templin, M. (1957). Certain language skills in children. *University of Minnesota Institute of Child Welfare Monograph Series 26*. Minneapolis: University of Minnesota Press.

Thelen, E., & Bates, E. (2003). Connectionism and dynamic systems: Are they really different? *Developmental Science*, *6*(4), 378-391.

Thelen, E., & Smith, L.B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge: MIT Press.

Thelen, E., & Ulrich, B.D. (1991). Hidden skills: A dynamic systems analysis of treadmill stepping during the first year. *Monographs of the society for research in child development*, *56*(1), 1-104.

Trista, J., McMaster, K. (2007). Productivity, Fluency, and grammaticality measures from narratives: Potential indicators of language proficiency? *Communication Disorders Quarterly*, *29*(1), 43-53.

Toronto, A.S. (1976). Developmental assessment of Spanish grammar. *Journal of Speech and Hearing Disorders*, *41*, 150-171.

Uccelli, P., & Paéz, M.M. (2007). Narrative and vocabulary development of bilingual children from kindergarten to first grade: Developmental changes and associations among English and Spanish skills. *Language, Speech, and Hearing Services in Schools*, *38*, 225-236.

Uchikoshi, Y. (2006b). English vocabulary development in bilingual kindergartners: What are the best predictors? *Bilingualism: Language and Cognition*, *9*(1), 33-49.

Uchikoshi, Y. (2006a). Early reading in bilingual kindergartners: Can educational television help? *Scientific Studies of Reading*, *10*(1), 89-120.

van Geert, P. (2008). The dynamic systems approach in the study of L1 and L2 acquisition: An introduction. *The Modern Language Journal*, *92*(2), 179-199.

van Geert, P. (2007). Dynamic systems in second language learning: Some methodological reflections. *Bilingualism: Language and Cognition*, *10*(1), 47-49.

van Geert, P. (1998). A dynamic systems model of basic developmental mechanisms: Piaget, Vygotsky, and beyond. *Psychological Review*, *5*, 634-677.

van Geert, P. (1994). *Dynamic systems of development: Change between development and chaos*. New York: Harvester Wheatsheaf.

van Geert, P., & van Dijk, M. (2002). Focus on variability: New tools to study intra-individual variability in development data. *Infant Behavior and Development*, *25*, 340-374.

Velten, H.V. (1943). The growth of phonemic and lexical patterns in infant language. *Language*, *19*(4), 281-292.

Wang, M., & Bodner, T.E. (2007). Growth mixture modeling: Identifying and predicting unobserved subpopulations with longitudinal data. *Organizational Research Methods*, *10*(4), 635-656.

Weismer, S.E., Murray-Branch, J., & Miller, J.F. (1994). A perspective longitudinal study of language development in late talkers. *Journal of Speech and Hearing Research*, *37*, 852-867.

Willet, J.B. (1989). Some results on reliability for the longitudinal measurement of change: Implications for the design of studies of individual growth. *Educational and Psychological Measurement*, *49*, 587-602.

Wu, W., West, S.G., & Hughes, J.N. (2008). Effect of retention in first grade on children's achievement trajectories over 4 years: A piecewise growth analysis using propensity score matching. *Journal of Educational Psychology*, *100*(4), 727-740.

Yip, V., & Matthews, S. (2006). Assessing language dominance in bilingual acquisition: A case for mean length utterance differentials. *Language Assessment Quarterly*, *3*(2), 97-116.

Zentella, A. (1997). *Growing up bilingual*. Malden: Blackwell.

Zwietering, M.H., Jongenburger, F.M., Rombouts, F.M., & Riet, K.V. (1990). Modeling of the bacterial growth curve. *Applied and Environmental Microbiology*, *56*(6), 1875-1881.

