MULTISYLLABIC WORD PRODUCTION OF SCHOOL-AGED CHILDREN WITH AND WITHOUT PROTRACTED PHONOLOGICAL DEVELOPMENT

by

Glenda Mason

BSc, The University of Alberta. 1976

MSLP, The University of Alberta. 1997

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Audiology and Speech Sciences)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

December 2015

Abstract

Few studies exist concerning multisyllabic word development, yet researchers suggest that evaluation of long words is essential for phonological assessment of school-aged children (e.g. Holm, Farrier, & Dodd, 2008; James, 2006). Furthermore, multisyllabic word (MSW) production may be related to literacy development (Carroll & Snowling, 2004). The goal of this thesis therefore was to further study the phonological acquisition of MSWs.

The dissertation begins with a meta-analysis (Chapter 2) that examines factors related to word level literacy skills in children with and without protracted phonological development (PPD). Fifty-two eligible studies were evaluated that had measured at least one word level literacy construct. Sixty-four independent samples were generated for evaluation of: *Phonological awareness*, *Decoding*, *Fluency*, *Nonword decoding* and *Spelling.* A *mean mixed weighted d* was compared on the *Q*-statistic using a random effects model. MSW evaluation was a relevant moderator of literacy for children with PPD as expected and motivated the remainder of the thesis.

A pilot study (Chapter 3) evaluated a new MSW metric based on nonlinear phonological and language processing frameworks. Six MSWs were analyzed for ten English-speaking typically developing (TD) 5-year-olds, and eight French-speaking 3- to 4-year-olds with PPD. Mismatches were tallied and ranked, and compared with tallies from Phonological Mean Length of Utterance (Ingram, 2002) and Percent Consonants Correct (Shriberg, Austin, Lewis, McSweeny, & Wilson, 1997). The number of different ranks was significant and systematically higher with the new metric.

Chapter 4 examines production of 20 MSWs in three sub-studies: (a) longitudinal: 44 TD children at ages 5 and 8 to 10 years; (b) cross-sectional by age: twelve age-matched 8- to 10-year-olds with and without PPD; and (c) cross-sectional by group: 62 TD 5-year-olds and the 12 8- to 10-year-olds with PPD. Lexical and phonological tallies decreased significantly between ages 5 and 8 years for the TD children. The 8- to 10-year-olds with PPD showed overall scores equivalent to those of the TD 5-year-olds although some different mismatch (error) patterns. In summary, the thesis further examines the link between MSWs and literacy and provides both a new whole-word metric for MSWs and phonological data for school-aged children.

Preface

The research presented in this thesis is original in conception, with research and writing for three chapters done by the author (Chapters 1, 4 and 5) and two (Chapters 2 and 3) done in collaboration with colleagues as indicated below.

A version of Chapter 2 has been published [Mason, G. K., & Bernhardt, B. M. 2014. The impact of protracted phonological disorders on literacy outcomes in children: A meta-analysis. In P. A. Ysunza (Ed.), *Speech, language and voice pathology: Methods, challenges and outcomes* (pp. 49-116). New York: Nova Publishers.]. I was the lead investigator, responsible for all major areas of concept formation, data collection and analysis and manuscript composition. Dr. Bernhardt, as supervisor for my dissertation, provided manuscript edits.

A version of Chapter 3 was published also [Mason, G. K., Bérubé, D., Bernhardt, B. M., & Stemberger, J. P. (2015). Evaluation of multisyllabic word production in Canadian English- and French-speaking children within a nonlinear phonological framework. *Clinical Linguistics and Phonetics, 29*(8-10), 666-685. doi 10.3109/02699206.2015.1040894.] I was the lead investigator, and was responsible for concept formation, and the majority of analysis and manuscript composition. Dr. Bérubé collected the French data and assisted with analysis and writing of the French part of the study. Drs. Bernhardt and Stemberger, as supervisor and committee member for my dissertation respectively, provided feedback on the evaluation methods and manuscript edits.

The research was conducted in accordance with the Behavioural Research Ethics Board approvals: Certificate #B06-0397, H11-00619 and H09-03206.

Table of Contents

List of Tables

List of Figures

Acknowledgements

The first acknowledgement must be of my supervisor, Dr. (B.) May Bernhardt. Dr. Bernhardt has mentored me along a path in the study of child phonology that began before 2009, and subsequently led to a PhD program of study. Dr. Bernhardt has shared a unique breadth and depth of knowledge, skills, and wisdom that have guided this dissertation. Her expertise and support were provided with a spirit of generosity and compassion such that every step could unfold. I am very grateful and humbled to have had this opportunity, to be under Dr. Bernhardt's patient and steady supervision.

I would also like to thank committee members Dr. J. A. Small and Dr. Joseph P. Stemberger (Linguistics). They have provided continuous direction and encouragement to think about the theories and data for the thesis and their support throughout the process is highly valued.

Another acknowledgement is due to a member of the original supervisory committee, Dr. Julie J. Masterson of Missouri State University. Dr. Masterson provided a link to early literacy, and supported dissertation projects from across the miles.

I would next like to acknowledge Dr. Nand Kishor, who provided valuable and patient instruction in statistics. Dr. Kishor empowered the discovery and practice of the statistical procedures needed to support the dissertation research.

The acknowledgements turn now to others less formally associated with the dissertation. The first is my friend and colleague, Dr. Daniel Bérubé, and secondly, Bonnie J. Craig, Director of Dental Hygiene (retired), who have been supports and mentors over the years.

I also offer sincere thanks to Sarah Wright, Christine Schretlen, Valerie Marshall, Sara Skolsky, Lauren Butters, Robyn Gordon and Jamie Russell for completing high quality work. I must also thank Christine Schretlen for willing collaborations and insights on our respective theses.

Funding sources must also be acknowledged: the Hampton Fund, the Pacific Century Graduate Scholarships, the University of British Columbia Faculty of Medicine Graduate Award, Gertrude Langridge Graduate Scholarship in Medical Sciences, Shaughnessy Hospital Volunteer Society Fellowship in Health Care, John R. Grace Fellowship, and the Kathleen Simpson Memorial Graduate Scholarship.

The school district in which data were collected (purposely left anonymous) must also be acknowledged. Their support and facilitation were crucial to the dissertation. The final acknowledgement is to the families and children who participated in the research. Their wish has been to improve the quality of life for other children who may experience an extended period of speech, language, and literacy development. I hope that I have honored their courage and generosity in the completion of this research.

Dedication

This dissertation is dedicated to my family, and to the memories of my mom and dad.

Chapter 1: Introduction

This thesis concerns the phonological acquisition of multisyllabic words (MSWs) in children of school age. Researchers have suggested that MSW production is an essential component of phonological assessment of school-aged children (Burrus, 2007; Carroll & Snowling, 2004; Holm, Farrier, & Dodd, 2008). There are two aspects to this claim. The first involves the connection of phonological development to literacy. The second concerns phonological development of MSWs, which might not be reflected in acquisition of the phonological system in short words. To address the first aspect, a meta-analysis was initially conducted in order to examine factors related to word level literacy skills of elementary school children with a history of PPD (Chapter 2), including production of MSWs. The reasons for this were two-fold: (a) both phonology and print encode the sounds of language; and (b) literacy acquisition co-occurs with the phonological development of MSWs. The results of the meta-analysis provided motivation for the second aspect of the claim and the remainder of the thesis, i.e. MSW development. A new measure of MSW production was first evaluated (Chapter 3). The research culminated in the main study (Chapter 4), in which the measure was used to examine the MSW phonology of children with and without PPD, who were attending elementary schools in the southern interior of British Columbia, Canada. The introduction expands this initial overview of the thesis, beginning with key definitions concerning PPD and literacy. Theoretical constructs are then reviewed that relate to the overarching research questions of the thesis, the latter of which appear in the final section of the introduction.

1.1 Definition of Protracted Phonological Development (PPD)

The term *protracted phonological development* (*PPD*) used herein, encompasses various other terms in the literature, including: (a) combinations of *Phonological, Speech Sound, (In)Consistent Speech, (Developmental) Speech, or Articulation* with one of *Disorder, Impairment, Delay, or Errors*; and in addition, (b) *Developmental Verbal Apraxia or Childhood Apraxia of Speech*. The view taken of children's phonological development is a continuum of acquisition of different interacting levels of phonological structure (Bernhardt & Stemberger, 1998; Rvachew & Brosseau-Lapré, 2012). Progression along the phonological continuum depends on the degree to which known or unknown biological and environmental factors impact neurodevelopmental processes (Rvachew & Brosseau-Lapré, 2012), and in turn, the pervasiveness and persistence of patterns of constraint on phonological components (Bernhardt & Stemberger, 1998). For the most part, phonological patterns observed in PPD also occur in typical phonological development, and there is the ability to explain atypical patterns in terms of components of the phonological system (Bernhardt & Stemberger, 1998). In that respect, the notion of a developmental continuum is also plausible.

1.2 Definition of Literacy in the Context of the Study

Literacy is the ability to construct and express meaning through reading, writing, and talking about text (British Columbia Ministry of Education, 2006; Murnane, Sawhill, & Snow, 2012). Two key skills of literacy, word recognition (decoding) and spelling, are related particularly to the phonological system because processing of print requires integration with the existing sound components of language (Duncan, 2009; Nathan, Stackhouse, Goulandris & Snowling, 2004; Perfetti, Landi, & Oakhill, 2005; Stackhouse & Wells, 1997; Wood, Wade-

Wooley & Holliman, 2009). Word decoding and spelling necessarily must be fluent (unconscious) and accurate; otherwise, if slow and effortful (conscious), and/or inaccurate, in theory, insufficient cognitive capacities will be available for text comprehension or written expression (Cunningham, 1998; Duncan & Seymour, 2003; USNICHHD, 2000).

The three language-based skills of literacy (reading, writing and speaking) develop in parallel and have reciprocal relationships across the life span (British Columbia Ministry of Education, 2006; Murnane et al., 2012). During the elementary school years, literacy development involves increasing exposure to multisyllabic words (MSWs), orally and in print (Cunningham, 1998). Grade Four students in British Columbia are expected to fluently read and respond to narrative and informational text, including the relevant MSW vocabulary (British Columbia Ministry of Education, 2006). The cognitive organization of spoken MSWs, therefore, must keep pace with vocabulary exposure in print, so that variously sized phonological components can be mapped reliably to relevant orthographic patterns, e.g. for syllables and sounds of base words, and morphological affixes (Apel, Wolter, & Masterson, 2006; Bahr, Silliman, & Berninger, 2009).

1.3 Literacy of Children with PPD

The connection between speech production and literacy provokes questions about the impact of MSW production on the literacy of children who show protracted phonological development. Identification of PPD for preschoolers and children of early school age, is typically determined by evaluating phonological production solely in short words (James, 2006; Skahan, Watson, & Lof, 2007), leading to suggestions that by school age, most children have resolved their PPD (Shriberg & Kwiatkowski, 1982). Nonetheless, for children with a history of PPD, the research is mixed as to success in acquiring word level literacy skills, such as phonological awareness, word reading, or spelling. On the one hand are studies that have reported average attainment of the respective skills (Apel & Lawrence 2011; Bishop & Clarkson, 2003; Puranik, Petscher, Al Otaiba, Catts, & Lonigan, 2008). On the other hand, children with and without PPD have differed significantly with regard to phonological awareness, word reading, or spelling (e.g. Bird, Bishop, & Freeman, 1995; Peterson, Pennington, Shriberg, & Boada, 2009; Nathan, Stackhouse, Goulandris, & Snowling, 2004, respectively). A widening gap in grade level reading achievement has also been evident between Grade 1 and Grade 3 (Puranik et al., 2008), and in addition, children with persistent PPD have been more at risk for reading disabilities (Nathan et al. 2004; Puranik et al., 2008; Raitano, Pennington, Tunick, Boada, & Shriberg, 2004).

1.4 Motivation for the Research

Resolution of PPD implies that foundational speech production skills of MSWs are in place for literacy acquisition. Scant information is available regarding MSW speech development, however; existing reports mainly concern Australian English-speaking children, often of culturally specific words (James, 2006). These circumstances suggest gaps in knowledge about the MSW phonological development of Canadian English-speaking children in general, and moreover, with regard to culturally relevant words (James, 2006). Furthermore, while James (2006) reported prevalence of individual phonological mismatch patterns across samples of children and words, such data do not quantify whole word phonological accuracy in comparison with adult targets. The current program of research, therefore, had a threefold purpose: (a) to determine the convergence of evidence about literacy acquisition for children with a history of PPD, and then, the relationship of MSW production to this achievement; (b) to develop a quantifiable whole word metric of MSW production; and (c) to examine the accuracy of MSW production in children with and without PPD, pre- and post-early literacy instruction. In sequence, the dissertation chapters present the research relative to these questions. The next section provides additional details about chapter content, and key aspects of the background theories relevant to spoken MSW production.

1.5 Overview of the Chapters

The three chapters that follow, i.e. chapters 2, 3, and 4, are presented in manuscript style. A progression of research is presented, but because of diverse purposes and methodologies, the studies in each chapter are introduced by a review of the relevant literature, and therefore, a literature review is not included in this introductory chapter. Chapter 2 addresses whether literacy levels are related to definitions of PPD, and in particular whether methods for defining PPD included the ability to produce MSWs. A meta-analysis is presented concerning the word-level literacy of school age children with PPD, in relation to ways in which PPD was defined: (a) stage of literacy at speech evaluation, i.e. pre- or post-early literacy instruction; (b) whether or not PPD was considered resolved; and (c) whether or not MSWs were assessed. The dissertation will suggest that MSW evaluation was relevant to literacy acquisition because ongoing PPD was identified. This hypothesis motivated the pilot study reported in Chapter 3, within which the presentation of English data is relevant to this thesis. That study, however, also presents data for French, that was part of a published version, because the aim was to describe application of the metric to more than one language spoken in Canada (Mason, Bérubé, Bernhardt, & Stemberger, 2015). For the thesis, the French data were retained because the integrity of the English data interpretation depends on the study as a whole.

The pilot study in Chapter 3 evaluated the reliability and validity of a whole word MSW metric, structured according to nonlinear phonological and parallel interactive language processing theories. The dissertation will argue for the potential of the metric to more finely rank-order variable word productions, and thus, to compare the relative degree of mismatch from adult-like targets.

Chapter 4 presents the main study for the thesis, which furthers the use of the metric to develop a criterion reference for MSW productions of 5- and 8- to 10-year-olds with typical speech (n = 124). The metric is also evaluated for its ability to differentiate MSW productions of 8- to 10-year-olds with PPD (n = 12) from age-matched children with typical speech, and from typically speaking, but developmentally younger 5-year-olds. The theoretical background for the studies is reviewed in the next section, initially, nonlinear phonological linguistic theory, and subsequently, theories of language processing.

1.6 Theory and Multisyllabic Word Production

A suitable theory of MSW production must include explanations that relate to several perspectives: (a) linguistic composition and the underlying phonological representations; (b) the course of childhood phonological development and variable mismatches from adult targets; and (c) language processing and cognitive efficiencies. First, linguistic theory must account for the multiple phonological elements of MSWs, i.e. number of syllables and stress patterns, in addition to consonant and vowel segments with their features. In terms of childhood phonological development, a second requirement is to account for James' (2006) findings about MSW acquisition and patterns of mismatches. Between ages 2;6 and 11 years, James identified three stages of developmental trade-offs between word prosody and phoneme accuracy. Initially, between ages 2;6 and 4 years, word length in terms of syllable number and words with initial stress increased in accuracy at the expense of syllable structure and feature sequences. Second, children between ages 4 and 6 years increased their segmental accuracy but stress patterns were more often sacrificed, especially when stress was not word-initial. Finally, between 7- and 10-years old, both prosodic and segmental components of MSWs were matched.

With respect to language processing, MSWs with their length and complexity, and lower frequency in English (cue validity) (Bernhardt & Stemberger, 1998), potentially tax cognitive resources (cue cost) to a greater extent than shorter words. A theory of MSW processing must explain how multiple linguistic cues promote combination and recombination of phonological components that surface as unique MSWs, and the trade-offs between word structure and segmental features in acquisition.

1.7 Nonlinear Phonological Theory

The theory of nonlinear phonology referenced for the dissertation is the hierarchical structure of the prosodic word suggested by Bernhardt and Stemberger (1998) (see Figure 1.1). In this version, the prosodic word is the highest level of single word representation, with a hierarchical structure of smaller phonological units below it. Represented by nodes in the hierarchy are feet, syllables, timing units, and segmental features. Association lines specify the relationships among these units. Key constructs that describe the relationships among phonological components are autonomy, adjacency, simultaneity, dependency and constraints. In the next sub-sections, these concepts will first be described as they relate to the hierarchical geometry in general before discussing the phonological components more specifically.

1.7.1 Key Constructs of the Phonological Hierarchy

There are several fundamental constructs of the phonological hierarchy: nodes, tiers, links, feature autonomy, adjacency, and simultaneity (Bernhardt & Stemberger, 1998). In the geometry, nodes of two types represent phonological components: content and organizing nodes. In theory, the former contain phonetic information about segmental features that allows realization of surface output, whereas the latter organize features into groups of similar elements (e.g. [Labial] includes [round] and [labiodental]). The phonological components represented by nodes are on independent tiers, with links between them that define the component relationships. Because manner, place and laryngeal feature components are on different tiers, they are autonomous in the sense that they can participate independently in phonological patterns (Harris, 2007).

Elements on the same tier occur in sequence and are adjacent if no other elements are between them (tier adjacency) (Bernhardt & Stemberger, 1998). If, in addition, two elements are in adjacent segments, they have Root node adjacency, and if in adjacent syllables, then syllable adjacency. Notably, consonants and vowels are on different tiers, and therefore are not adjacent at the Root node level, but can be adjacent on the timing unit tier (timing unit adjacency). This notion accounts for compensatory lengthening of a vowel when a consonant apparently deletes; i.e. the features of the consonant delete while its timing unit remains. The features linked to an adjacent vowel may thus link up to the preserved consonant timing unit, and as a result, the vowel is lengthened.

Relationships also exist between elements on different tiers. Direct or indirect dependencies between lower and higher components are a result of associations between organized nodes (Harris, 2007). These dependency relationships can be described as follows: (a) direct linking of two elements, A and B; and (b) direct linking of two elements, A and B, and also of A to C, and therefore, indirect linking of B to C. As a result of either dependency relationship, two elements on different tiers can occur simultaneously.

Phonological patterns describe altered associations in the hierarchy as a result of constraints in the phonological system (Bernhardt & Stemberger, 1998). Constraints are responsible for insertion or deletion of components, entailing deletion or insertion of the relevant association line. Association line deletion (delinking) or insertion (relinking) can also be independent of a relevant phonological component. That is, an association line might delink entirely, or alternatively, delink from one element and relink to another existing element (spreading). Because organizing nodes and association lines group elements, a pattern of deletion or insertion might apply to the group as a whole, and as a result, explanations for patterns are simplified. Next, specific to the prosodic word, phonological components and their relationships will be elaborated. Where relevant, implications for phonological development are included.

1.7.2 The Phonological Hierarchy of the Prosodic Word

The prosodic word dominates the next largest phonological component, the foot. Feet are combinations of syllables, patterned by stress (Harris, 2007; Kager, 2007). One syllable within the foot has greater (primary) stress, or prominence, than the remainder of syllables, which are unstressed. The prosodic word may comprise one or more feet. In the latter case, each foot contains a prominent syllable but the stressed syllable in one foot has greater prominence (primary stress) than the stressed syllable in the other (secondary stress). Primary stress can exist within the first or second foot, classifying the prosodic word as either left- or right-prominent, respectively. Word length in combination with stress often determine left or right prominence, as follows: (a) right-prominence in two-footed words with a single syllable in each foot, e.g. *shampoo* /ˌʃæm.ˈpʰuː/, or with a final foot of two syllables, e.g. *umbrella* /ˌʌ̃m.ˈbɹɛ.lə/; and (b) left-prominent when the first foot has two syllables, and the last foot, a single syllable, e.g. *dinosaur* /ˈdaɪ.nə.ˌsɔɹ/. MSWs of English are usually no longer than two feet, but more commonly are one-footed disyllables. Primary stress of a disyllabic word is more often on the first syllable (*trochaic* stress, e.g. *BAby*) as opposed to the last (*iambic* stress, e.g. *giRAFFE*), i.e. trochaic stress is preferred (Shockey, 2003). For young children or children with PPD, mismatches at the prosodic word level might result from interactions of stress and foot length. For example, a maximal word preference might constrain length to a single foot (Bernhardt & Stemberger, 1998), such that the non-prominent foot is deleted. Iambic stress of a two-syllable foot might be avoided, and surface variously, e.g. syllables of equal stress, primary stress shifted to the unstressed syllable (trochaic stress), or deletion of the weakly stressed syllable (Kehoe, 2001).

At the next level are syllables organized in terms of the onset and rime (Zec, 2007). The syllable rime links to the syllable nucleus (peak) and to the coda. The nucleus includes one or more vowel or syllabic consonant slots, and the onset and coda, consonant slots. Slots in the onset and coda refer to consonants that precede or are subsequent to the vowel, respectively. Because a syllabic consonant also determines the syllable nucleus, it is positioned in a vowel slot. A vowel or consonant slot indicates the presence of a segment in terms of the amount of time allotted for its realization in the output. Slots do not contain the phonetic information of segmental features. The independence of segmental slots from features allows the segmental length to be realized in spite of features failing to surface, e.g. compensatory vowel lengthening when a following consonant deletes. Segmental lengthening might also result from syllable components linking to two slots, e.g. customary phonetic lengthening of high tense and low vowels of English, (i, u) or (æ, ɑ); and non-customary vowel lengthening, as a result of a constraint on an element of the rime, the mora, discussed in more detail below.