APPENDICES

APPENDIX A

GLOSSARY OF GREEK TERMS (CROMLEY, 2008)

GREEK TERM	ENGLISH DEFINITION
$\mathbf{Y_{ij}}$	Each person's (i) score at each time point (j)
$\boldsymbol{\pi}$	Level 1 (over time) coefficients
$\boldsymbol{\gamma}$	Level 2 coefficients (for person or environmental characteristics)
$\boldsymbol{\gamma_{00}}$	Grand mean intercept (when all other predictors, such as Male, Group, etc. are at 0). Could also be described as an unconditional intercept.
$\boldsymbol{\gamma_{01}}$	Effect of a substantive predictor (effect of X) on the intercept. Could also be described as the differential in the intercept for being in Group = 1 for a 0/1 categorical predictor.
$\boldsymbol{\gamma_{10}}$	Grand mean slope (when all other predictors, such as Male, Group, etc. are at 0). Could also be described as an unconditional slope.
$\boldsymbol{\gamma_{11}}$	Effect of a substantive predictor (effect of X) on the slope. Could also be described as the differential in the slope for being in Group = 1 for a 0/1 categorical predictor.
$\boldsymbol{\varepsilon_{ij}}$	The residual around a fitted line for each person's (i) score at each time point j; Level-1 error term
$\boldsymbol{\zeta_{0i}}$	The residual between each person's (i) actual intercept and the γ_{00}; Level-2 error term
$\boldsymbol{\zeta_{1i}}$	The residual between each person's (i) actual slope and the γ_{10}; Level-2 error term
$\boldsymbol{\sigma}$	Variances and covariances of Level-1 (ε) and Level-2 (ζ) error terms

APPENDIX B

THE UNCONDITIONAL MEANS (UM) MODEL

A sample UM model formula:

Level-1: Within-person variation $Y_{ij} = \pi_{0i} + \varepsilon_{ij}$

Level-2: Between-person variation $\pi_{0i} = \gamma_{00} + \zeta_{0i}$

At Level-1 ($Y_{ij} = \pi_{0i} + \varepsilon_{ij}$), $\mathbf{Y_{ij}}$ is the outcome measure for each participant (subscript ***i***) at each wave of measurement (subscript ***j***). The outcome measure is composed of the expected true intercept (i.e., initial status) of each participant $\boldsymbol{\pi_{0i}}$, and the Level-1 (within-person) residual variance ($\boldsymbol{\varepsilon_{ij}}$ [σ_ε^2] term, *variance component*) across all time points. Level-2 ($\pi_{0i} = \gamma_{00} + \zeta_{0i}$) attempts to simultaneously account for group-level differences and to predict differences generated from Level-1. For the UM-model, this includes the intercept for each participant $\boldsymbol{\pi_{0i}}$ (notice that this parameter comes from Level-1 and is now an outcome to be predicted), which is composed of the mean sample intercept ($\boldsymbol{\gamma_{00}}$ term, *fixed effect*), and Level-2 (between-person) residual variance ($\boldsymbol{\zeta_{0i}}$ [σ_0^2] term, *variance component*), which is the residual between each person's Level-1 intercept (π_{0i}) and the mean sample intercept (γ_{00}). Therefore in the UM-model, each participant's outcome measure is a combination of:

- γ_{00} = sample mean intercept (i.e., initial status)
- Two sources of residual variance

- ε_{ij} [σ_{ε}^2] = Level-1 within-person residual variance over time
- ζ_{0i} [σ_0^2] = Level-2 between-person residual variance in intercepts

APPENDIX C

THE UNCONDITIONAL GROWTH (UG) MODEL

A sample UG model formula:

Level-1: Within-person variation $Y_{ij} = \pi_{0i} + \pi_{1i} (TIME_{ij}) + \varepsilon_{ij}$