Moras are timing units that link to components in the rime, and are related to stress assignment. In English, lax or short vowels and coda consonants are monomoraic and long vowels and diphthongs are bimoraic. Moras encode syllable weight, such that a heavy syllable has two or more moras, and is therefore more likely to attract prominent stress. A light syllable is unstressed because it links to a single mora. If heavy syllables are preferred in the phonological system, the weight of a light syllable could be increased by segmental insertion in the rime, thereby adding a mora, e.g. consonant insertion that adds a coda to an open (CV) syllable; inherent length added by a high tense or a low vowel insertion. In the example below for *hippopotamus*, because of timing unit insertion, /ɪ/ is lengthened, and as a result, the initial syllables of the non-prominent foot /ˌhɪːpə/, and the prominent foot /ˈpʰɑːɾəməs/, have equal weight.

hippopotamus ˌhɪpəˈpʰɑːɾəməs ˌhɪːpəˈpʰɑːɾəməs

Finally, at the bottom of the hierarchy are features, the smallest units. Feature nodes are the only content nodes in the hierarchy, because they contain the phonetic information for segmental realization. Features are also autonomous elements, and therefore link together in various combinations to represent segments. Patterns are thus explicable in terms of features that segments have in common. Feature combinations are organized at the Root node, which links directly to the various Manner features, and in addition, to organizing nodes for Place and Laryngeal (Figure 1.1). Manner features are therefore terminal because they have no other dependent elements. The Place node links to dependent content nodes representing the major places of articulation, Labial (lips), Coronal (tongue tip or blade) and Dorsal (tongue body). These major place nodes link in turn to secondary, or terminal features that further specify place of articulation. Finally, Laryngeal organizes terminal features for the larynx. The content nodes of the phonological hierarchy therefore encompass the manner nodes, major and secondary place nodes, and the laryngeal feature nodes (Table 1.1).

Consonants features also interact with syllable structure to determine allowable consonant sequences in the onset or coda. The sonority hierarchy of consonant classes and vowels generally govern such sequences. That is, in the onset, consonants increase in sonority between the syllable edge and the vowel; conversely, in the coda, consonant sonority decreases between the vowel and the syllable edge. For example, the general sonority sequence from least to most sonorous is stops, fricatives, nasals, liquids, glides and vowels. A particular exception to the sonority sequence is with respect to /s/+stop clusters, e.g. /st/, because the fricative /s/ has greater sonorance than a stop. To resolve this violation, /s/ might be *extrametrical*, outside the syllable that dominates the stop. In the situation of an intervocalic consonant, there is possibly joint linking to the adjacent coda and onset, with the conditions of the sonority hierarchy then met in both prosodic positions. When the second syllable is stressed, however, intervocalic consonants can be assigned to the onset. At the boundary of two syllables, the maximum number of consonants that form an allowable sequence are generally positioned in the onset; the remainder are situated in the preceding coda.

Concerning vowels, certain traditional features incorporate acoustic with articulatory properties (Ladefoged, 2006), e.g. tenseness, height and backness (Table 1.2). Nevertheless, in the phonological hierarchy, vowel features follow consonant feature organization in general. The dependency relationships for the manner, place and laryngeal nodes of vowels are as follows: (a) tenseness with the Root node; (b) height, backness and rounding with Place; and (c) [+voice] with Laryngeal. In the nonlinear hierarchy, vowel and consonant features can be viewed as positioned on separate planes, i.e. are not adjacent, and therefore may show independent patterns. For example, phonological patterns across consonant sequences can exclude intervening vowels. Vowel and consonant interactions can occur, however, as a result of links to higher adjacent nodes, i.e. vowel adjacency because of syllable node adjacency; consonant adjacency because of timing unit adjacency. Constraints on the higher tiers will interact with production at the lower tiers.

Further to feature relationships, as mentioned previously, nodes group features together and association lines determine their simultaneous occurrence. Features can therefore *spread* as a group over more than one segment. *Spreading* re-aligns the underlying timing such that elements begin earlier (leftward spreading) and/or end later (rightward spreading). *Spreading* is one of the principal arguments in support of feature grouping (Bernhardt and Stemberger, 1998) because the process of spreading explains assimilatory patterns. For example, a nasal consonant assimilates the place of various successive consonants. The most parsimonious explanation for this pattern is spreading of the Place node, as opposed to spreading of separate dependent place features.

Feature assimilation occurs most often between adjacent segments, because an intervening segment can effectively block feature spreading. Relinking an association line is blocked because of cross over with the association line of the intervening segment. For example, Place may be blocked from spreading between syllables when there are several intervening segments.

Features may be affected by the relative strength of higher prosodic elements. Certain prosodic positions are *weak* with respect to stress and syllable structure. These include: non-prominent feet; unstressed syllables, and in particular, word initial and within word weak syllables (James 2006; Kehoe, 2001); codas; and the second consonant of clusters in syllable onsets. Weak prosodic positions are more subject to feature constraints. First, certain features might be prevented in weak prosodic positions, usually those that are less frequent/more marked/nondefaults. Second, features might be restricted to solely the onset, coda or intervocalic position. Third, to approximate optimal syllable structure, only preferred features might occur in the onset or coda. For instance, in English, codas may appear that have features related to vowels ([+sonorant], [Dorsal]). Finally, constraints might apply to sequences, of the same or of different features, or of the same or different stress levels.

With respect to the representation of the multiple components of MSWs, nonlinear, as opposed to linear, phonological theory appears to be particularly useful. For example, from a linear standpoint, similarities between prosodic words such as *cash register* /ˈkʰæːʃˌɹɛdʒɪstɚ/, and *hippopotamus* /ˌhɪpəˈpʰɑːɾəmɪs/, seem few because segments differ as to number (ten and eleven, respectively), and type (e.g. affricate or fricative; rhotic or nasal), and in addition, in terms of developmental acquisition, i.e. are primarily late or early, respectively (McLeod, 2009). Solely the vowel, /ɪ/, and the consonant /s/, are common to both words. The words, *cash register* /ˈkʰæːʃˌɹɛdʒɪstɚ/, and *hippopotamus* /ˌhɪpəˈpʰɑːɾəmɪs/, however, are apparently more similar from a nonlinear perspective. For instance, the words are two-footed. The final feet, /ˌɹɛdʒɪstɚ/ and /ˈpʰɑːɾəmɪs/, each contain three syllables, the first of which has prominent stress, while the final two are unstressed. The first two syllables, /ˌɹɛ/, and /ˈpʰɑː/, also share open (CV) structure, i.e. onset-nucleus without a coda. An additional possibility, in /ˌɹɛdʒɪstɚ/, is that /s/ is ambisyllabic, such that a foot in each word contains a closed syllable, CVC, i.e. /dʒɪs/ and /mɪs/, and therefore, the rime, /ɪs/, is common to both. The advantage of construing word components as nonlinear is that phonological components are organized in linked *chunks* that might facilitate mapping onto the timing of surface output, possibly bootstrapping MSW phonological learning. Conversely, a phonological system with insufficient hierarchical organization, might contribute to protracted realization of adult-like output, even if a word such as *hippopotamus,* is composed of earlier acquired segments (James, 2006).

In order to further explain the mechanisms involved in organizing phonological components and their timing during word production, various language processing theories are available. A suitable theory, however, must also account for aspects of phonological development, which are summarized in the next section, before focusing on two general categories of theories: serial modular and parallel interactive.

1.8 Child Phonology and Language Processing Theories

In relation to phonology, a theory of language processing explains how word meaning is mapped to output in the form of a series of speech sounds. For an adequate explanation, there are considerations of relevance to child phonology in general (Bernhardt & Stemberger, 1998;

Presson & MacWhinney, 2010), and more specifically to the phonology of MSWs (Bernhardt & Stemberger, 1998; James, 2006; Kehoe, 2001). Beginning with general considerations, processing models should provide a single cohesive explanation for adult and child phonological output, whether typical or atypical. Second, theories must account for acquisition, and the variability of output across and within children.

Considerations specific to MSWs relate to linguistic structure and available developmental information. First, a model of language processing must account for the nonlinear structural complexity of MSWs, and therefore, the output of multiple phonological elements, i.e. feet, number of syllables and stress patterns, syllable structure, and in addition, consonant and vowel segments with their features. Second, the stages of acquisition and patterns of mismatches reported by James (2006) and Kehoe (2001) should be explicable. These stages have concerned developmental trade-offs between matching word prosody, and phonemes (James, 2006). That is, between ages 2;6 and 4 years, children increasingly matched consonant and vowel structure, and the segmental features of monosyllables. In contrast, for MSWs, syllable number and stress patterns (with greater prominence on the first syllable) increased in accuracy at the expense of syllable structure and feature sequences. More syllable and segmental deletions were related to word-initial and within-word unstressed syllables, phenomena that Kehoe (2001) also reported. Next, children between ages 4 and 6 years increased their production of weakly stressed syllables. Weak syllables within words, however, remained most vulnerable to mismatches, and in particular, those with sonorant onsets or posterior to anterior place sequences. Because segmental deletions were fewer, a more frequent pattern was assimilation. When segments were accurate, however, stress (with greater prominence on a syllable other than the first) was more often sacrificed, as a result of either schwa insertion or tensing. Finally, between 7- and 10-years old, both prosodic and segmental components of MSWs were matched.

The next sections focus on two general categories of language processing theories: serial modular and parallel interactive, each of which is discussed in turn, and then implications for MSWs are suggested.

1.9 Overview of Serial Modular and Parallel Interactive Theories

Several models of language processing fall into two large categories: serial modular (e.g. Levelt, Roelofs, & Meyer, 1999; Shattuck-Hufnagel, 1992), and parallel interactive (e.g. Dell & Kittredge, 2010; Stemberger, 1985; Wheeler & Touretzky, 1997). To explain language output, these views have differed with regard to units of processing and their relationships, and in addition, as to the cognitive representation of words. Theory verification has also related to diverse age groups and integrity of the phonological system, i.e. childhood typical or protracted development (Bernhardt & Stemberger, 1998; Menn & Matthei, 1992), adult phonology in experimental paradigms (Levelt, et al., 1999; Shattuck-Hufnagel, 1992; Wheeler & Touretzky, 1997), and adult acquired phonological disruptions (Dell & Kittredge, 2010; Presson & MacWhinney, 2010). Within these theories, some have taken into account continuity across the lifespan (Bernhardt & Stemberger, 1998; Dell & Kittredge, 2010; Presson & MacWhinney, 2010; Stemberger, 1992), and certain others have implemented computational modeling (Levelt, et al., 1999; Shattuck-Hufnagel, 1992; Wheeler & Touretzky, 1997).

Further to units and processes in the two categories of models, those that are serial modular have identified discrete specific language domains, e.g. phonological rules as opposed to motor programs (Menn & Matthei, 1992); and lexical features as opposed to lexical form

(Levelt, et al., 1997). Parallel interactive models have focused on the coordination of cognitive processes in terms of patterns of unit activation (Presson & MacWhinney, 2010; Stemberger, 1992). Direction of processing has also been suggested. The specialized cognitive modules of traditional theories have operated serially, combining information incrementally to assemble a form of word output. In contrast, parallel interactive models have attempted to integrate theories of linguistic structure with cognitive and neurophysiological knowledge. Such models include cross-level interactions and multiple connectivities, and thus align with other explanations of human learning (Bernhardt, Stemberger & Charest, 2010; Elman, 1999; Stemberger, 1992). Connectionist models also characterize language processing within more generalized cognitive systems of hierarchical networks with multidirectional interactions. Specialization has related to patterns of activation across multiple nodes in the hierarchy, rather than to modules.

Another contrast between serial modular and parallel interactive theories concerns whether cognitive word representations are stored and how they are accessed. In serial modular theories, long-term storage of representations has been presumed to speed access and selection, because an output form is not reconstructed each time it is produced (Menn & Matthei, 1992; Stemberger, 1992). Parallel interactive theories, however, have proposed that the strength of connections across a network facilitates access and selection, with distributed effects (Stemberger, 1992). With this general background to the two overarching models, additional details about serial modular and parallel interactive theories will be presented below, and then the relevance of their similarities and differences to MSW production will be considered.

1.9.1 Serial Modular Language Processing Models

In earlier theories of childhood phonology, language processing occurred between input and output components, and was strictly unidirectional (Menn & Matthei, 1992). Knowledge of word forms was stored in modular components such that application of the child's phonological rule system took place between them. Serial processing was posited as either on-line or off-line in different approaches, i.e. in relation to the point in time at which an underlying word representation was selected for output. Processing of output in real time was on-line. In the two-lexicon approach, creation of the child's pronunciation from the adult form took place off-line, at some other point in time, e.g. once a word had been processed on-line for the first time (Menn & Matthei, 1992). The notion of off-line processing accounted for more phenomena in the phonological output of young children, and addressed the issue of efficiency of processing. In turn, however, reduction in the amount of on-line processing increased the number of storage components in the system, to which the discussion now turns.

The construct for storage of a child's underlying representations, the lexicon, was a principal component of early serial models. Initial single lexicon models proposed storage of underlying representations that matched the adult target, presupposing accurate word perception. On-line application of phonological rules reduced the underlying representation to the child's output form (Bernhardt & Stemberger, 1998; Menn & Matthei, 1992; Stemberger, 1992). The single lexicon approach received two main criticisms. The first concerned the assumption of adult-like underlying representations. The second concern was the lack of an account for variable segmental production within children, i.e. exceptions in the output with components that were less accurate, or alternatively, more accurate in comparison with phonological patterns in general.

To address these criticisms of single lexicon models, two-lexicon models proposed an additional storage module, the output lexicon (Menn & Matthei, 1992). In comparison with single lexicon models, the input lexicon stored phonological representations as perceived by the child, which may be adult-like in some ways but non-adult-like in others. Representations in the input lexicon were used in word recognition, but also in on-line speech production, the first time a word was produced. The output lexicon stored representations that resulted from a one-time, on-line application of the child's phonological rules to the input representation. Once stored in the output lexicon, representations could be directly selected for production without further on-line processing of the input-lexicon representation. Surface variability could occur because of access to various permanently stored output representations. Cognitive processing resources were also conserved because representations were not repeatedly constructed on-line. Nevertheless, the selection of one representation over another, and the eventual stabilization of an adult-like representation, was not explained by this version of the two-lexicon model (Bernhardt & Stemberger, 1998; Stemberger, 1992).

Two-lexicon models were later embedded into sequences of numerous modules of psycholinguistic processing, from speech recognition to motor execution (Vance, Stackhouse, Pascoe, & Wells, 2007). The purpose was to identify breakdowns in off-line processes involved in phonological output. In particular, processes for encoding underlying phonological representations and output motor programs were considered in relation to semantic representations as verified on *word* elicitation tasks, i.e. picture naming, and word or non-word repetition. For example, for picture naming, underlying phonological representations fed forward by two routes, directly to output motor programs, and also indirectly via the semantic representation. In contrast, word repetition and nonword repetition by-passed the semantic system, and for nonword repetition, the phonological representation was by-passed in addition. As a result, processes in the semantic and phonological components could be sharply disengaged depending on the task.

Serial modular theories also developed from analyses of errors in adult speech, primarily in experimental tasks. Computational modeling was used to replicate the construction of phonological output for theory verification. In these theories, a unit of the morpho-syntactic organization for a lexical concept, the *lemma*, was retrieved before the phonological representation, or *lexeme* (Levelt et al., 1999). The phonological representation was stored as an entire unit, but was broken down into segments for processing word output.

A key aspect of computational serial models was that processing of segmental sequences was in relation to larger linguistic word constituents, i.e. encoding the prosodic word as opposed to the lexical word (Levelt et al., 1999; Shattuck-Hufnagel, 1992). The phonological form of the base word included separate components for metrical structure, e.g. main stress or number of syllables, and segmental composition. In the model of Levelt et al., the metrical frame was retrieved before the segmental units, and then segments filled the frame in successive order. Levelt et al. argued for strictly serial processing but acknowledged that the feed forward of information from metrical through segmental levels could overlap. In that respect, information spread to some extent, but without cascading freely.

Shattuck-Hufnagel (1992) expanded on the use of elements in the metrical structure of a word. Segments were assigned to place holders, or slots, for syllable structure components, i.e. the nucleus, onset and coda. Selection processes were anticipatory in terms of scanning ahead in the buffer of activated candidate segments. The first segments identified were those needed for stress assignment, i.e. vowels for assignment to the syllable nucleus. In the next step, consonants were serially assigned to the onset. In the last step, the scan ahead process identified a vowel available for assignment to a subsequent syllable. If none was found, consonants from among those still held in the buffer were sequentially assigned to the coda. Because of the potential for incorrect segmental selection, Shattuck-Hufnagel (1992) proposed two additional components: a *check off* monitor that *kept track* of segments that had been selected and prevented re-selection; and an *error checking* module that detected and deleted input-output mismatches. The addition of modules external to the selection process likely took up cognitive resources, however, reducing as opposed to improving system capacity. In the next section, parallel interactive models are described that propose simultaneous as opposed to serial processing.

1.9.2 Parallel/Interactive Language Processing Models

The theories included in this section also have a basis in computational modeling. First, Parallel Licensing Theory (Wheeler & Touretzky, 1997) is described, and then connectionist models of language processing (Bernhardt & Stemberger, 1998; Dell & Kittredge, 2010; Stemberger, 1992). Parallel Licensing theorists (Wheeler & Touretzky, 1997) expanded on the components of metrical structure included in serial computational theories (Levelt, et al., 1990; Shattuck-Hufnagel, 1992). All phonological units were considered to be licensed, i.e. to belong to higher prosodic categories; without licensing, an element could not be produced. Processing was parallel in the sense that each licensing node, e.g. syllable, rime, and nucleus, simultaneously attempted to license any and all segments within its domain. A licensing node could therefore license several segments and segments could have multiple licensers. One or more activated candidates in the buffer could be active at a given point in processing. Linear Output Constraints (LOCs) prevented unallowable segmental assignments on the basis of the amount of activation in the lexical string when licensing began. A segment could therefore be assigned without activation of all of its features, provided the information available to the LOCs met sonority principles and other phonotactic constraints. This was key because underspecification could lead to competition among units in the buffer. Errors in selection could occur, but segments that later violated the LOCs could also be dropped from the representation. Similar to Shattuck-Hufnagel's (1992) model, vowels were assigned to the nucleus first; however, processes for segmental assignment to the coda and onset were not serial, but were simultaneous.

The next parallel/interactive theories considered in the thesis, connectionist models of language processing, have also been applied to phonological processing of adults (Dell & Kittredge, 2010), and in addition, to children (Bernhardt & Stemberger, 1998; Stemberger, 1992). In these models, output processing always occurred on-line, in some versions of the theory with two steps (Dell & Kittredge, 2010) and in other versions, with one (Stemberger, 1992). In two-step models of single word production, processing began with serial activation of the lexical and phonological networks; the latter in turn mapped to the motor representation (Dell and Kittredge, 2010). In a one-step model, a phonological form was created on the basis of the word's meaning, such that meaning was mapped to the motor representation (Bernhardt & Stemberger, 1998; Bernhardt, Stemberger & Charest, 2010). The phonological system was considered to have constraints related to the difficulty of producing elements in the output. Mapping from meaning to the sound based representation provided just enough information to adjust the constraints, in order to match the output to the detailed perceptual representation. The detail of the perceptual representation included predictable (unmarked) and unpredictable (marked) phonological-phonetic information. In the meaning to sound mapping, predictable phonological information was not specified. The system was assumed to be biased to output high-frequency information that did not have to be learned lexically; however, low-frequency unpredictable information was assumed to be learned lexically, in order to overcome system processing biases. Until learned, the latter could surface as elements that were less marked, i.e. mismatched in comparison with the perceptual representation.

The structure of a connectionist model was represented in terms of a network of hierarchically organized units connected by links that could vary in strength (Dell & Kittredge, 2010). A unit, or node, had an activation value that changed over time as it is passed through the connections to other units. The processes that mapped input to output occurred by spreading activation throughout the network of units. Activation of one word in turn activated a *gang* of words and syllables that were similar to the intended word to greater or lesser extents, both semantically and phonologically. Within the gang, bi-directional interacting activations of components could influence the selection of components for the target word. Correct selection required appropriately timed patterns of competing levels of activation and inhibition among units (Bernhardt et al., 2010; Presson & MacWhinney, 2010). Repeated selection of units created strong connections that networked larger combinations of units, and reduced the load on cognitive capacity. Learning was explained by the strengthening of connections as a result of selection. Units that were repeatedly selected remained at a higher resting level of activation, and were more available for re-selection.

1.9.3 Language Processing Theories in Relation to MSWs

In the previous two sections, the discussion has focused on two categories of language processing theories: serial modular (1.9.1) and parallel interactive (1.9.2). The models differ as to their bases in observations of childhood phonological development (Bernhardt & Stemberger, 1998; Menn & Matthei, 1992; Stemberger, 1992), or adult phonological errors and reaction times in experimental tasks (Levelt et al., 1999; Shattuck-Hufnagel, 1992; Wheeler & Touretzky, 1997). In this section, the suitability of the models for MSW production in school-age children will be deliberated, in terms of certain parameters. The first of these relate to continuity between childhood and adult phonological output, both typical and atypical; the relationship of phonology to meaning; the nature of underlying representations and cognitive efficiency; and also, variability in early childhood phonological development. Other considerations relate specifically to MSWs, i.e. the linguistic structure of multiple nonlinear phonological components, and in addition, stages of childhood acquisition and phonological patterns (James, 2006; Kehoe, 2001). Serial, and then parallel interactive models are discussed in turn below, with respect to their limitations.

With the exception of the considerations specific to MSWs, the limitations in the preceding paragraph arose from criticisms of single and two lexicon models (Bernhardt & Stemberger, 1998; Menn & Matthei, 1992; Stemberger, 1992), as reviewed in section 1.9.1. Later serial models (Levelt et al., 1999; Shattuck-Hufnagel, 1992) with their focus on adults, also had limited ability to account for change and variability in childhood phonological acquisition (Bernhardt et al., 2010). A second weakness was the serial relationship between stored meaning and phonological components that precluded the interaction of one word's form with that of other lexically and phonologically related words. As a result of these drawbacks, this group of serial models will not be referred to further.