Level-2: Between-person variation $\pi_{0i} = \gamma_{00} + \zeta_{0i}$

$\pi_{1i} = \gamma_{10} + \zeta_{1i}$

At Level-1 ($Y_{ij} = \pi_{0i} + \pi_{1i} (TIME_{ij}) + \varepsilon_{ij}$), the difference from the UM-model, is that the UG-model adds the effect of time. This is the expected true slope $\boldsymbol{\pi_{1i}}$ (i.e., growth rate) experienced by each participant over linear time ($\mathbf{Time_{ij}}$). The interaction of the intercept $\boldsymbol{\pi_{0i}}$ (i.e., initial status) and the slope $\boldsymbol{\pi_{1i}}$ (i.e., rate of change) x $\mathbf{Time_{ij}}$ (time) determines true growth trajectories for each participant in a longitudinal dataset. Level-2 of the UG ($\pi_{0i} = \gamma_{00} + \zeta_{0i}$; $\pi_{1i} = \gamma_{10} + \zeta_{1i}$) model adds the slope for each participant $\boldsymbol{\pi_{1i}}$ (again, this parameter comes from Level-1) which is composed of the mean slope (i.e., constant growth rate in a linear model; instantaneous acceleration in a quadratic or cubic model at initial status) for the entire sample ($\boldsymbol{\gamma_{10}}$ term, *fixed effect*), and the Level-2 between-person residual variance ($\boldsymbol{\zeta_{1i}}$ [σ_1^2] term, *variance component*), which is the residual between each person's Level-1 (π_{1i}) slope and the mean sample slope (γ_{10}). The mean slope (γ_{10}) is the growth rate for the participants in the sample. Therefore in the UG-model, each participant's outcome measure is a combination of:

- γ_{00} = Sample mean intercept (i.e., initial status)

- γ_{10} = Sample mean slope (i.e., growth rate)

- Four sources of residual variance

    - ε_{ij} [σ_{ε}^2] = Level-1 within-person residual variance over time

    - ζ_{0i} [σ_0^2] = Level-2 between-person residual variance in intercepts

    - ζ_{1i} [σ_1^2] = Level-2 between-person residual variance in slopes

    - [σ_{01}] = Level-2 residual covariance between the true intercept (π_{0i}) and true slope (π_{1i}) across all participants in the sample

APPENDIX D

THE CONDITIONAL GROWTH (CG) MODEL

A sample CG model formula:

Level-1: Within-person variation $Y_{ij} = \pi_{0i} + \pi_{1i} (TIME_{ij}) + \varepsilon_{ij}$

Level-2: Between-person variation $\pi_{0i} = \gamma_{00} + \gamma_{01} (GENDER_i) + \zeta_{0i}$

$\pi_{1i} = \gamma_{10} + \gamma_{11} (GENDER_i) + \zeta_{1i}$

At Level-1, given that a time-varying covariate is not added, the UG and the CG-models are the same. At Level-2 ($\pi_{0i} = \gamma_{00} + \gamma_{01} (GENDER_i) + \zeta_{0i}$; $\pi_{1i} = \gamma_{10} + \gamma_{11} (GENDER_i) + \zeta_{1i}$), the difference from the UG-model, is that the CG-model adds the mean effect of a time-invariant covariate (i.e., gender, which can be specified as the effect of being female or male) on the intercept ($\boldsymbol{\gamma_{01}}$ term, *fixed effect*) and/or on the slope ($\boldsymbol{\gamma_{11}}$ term, *fixed effect*). The mean effect of adding a time-invariant (or a time-varying) covariate can further explain intercept and/or slope variance. Therefore in the CG-model, each participant's outcome measure is a combination of:

- γ_{00} = Sample mean intercept (i.e., initial status)
- γ_{01} = Mean effect of an additional covariate (e.g., gender) on intercept
- γ_{10} = Sample mean slope (i.e., growth rate)
- γ_{11} = Mean effect of an additional covariate (e.g., gender) on slope

- This parameter is optional

- Four sources of residual variance
  - ε_{ij} [σ_{ε}^2] = Level-1 within-person residual variance over time
  - ζ_{0i} [σ_0^2] = Level-2 between-person residual variance in intercepts
  - ζ_{1i} [σ_1^2] = Level-2 between-person residual variance in slopes
  - [σ_{01}] = Level-2 residual covariance between the true intercept (π_{0i}) and true slope (π_{1i}) across all participants in the sample