The remainder of the discussion will concern parallel interactive theories, including parallel licensing (Wheeler & Touretzky, 1997) and connectionist theory (Bernhardt &

Stemberger, 1998; Stemberger, 1992). These theories, adapted from cognitive information processing, have appeal because language is a complex and distributed cognitive process. Across and within large- and sub-scale neural networks, language integrates domain-general multi-dimensional capacities, such as neural connectivity, working memory and attention (Presson & MacWhinney, 2010). The dynamic processes proposed in parallel/interactive models allow generalization beyond childhood, and to atypical phonology. Change and variability in phonological acquisition are also accommodated by parallel interactive processing among hierarchically organized components with varying amounts of activation. Connectionist models suggest graded interaction from larger to smaller linguistic components. Parallel licensing also occurs within hierarchical levels, in which the relationships between segments and higher levels can be one-to-many or many-to-one. Third, cognitive efficiency is also addressed because phonological representations are not stored as static units but constructed each time a word is spoken. Direct mapping of meaning to output is possible because learning is in terms of recurring patterns of activation and strong dynamic connections between units. Strong connections, in turn, create dynamic localized networks of multiple levels of word structure, e.g. syllables with onset, rime and coda units; such networks are accessed quickly and facilitate timing of selection and production processes overall. Cognitive capacities are not taken up by storage of static representations in long-term memory (Bernhardt & Stemberger, 1998; Bernhardt, Stemberger, & Charest, 2010; Presson & MacWhinney, 2010). Finally, error is accounted for by variability between input and output of dynamic networks, i.e. processes of pruning network components (Doidge, 2007; Elman, 1999). There is therefore no need for external mechanisms with unique *checking-off* operations (Shattuck-Hufnagel, 1992) that in all likelihood demand more cognitive capacities. In the next section, application of parallel processing theories is discussed in relation to developmental information about MSWs, first concerning stages of acquisition and then, patterns of acquisition (James, 2006; Kehoe, 2001).

Relating these theories to development, some data are available in James (2006) that suggests stages in MSW development with initial trade-offs between prosodic structure and phonological features, as follows: (a) stage one, between ages 2;6 and 4 years, children matched syllables and stress patterns (with greater prominence on the first syllable), but phonemes were inaccurate; and (b) stage two, at the expense of prosodic components, children between ages 4 and 6 years matched phonemes. These stages suggest apparent regression in stress and syllable matching between stage one and stage two. By stage two, the regression appears to be with respect to phonemes. Regressions where children appear to progress along one phonological dimension at the expense of another, have also been reported in phonological development of short words (Bernhardt & Stemberger, 1998). It is possible to accommodate these trade-offs within parallel processing models in terms of fluctuating levels of activation. Fluctuations can occur in order to learn new information, i.e. a cue (Presson & MacWhinney, 2010), in terms of the competition to select it. Selection will depend on the accessibility of the cue, in terms of the importance of the information (validity), and in addition, the reliability with which the cue will lead to the correct output form. An additional dimension is cue cost, the amount of processing capacity needed to make use of the cue. For cues to compete, they must be maintained together on the short-term, in some type of working memory, i.e. the buffer (Presson & MacWhinney, 2010; Wheeler & Touretzky, 1997). The strength of a given competitor is determined by its resting activation and additional activation received from other components in the network.

In the phonological system, segmental features provide more information for distinguishing among words than the frames of metrical units, and therefore, feature cues have importance for successful communicative attempts. During James' first stage, the perceptual representations of stressed syllables are presumably more detailed as a result of the salience provided by their greater acoustic energy (Ladefoged, 2006). The predictability of stressed syllables in English is a result of the combination of salience with high frequency in the language (Bernhardt & Stemberger, 1998; Ladefoged, 2006). Being less marked, their cue cost is less, i.e. resting activation levels are lower in comparison with syllables of weak stress. The metrical nodes of stressed syllables therefore reach sufficient activation for licensing of segmental features to begin. As the connections between licensing nodes within stressed syllables strengthen, there is more drive in the system toward faithfulness to weak syllables. Interactive processes will therefore send more activation to weak syllables nodes, word-final before within-word syllables, as a result of the greater perceptual salience of the former. Segmental features active in the buffer(s) will receive inhibitory activations in order that weak syllable nodes can be licensed. Because of the rapidity of speech productions and the consequent competing activations, segmental licensing is more likely to begin before all features have been activated, such that mismatched segments will surface. When features are not licensed, and therefore not produced, segmental *substitutions* will be realized. Assimilation will surface because features that are insufficiently inhibited will be incorrectly licensed by more than one segmental root node. During processing, there is a trade-off, where activations are adjusted to allow faithfulness to more segmental features at the expense of the recently learned (and presumably still vulnerable) weak syllables. Finally, during stage three, the connections among stress, syllables and segmental features reach the appropriate weights for all components of the words to be produced.

In summary, a model of parallel interactive language processing in combination with a linguistic model of nonlinear phonology seem best suited for explaining the complex dimensions of MSW phonological output, and the changes and variability across the course of phonological development. For the purpose of the current research, the relationships of MSW production to literacy, also a language-based system, can be expected to be equally complex. Assuming that MSW production is important to examining the phonological system of children with PPD, the models suggest the difficulty of deconstructing production patterns into many smaller parts. All of the components of a word are relevant in its output, and in MSWs, there are many aspects to each word. As a result, methods that examine the system as a whole would appear to have greater utility than those that might look at individual aspects of word production. These issues are considered in the research questions posed in the next section.

1.9.4 Questions for the Research

The issues discussed in the previous section with respect to frameworks of nonlinear phonology and parallel interactive language processing, in addition to issues related to MSW acquisition, led to several considerations for the current research. The first concerned the risk of children with PPD for literacy, and the connection to spoken MSW phonology. The second was with regard to the structuring of a metric that could quantify whole word interactions of the multiple components representing MSWs. These considerations led to the following research questions.

1. Are children with PPD more at risk for literacy than children without, and if so, can this be predicted in the post-early literacy school years, at least in part by the ability to produce long words (MSW)?

MSWs were predicted to be a moderator for such risk in a meta-analysis (Chapter 2) because such words have high frequency in text to be decoded post-early literacy instruction (Cunningham, 1998), and because studies of children with PPD have suggested that children with PPD are at risk (Bird, Bishop, & Freebairn, 1995; Nathan, Stackhouse, Goulandris, & Snowling; 2004; Peterson, et al., 2009). The next problem was how to characterize the complex productions of these long words in a way that captured all word components. This led to the next question.

2. Does a whole word metric that integrates nonlinear phonological theory with parallel interactive models of language processing, more finely differentiate various productions of children across MSW words, than other existing measures?

   A whole word metric was predicted to be a reliable and valid measure of MSWs compared with other current methods, on the basis of its theoretical grounding in nonlinear phonology (Bernhardt & Stemberger, 1998) and parallel interactive language processing (Bernhardt & Stemberger, 1998; Stemberger, 1992; Wheeler & Touretzky, 1997). The major objectives for the thesis then were to establish criterion reference data for a typically developing cohort, pre- and post-early literacy instruction, and to compare their MSW productions with children already identified with a history of PPD. The following questions pertain to Canadian English-speaking children.

3. Is there a difference between the MSW accuracy of children with typically developing literacy, pre- and post-early literacy instruction?

   A developmental progression in MSW accuracy was predicted, given James (2006) findings of such a progression for Australian English-speaking children.

4. Is there a difference between the MSW accuracy of children with a history of PPD, post-early literacy instruction, and children with typically developing speech, either pre- or post-early literacy instruction?

   Significant between-group differences were predicted between MSW accuracy of children with and without PPD. In comparison with typically developing children pre-early literacy instruction, the prediction was that MSW accuracy of children with PPD would be higher, with a moderate difference. In comparison with typically developing children post-early literacy instruction, the prediction was that MSW accuracy of children with PPD would be lower, with a small difference. Children with PPD were expected to demonstrate a developmental progression in MSW accuracy, in accordance with James' (2006) findings, but at a slower rate, such that gaps would still exist in comparison with typically developing children.

In order to establish the importance of answering the research questions posed above, Chapter 2 presents the results of a meta-analysis that examined the relationship between MSW production and word level literacy of school age children. MSW production was tested as a moderator of literacy, and suggested a meaningful relationship. This led to the pilot study of a MSW metric, reported in Chapter 3, because no theoretically based, quantitative whole word MSW measure was available. Having established initial reliability and validity, the metric was used in the main study of the research questions, reported in Chapter 4. The Conclusion of the thesis then considers the results of the research overall.

Figure 1.1 Prosodic word hierarchy

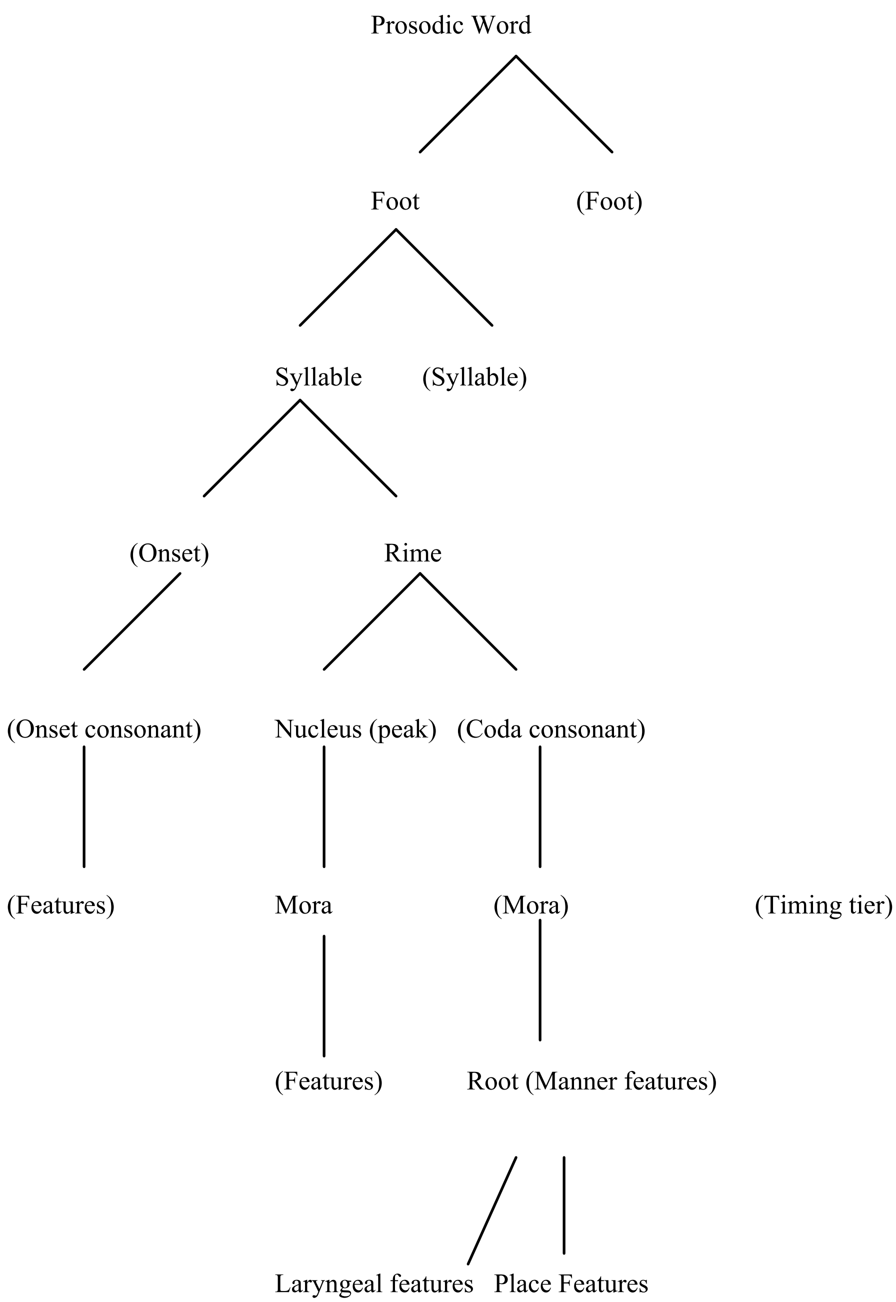


Table 1.1 Nonlinear consonant features

	p	t	k	b	d	g	f	θ	s	ʃ	ʧ	v	ð	z	ʒ	ʤ	w	l	j	ɹ	m	n	ŋ	h	ʔ
Manner																									
consonan-tal	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	-	+	+	+	+	+	+
continuant	-	-	-	-	-	-	+	+	+	+	-/+	+	+	+	+	-/+	+	+	+	+	-	-	-	+	-
sonorant	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+
nasal	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	-	-
lateral	-	-	-	-		-	-	-	-		-	-	-	-	-	-	-	+	-	-	-	-	-	-	-
Laryn-geal																									
voiced	-	-	-	+	+	+	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	-	-
spread glottis	+	+	+				+	+	+	+	+													+	
Place																									
Labial	✓			✓			✓					✓					✓			✓[a]	✓				
Coronal		✓			✓			✓	✓	✓	✓		✓	✓	✓	✓		✓	✓	✓		✓			
anterior		+			+			+	+	-	-		+	+	-	-		+	-	-		+			
grooved		-			-			-	+	+	+		-	+	+	+									
Dorsal			✓			✓											✓	✓	✓				✓		

[a] Onset only.
Note. Adapted from Bernhardt and Stemberger (1998, 2000), where the rhotic is referred to as [-consonantal]. [+spread glottis] is only at the beginning of a stressed syllable.

Table 1.2 Nonlinear vowel features

	i	ɪ	eɪ*	ɛ	æ	u	ʊ	oʊ*	ɔ	ʉ	ə	ʌ	ɑ
+tense	+		+/		+	+		+/				+	+
-tense		+	/-	+			+	/-	+		+		
Labial						+	+	+	+	+			
Coronal	+	+	+	+	+								
Dorsal	+	+	+	+	+	+	+	+	+	+	+	+	+
+high	+	+	/+			+	+	/+		+			
-high			+/	+				+/	+		+	+	
-low			+	+				+	+		+	+	
+low					+								+

Note. Adapted from Bernhardt and Stemberger (1998, 2000).
*The (+) to the left of the slash refers to the first vowel of the diphthong and vice versa; in two instances (-) was used in place of the (+) for the feature [tense].

Chapter 2: A Meta-analysis of Protracted Phonological Disorders and Word Level Literacy in School-Aged Children

2.1 Synopsis of Study

A meta-analysis evaluated word level literacy in children with typical and protracted phonological development (PPD). Fifty-two eligible studies had measured at least one word level literacy construct in school-aged children with and without PPD, generating 64 independent samples: *Phonological awareness* (PA: n = 20), *Decoding* (DEC: n = 17), *Fluency* (FL: n = 4), *Nonword decoding* (NWD: n = 9) and *Spelling* (SP: n = 14). The *Effect Size (ES), the mean mixed weighted d,* was compared on the *Q*-statistic using a random effects model (the small n excluded *FL*). Age at literacy evaluation, severity/sub-type and resolution of PPD significantly moderated *NWD* (*Q* = 14.21) and *SP* (*Q* = 8.00). Contradictory outcomes existed about literacy outcomes for children with PPD, possibly because definitions of PPD used for participant selection differed. One relevant definition concerned phonological context assessed (i.e. word complexity; single words versus connected speech). Speech context, particularly multisyllabic word evaluation, significantly moderated *PA* (*Q* = 4.85), *DEC* (*Q* = 4.14) and *SP* (*Q* = 5.66).

2.2 Introduction

Over the past two decades a number of studies have evaluated the relationship between phonological development and literacy outcomes (Apel & Lawrence, 2011; Bird, Bishop & Freeman, 1995; Foy & Mann, 2012; Holm, Farrier, & Dodd, 2008; Kirk & Gillon, 2007; Meredith, 2002; Overby, Trainen, Smit, Bernthal, & Nelson, 2012; Peterson, Pennington, Shriberg, & Boada, 2009; Preston & Edwards, 2007; Puranik, Petscher, Al Otaiba, Catts, & Lonigan, 2008; Raitano, Pennington, Tunick, Boada, & Shriberg, 2004; Rvachew, 2007; Webster, Plante, & Couvillion, 1997; Wellman, Lewis, Freebairn, Avrich, Hansen, & Stein, 2011). Research evidence concerning the relationship between phonological and literacy development provides a foundation for clinicians providing therapy to children with difficulties in either or both areas. A meta-analysis was conducted for the current chapter on this topic, examining studies that included children with typical versus protracted phonological development. For the various studies, researchers typically sampled SLP caseloads of children with clinical diagnoses of protracted phonological development (PPD[1]), the majority having received phonological intervention. These methodological differences include broader developmental considerations such as age at literacy evaluation and co-occurrence of other conditions, such as Protracted Language Development (PLD). In particular, children designated with severe or persistent PPD appeared to have lower literacy outcomes than those with less severe PPD (Leitão & Fletcher, 2004; Lewis & Freebairn, 2000). Thus, differences in participant selection, possibly reflecting researchers' definitions of PPD, might subsequently have contributed to the apparent contradictions.

Thus, for the meta-analysis discussed in the current chapter, a variety of methodological parameters were examined relative to age of literacy acquisition and characterization of PPD through traditional severity rankings, sub-types, resolution, and the complexity of speech context assessed. While accounting for association with PLD, therapy history and other differences in

[1] *PPD* includes terms such as *Phonological/Speech Sound/(In)Consistent Speech/(Developmental) Speech/Articulation + Disorder/Impairment/Delay/Errors* or *Developmental Verbal Apraxia*

study design, the analysis aimed to (a) summarize and quantify literacy outcomes, and (b) statistically explain the variation in these outcomes in terms of age and the selected PPD parameters.

In addition to selecting participant variables related to age and characterization of PPD, relevant literacy outcomes were identified. Spoken phonology has been regarded as directly related to word level literacy and only indirectly to comprehension (Perfetti, Landi, & Oakhill, 2005; Wood, Wade-Woolley, & Holliman, 2009). Thus, the meta-analysis examined the literacy constructs most likely connected to the organization of the speech sound system: (a) phonological awareness, (b) word decoding, (c) fluency, (d) nonword decoding, and (e) spelling. As background, the following sections describe key issues observed: characterization and evaluation of PPD and the literacy constructs included, and literacy outcomes for children with protracted versus typical phonological development.

2.2.1 Characterization of PPD

Authors have noted an overlap of PPD, Protracted Language Development (PLD), and lags in literacy attainment (Bird et al., 1995; Lewis, Freebairn, Hansen, Iyengar, & Taylor, 2004; Schuele, 2004), which Peterson, McGrath, Smith and Pennington (2007) have suggested could arise from a shared risk factor for phonological deficits but also factors unique to each. Studies of children with PPD in the absence of co-occurring PLD, have demonstrated inconsistency relative to literacy outcomes, with some results showing low risk (Catts, 1995; Meredith, 2002) and others, high risk (Catts et al., 2002; Peterson et al., 2009). Researchers have reported the lowest literacy outcomes for children described as having more severe or persistent PPD

(Larrivee & Catts, 1999; Leitão & Fletcher, 2004; Lewis & Freebairn, 2000). However, the characterization of PPD and its relative severity differ across the studies.

These differing perspectives on PPD have concerned the relevance of: (a) continuous (Bernhardt & Stemberger, 1998) versus discrete categories of severity (Shriberg & Kwiatkowski, 1982) versus sub-types of PPD (Dodd, 2005; Lewis, 2007; Snowling & Stackhouse, 1983); (b) resolution or persistence of PPD (Nathan, Goulandris, Stackhouse, & Snowling, 2004; Puranik et al., 2008); and (c) evaluation method with respect to both elicitation type (single-word versus connected speech), and word complexity (primarily mono- and di-syllabic words versus multisyllabic). Word complexity might also be relevant for characterization of PPD as 'resolved'; if insufficient numbers of MSWs were tested, ongoing difficulty with timing and sequencing of both syllables and speech sounds may be missed, at least up to the age of 7 years (James, 2006).

Further to word complexity, production of MSWs may be a key component in evaluation of the relationships between spoken phonology and literacy. Researchers have shown relationships between MSW production, phonological awareness skills, fluent word or nonword decoding, and spelling. For example, Core (2004) identified positive correlations between phonological awareness scores and Percent Consonants Correct (PCC) on a MSW naming task in 5- and 6-year-olds. Larrivee and Catts (1999) observed that MSW and nonword repetition in children with severe PPD at the end of kindergarten predicted outcomes in decoding (ability to pronounce words in print) and nonword decoding at the end of grade 1 and explained the greatest variance in decoding (including MSW *DEC*) in grade 3 (Ekelman, 1993). For ages 5 to 8;7 years, PCC of MSWs explained a larger significant unique variance in composite real and nonword decoding when included in a model with composite phonological awareness and lexical retrieval scores (McDowell, 2004).

In addition to differences in characterization of PPD, diversity in terms of literacy construct and/or measurement method are relevant in evaluation of the relationship between PPD and literacy, as discussed below.

2.2.2 Characterization of Literacy and the Included Constructs

In the following section, stages and theory of literacy acquisition are discussed, and the characterization and evaluation methods of the literacy constructs considered for the meta-analysis are described. First, in terms of literacy evaluation, the stage of acquisition is relevant. In earlier elementary school, formal literacy instruction begins and focuses chiefly on the letter names, sound-symbol correspondence and basic spelling patterns needed to decode frequent words. In later elementary school, literacy instruction places a greater emphasis on extracting meaning from text, which comprises words with decreasing frequency. In the high school years and beyond, students must decode and comprehend infrequent multimorphemic words, and comprehend increasingly abstract text in the context of less explicit literacy instruction (Chall, 1983). Theories of literacy strive to account for student advancement across these stages, with word level literacy considered a foundation for all of them (Duncan, 2009), as explained briefly below.

A major discussion of theories of word level literacy is beyond the scope of this paper. However, most theories assume spoken and print language to be connected through some aspect of phonological processing, particularly the representation of smaller speech units by letters (Frith, 2002; Hatcher & Snowling, 2005). The commonly described *Dual Route* models (e.g. Coltheart, 2005; Romani, Olson, & Di Betta, 2005) suggest that reading and spelling words involve two separate cognitive procedures, decoding and automatic sight recognition. The models aim to explain how knowledge of word sub-units (phonological awareness) enables reading of familiar versus unfamiliar words (word and nonword decoding) and their orthographic encoding (spelling). However, the constructs and evaluation of phonological awareness, decoding and spelling have varied across studies, as discussed in the following sub-sections.

2.2.2.1 *Phonological awareness (PA)*

Phonological awareness (PA) refers both to knowledge (implicit and explicit) that words are composed of sub-units (e.g. syllable, onset, rhyme, phoneme), and the ability to consciously manipulate these units mentally (Holm et al., 2008; Rvachew, 2003; Sutherland & Gillon, 2007). Studies have shown robust relationships between word level literacy and *PA* (Compton, 2002; Ehri, 1992; Perfetti et al., 2005), with a minimum proficiency deemed necessary for acquiring phonological to orthographic correspondences for monosyllabic (Tunmer & Hoover, 1992) and MSW word recognition (Lee, 2007). Phonological awareness is possibly promoted by strong cognitive networks that activate components of information in the output phonological system (Bernhardt & Stemberger, 1998). Cognitive resources are therefore more available for associating spoken phonology with orthographic chunks, or mental orthographic images (Apel, Wolter & Masterson, 2006).

Characterization and evaluation of *PA* have differed in conjunction with childhood cognitive growth. Task characterization has varied in terms of levels of meta-cognitive awareness and working memory, language recognition versus production and size of phonological units (Chabon, 1980; Rvachew, 2012; Snowling & Carroll, 2004; Webster et al., 1997). Studies have utilized standardized tests (e.g. Apel & Lawrence, 2011; McDowell, 2004; Sutherland & Gillon, 2007), nonstandardized tasks (e.g. Carroll & Snowling, 2004; Lewis &

Freebairn, 1992; Rvachew, 2007), or a combination thereof (e.g. Peterson et al., 2009; Raitano et al., 2004). Rhyme and onset matching tasks have predominated through age 6 years. For 7- to 8-year-olds, segmenting and blending have been common, sometimes of syllables and onsets but more often, phonemes. Phoneme manipulation tasks have frequently been administered from age 7 years, initially elision, and progressing to phoneme reversals at ages 10 through 13 years. The latter tasks quite possibly are a better reflection of working memory capacity (Rvachew, 2007). Children typically begin to decode and spell short words between 6 and 7 years of age, before they are able to perform phonological awareness tasks such as elision or reversal. Language processing theories would suggest that awareness of components of the prosodic word (syllables, stress, segments) are more relevant (Duncan, 2009).

2.2.2.2 *Decoding (DEC), Nonword decoding (NWD)* and *Fluency (FL)*

In terms of word reading, Ehri (1992) described the relationship between the dual routes for recognizing words, i.e. recoding, or *decoding (DEC)* and automatic sight recognition, *fluency (FL). DEC* is assumed to rely on intermediate application of grapheme-to-phoneme rules for translation of orthographic representations into suprasegmental structure and phonological segments (Duncan, 2009) before accessing meaning. As reading proficiency increases, readers learn progressively larger orthographic chunks, including syllabic patterns, with few exposures to a word. This enhances *FL*, suggesting stronger and faster cognitive connections from orthography to phonology and meaning (Ehri, 1992). All orthographic words may in fact connect to meaning and pronunciation concurrently (Bowey, 2005; Coltheart, 2005; Ehri, 2005), emphasizing the foundation of accurate spoken phonology for translating print into speech (Hester & Hodson, 2004; O'Connor, Bell, Harty, Kappel, & Sackor, 2000).

In *nonword decoding* (*NWD*), knowledge of grapheme-to-phoneme rules is required to read pseudowords, so is largely accomplished without the contribution of meaning (Bowey, 2005). As children develop representations of larger chunks of orthographic patterns, however, word neighborhoods might aid in recognition of parts of nonwords (Duncan, 2009). In child literacy development, the importance of *NWD* has come from evidence that acquisition of phoneme-to-grapheme correspondences depends on factors beyond phonological awareness (e.g. some children have performed well on *PA* tasks but not on tasks involving phonological to orthographic correspondence; Tunmer & Hoover, 1992).

In comparison with the variety of *PA* assessments, *DEC* and *NWD* have been primarily assessed on word lists from standardized tests (e.g. Carroll & Snowling, 2004; Harris, Botting, Myers, & Dodd, 2011) and infrequently on nonstandardized tasks (e.g. Bird et al., 1995; Harris et al., 2011). Standardized tests of *DEC* have been created in several contexts: North America (e.g. Peterson et al., 2009; Wellman et al., 2011), Great Britain (e.g. Bird et al., 1995; Snowling & Stackhouse, 1983), Australia (Holm et al., 2008); and New Zealand (Kirk & Gillon, 2007). Uniquely, Overby et al. (2012) determined the ecological validity of measures from the Templin Archive (2004) of longitudinal data from 1960-1972. Irrespective of country, the standardized measure that was used to evaluate *NWD* was the *Word Attack* subtest of the *Woodcock Reading Mastery Test*[2] (e.g. Apel & Lawrence, 2011; Catts, 1993; Kirk & Gillon, 2007; Wellman et al., 2011). For the few studies investigating *FL*, contrasting exemplars were a large database of passage reading assessments (Puranik et al., 2008) and a small sample that used word and nonword lists (Rvachew, 2007). Decontextualized single word and nonword decoding has

[2] Woodcock (1973, 1987, 1998)

allowed examiners to control for semantic and morphosyntactic cues that contribute to decoding in text; as a result, performance could vary on the differing tasks.

2.2.2.3 *Spelling (SP)*

Turning to the final construct, *spelling (SP)* requires representation of spoken words with conventional orthography (Kemp, 2009). Neuroimaging studies about accuracy and reaction time for regular, irregular and nonword *SP* recognition have supported both phonological and lexical routes for spelling (Norton, Kovelman, & Petitto, 2007). Because accurate, as opposed to phonologically legal spelling, must be generated (Critten & Pine, 2009; Romani et al., 2005), *SP* possibly relies on more detailed, explicit phonological to orthographic connections than reading. Ehri (1992) considered such full alphabetic skill to be hierarchically organized, with automaticity gained from strong long-term memory connections between *SP* and pronunciation. Representing words in print is further taxed by demands for coordination with fine motor output, integration with morphosyntactics and, if in text versus list form, organization of ideas.

In studies of children with PPD, *SP* has been uniformly assessed using word dictation but with various measures, including those studies using recently standardized tests (e.g. Burrus, 2007; Lewis & Freebairn, 1992; Meredith, 2002), older standardized tests (e.g. Bird et al., 1995; Holm et al., 2008; Overby, 2012; Stackhouse, 1982), or nonstandardized tasks (Apel & Lawrence, 2011; Kirk & Gillon, 2007; Nathan et al., 2004).

Keeping in mind the diversity in characterization and evaluation of PPD and literacy, the following section provides an overview of literacy comparisons between children with and without PPD as background for the meta-analysis. The term PPD will pertain to any diagnosed history, whether or not considered ongoing at the point of literacy measurement. Thus, the groups studied will be stratified only by age to highlight progression with literacy.

2.2.3 Literacy Outcomes of Children with PPD

The sections below will review the literacy outcomes for children with PPD with respect to the constructs related to word level literacy discussed in section 2.2, *phonological awareness*, *word decoding*, *nonword decoding*, *fluency* and *spelling*.

2.2.3.1 Outcomes in *Phonological Awareness (PA)*

Studies of *PA* have compared children with PPD versus TD in the early and late elementary school years, i.e. *learning to read* and *reading to learn* phases, respectively (Chall, 1983). Preschoolers with PPD (3- to 6-year-olds), have shown lags on composite score analyses (Core, 2004; Raitano et al., 2004; Rvachew, Ohberg, Grawburg, & Heyding, 2003) and awareness or matching subtests at the rhyme and word initial phoneme levels (Carroll & Snowling, 2004; Holm et al., 2008; Rvachew, 2007), with significance on whole word segmentation and blending (Chabon, 1980; Sutherland & Gillon, 2007). For preschoolers at 5 to 6 years of age, Raitano et al. (2004) demonstrated that persistent PPD had the strongest relationship with *PA* and a moderate *effect size* (*ES,* the magnitude and direction of actual differences).

At school age, studies have focused on the *learning to read* phase. Six- to eight-year-olds with PPD history have shown equivalence to children with TD on rhyme and phoneme awareness but significantly lower scores on various phoneme segmentation and blending tasks (Cowan & Moran, 1997; Harris et al., 2011; Webster et al., 1997). On the latter tasks, children up to age 9 years have scored within the average range, yet shown significant differences compared with TD children (Apel & Lawrence, 2011). Peterson et al. (2009) also found a significant difference and large *ES* on a *PA* composite for 7- to 9-year-olds; furthermore, controlling for nonverbal IQ, persistent PPD at age 4 years predicted ongoing *PA* difficulty.

Fewer studies have compared *PA* outcomes for older school-aged children or adolescents with PPD history versus TD. Also uncommon are studies evaluating longer words or manipulation tasks, such as elision or substitution of segments. School-aged children with and without PPD, predominantly 8- to 11-year-olds, have performed equally well on blending and segmentation tasks but with ceiling effects (Ekelman, 1993; Lewis & Freebairn, 1992; Meredith, 2002). On more advanced *PA* tasks of manipulation, such as reversed phoneme segmentation or Pig Latin, adolescents with PPD have lagged behind TD peers, irrespective of significance levels (Bird et al., 1995; Preston & Edwards, 2007); however, the relationship of such tasks to literacy is unclear.

Overall, children with PPD history from preschool through adolescence appear to lag behind TD children on *PA* composites or subtests of increasing complexity. *ESs* suggest moderate to large differences. Finally, PPD in early childhood is associated with lower *PA* outcomes, which in turn have predicted ongoing *PA* difficulty.

2.2.3.2 Outcomes in Decoding: *Word Decoding (DEC), Nonword Decoding (NWD)* and *Fluency (FL)*

Outcomes for decoding skills have been reported for the *learning to read* phase, and to a lesser extent, the *reading to learn* phase. In each phase, whether or not significant, lower scores have been reported for children with PPD than for those without. In the initial phase, despite ceiling effects, 5- to 7;6-year-olds with PPD have shown significantly lower scores on *DEC* and *NWD* than TD peers (Bird et al., 1995; Harris et al., 2011; Silva, 1984). Other studies have found average, but lower scores for children with PPD compared with those of their TD peers, with both significant (Overby et al., 2012) and non-significant differences (Catts, 1991; 1993; Holm et al., 2008).

Within wider age ranges, 6- to 9-year-olds with PPD have also had average achievement on *DEC* (Apel & Lawrence, 2011; Peterson et al., 2009), and in comparison with children with TD, superior performance has been reported for *NWD* subsequent to phonological awareness therapy, with a large *ES* (Kirk & Gillon, 2007). More broadly, small samples of school children ages 7 to 11 years, showed significantly lower *DEC* scores (Stackhouse, 1982; Snowling & Stackhouse, 1983) but using an upper limit of 13 years, *DEC* and *NWD* have been within average (Bishop & Clarkson, 2003; Wellman et al., 2011). Adolescents aged 12 to 17 years have also obtained average, but consistently lower scores on *DEC* and *NWD* (Lewis & Freebairn, 1992).

A final aspect of decoding is *FL*. For children with PPD, few studies of this construct exist, and those that do have focused mainly on the *learning to read* phase. For 5- to 6-year-olds with PPD, performance on words and nonwords has been in the average range, but significantly lower for the nonwords (Rvachew, 2007). Meredith (2002) found average word *FL* for 8- to 9-year-olds with PPD but with significant differences and large *ESs* compared with children with TD (Meredith, 2002). Finally, by grades 2 to 3, *FL* in connected text has been negatively correlated with persisting PPD (Peterson et al., 2009; Puranik et al., 2008).

To summarize the findings for the decoding attainment of school-age children with PPD history, average performance has been demonstrated on *DEC*, *NWD* and *FL*, although this performance may lag behind that of groups with TD. As was implied for *PA*, *ESs* suggest moderate to large magnitudes of difference for *DEC*, *NWD* and *FL*.

2.2.3.3 Outcomes in *Spelling (SP)*

More consistent results have been reported regarding *SP* proficiency, where children with history of PPD have performed significantly lower than their TD counterparts (Bird et al., 1995; Nathan et al., 2004; Overby et al., 2012), even during adolescence (Bishop & Clarkson, 2003; Lewis & Freebairn, 1992). A history of PPD has also predicted unique significant variance in *SP* (Apel & Lawrence, 2011). Nonetheless, some studies have shown performance within the average range for children with history of PPD (Bishop & Clarkson, 2003; Meredith, 2002; Overby et al., 2012).

Assuming equivalence between children with and without PPD in other developmental domains, the overall disadvantages in word level literacy of children with PPD are important for at least two reasons: (a) decoding of longer infrequent words and comprehension of increasingly abstract text might become more challenging, and consequently, (b) the range of available educational, career or employment choices might be narrower than for children with TD.

2.2.4 Questions for the Meta-Analysis

The review above suggests that overall, children with PPD history have lower literacy outcomes than their TD peers, even when their scores fall within the average range of tests normed on typically developing children. However, questions remain concerning the scale of difference for the selected word level constructs, the relative contributions of phase of childhood acquisition (earlier: *learning to read* versus later: *reading to learn*), and characterization and evaluation of PPD. The meta-analysis was thus undertaken in order to address those topics in terms of the following questions:

1. Are children with PPD consistently disadvantaged in word level literacy acquisition compared with TD peers, and if so, to what extent? Moderate to large magnitudes of difference were predicted between children with and without PPD, for each literacy construct: *Phonological Awareness, Decoding, Nonword Decoding* and *Spelling,* with the greatest group differences in *Spelling*. More protracted phonological development was expected to slow literacy acquisition because of the presumed connection of spoken phonology to word level literacy. Previous research has also shown that children with PPD have had lower performance than children with typically developing phonology on the following literacy constructs: phonological awareness, word decoding, nonword decoding, fluency and spelling. Lower performance for nonword decoding was predicted because the lack of a semantic component places greater reliance on the phonological to orthographic representations, in comparison with word decoding.
2. Are there between-group differences in both the early (*learning to read*) and later (*reading to learn*) childhood phases of literacy acquisition? Differences were predicted in both phases because research has shown that gaps in reading fluency achievement increase across the elementary school grades between children with and without PPD. Research has also suggested a higher incidence of reading disabilities for children with persistent PPD. Methods of characterizing PPD that reflected persistence were therefore expected to explain differences.
3. Are there differences between children with and without PPD that vary according to the characterization of PPD? The degree of protracted phonological development at time

of literacy measurement was predicted to explain differences between groups whether envisioned as discrete categories of severity and sub-type, or of resolution of PPD. The prediction was made on the basis that the characterizations of PPD were possibly equivalent constructs. Variation in methods used to characterize PPD in school-aged children in terms of context of assessment (i.e. MSWs and connected speech) was predicted to explain between-group differences. Speech context was expected to be significant because of the added length and complexity of word structure and phoneme sequences in MSWs or connected speech.

2.3 Method

This section first briefly describes the meta-analytic objectives, then outlines steps for the current analysis, including: search methods and study inclusion criteria, coding of moderators and reliability, data selection and transformation for *effect size* (*ES*) calculations, and methods for obtaining descriptive and analytic results.

2.3.1 Meta-analysis Objectives

Meta-analysis provides a statistical synthesis of comprehensive research findings obtained from different study designs and/or from sub-sets of data within studies (Hedges & Olkin, 1985; Lipsey & Wilson, 2000; Rosenthal, 1991). The included studies/data serve as the *participants* upon which hypotheses are tested. *Effect sizes* from selected data are standardized in order to aggregate and compare them. The method involves transformations, weighting and fitting of data to a statistical equation that takes into account study influence by sample size. These processes address bias arising from the sampling of studies, and/or the populations from which samples within studies are taken. While primary investigations explain variability in results according to dependent variables, meta-analyses examine independent variables (moderators) of interest and potentially specify gaps in research.

2.3.2 Study Selection

To obtain a comprehensive sample of studies of literacy outcomes for children with PPD, search terms included school age/ children/ childhood/ SSD/ speech/ articulation/ phonological/ disorders/ disabilities/ impairment / apraxia of speech/ phonological awareness/ literacy/ decoding/ word decoding/ word recognition/ fluency/ nonword decoding/ spelling/ comprehension/ reading/ dyslexia. The search for peer reviewed primary articles, conference proceedings and dissertations for any year included the main speech-language pathology, applied linguistics and education databases: Academic Search Complete, Cochrane Systematic Reviews, Communication and Mass Media Complete, CINAHL (Cumulative Index to Nursing and Allied Health Literature), Dissertation Abstracts International, Education Full Text (H.W. Wilson), Education Index Retrospective: 1929-1983 (H.W. Wilson), Education Research Complete, ERIC (Education Resource Information Center), Linguistics and Language Behavior Abstracts (LLBA), PsycARTICLES, PsycBOOKS, PsycINFO, PsycEXTRA, and Pub Med (1946-present). Subsequently, 5128 citations abstracts were reviewed (including duplicates in the various databases), from which 52 papers were evaluated, and their reference lists searched for additional candidates.

2.3.2.1 Inclusion Criteria

Inclusion criteria addressed aspects of study design, participant socio-behavioral characteristics, and available *effect size (ES)* data. Included studies measured at least one literacy construct of interest, in school-aged monolingual English-speaking children with history of PPD (without other notable developmental conditions) and a typically developing comparison group (TD).

With respect to data availability, *ESs* (or means and standard deviations from which to calculate *ESs*) either had to be published or obtainable from authors. Thirty-four studies, reported from 1980 to 2012, met the inclusion criteria, generating 64 independent samples (list available from author) distributed across constructs accordingly: *PA*, 20; *DEC*, 17; *FL*, 4; *NWD*, 9; and *SP*, 14. A total of 4007 children from 4 to 17 years of age (1496 with PPD and 2511 TD) was represented, with construct and corresponding sample size as: *PA*, 994 (527 with PPD; 467 TD); *DEC*, 1717 (410 with PPD; 1307 TD); *FL; NWD*, 479 (206 with PPD; 273 TD); and *SP*, 817 (353 with PPD; 464 TD). Few age spans were as narrow as one year: an age span of 3 to 3½ years and 4½ to 5½ years was covered in 50% and 40% of studies, respectively; in one study, the range was 6½ years.

In some cases, data clarifications and adjustments were needed before *ESs* could be determined. Clarifications that allowed data inclusion in the current meta-analysis were: (a) the independence of samples in the two sub-studies of Holm et al., 2008 (A. Holm, personal communication, November 22, 2011); (b) correction of a misprinted reading correlation value in Overby et al., 2012 (M. Overby, personal communication, February 27, 2012); and (c) obtaining unpublished data for Apel and Lawrence (2011), Sutherland and Gillon (2006), and Wellman et al. (2011). Catts (1991) and (1993) were treated as one longitudinal study because the second study followed up on the sample of children in the first. Finally, to improve equivalence of data and reduce the number of inferential tests for *PA,* composite scores were calculated if unreported (e.g. Peterson et al., 2009).

2.3.2.2 Coding of Included Studies

Coding was pertinent to the inclusion criteria and also to: publication identity, participant socio-behavioral attributes, literacy measures, and PPD characterization with consideration of co-occurring PLD. Coded socio-behavioral attributes were those used by researchers to equate comparison groups (and in some cases, tested for equivalence post hoc, e.g. Bird et al., 1995; Raitano et al., 2004), i.e. age, and one or more of the following: grade, classroom, IQ, language skills, reading/spelling level (when not a dependent literacy variable), ethnicity and SES/maternal education. Some indices combined these factors, e.g. Peterson et al. (2009) and Rvachew (2007). Coding was also done for specific tests of each literacy construct.

In relation to characterization of PPD, additional exclusion of studies was considered because of potential bias arising from co-occurring PLD. Language evaluation methods were reviewed in conjunction with coding the proportion of co-occurring PLD within included samples. Whereas a small percentage of studies referred to a clinical history without PLD diagnosis, 85% directly assessed language performance across a variety of comprehension and/or production skills. In the latter case, co-occurring PLD was defined by accepted clinical criteria (1.5 SD below the mean or criterion cut-off) and/or a significant post hoc comparison (half of included studies) between groups with and without PPD. A third of included studies (40% of those that directly evaluated PLD) reported no more than 25% co-occurring PLD. From this analysis, the coders agreed that there was reasonable control for confounding of literacy outcomes by co-occurring PLD and no further studies were excluded on that basis. Following examination of co-occurring PLD, PPD was characterized in terms of status and evaluation parameters. Statuses at time of diagnosis and at literacy measurement were coded with respect to: (a) traditional severity ratings, (b) subtype frameworks such as pattern consistency or *Childhood Apraxia of Speech* (*CAS*), and (c) resolution or persistence. With respect to resolution, mixed groups were coded as persistent if investigators reported few children with resolved PPD (e.g. 3 of 15 in Catts, 1993; 7 of 86 in Peterson et al., 2009). Coded evaluation parameters were the specific testing methods used, elicitation type (single words, connected speech) and word complexity (mono- and disyllabic versus multisyllabic). Coding of word complexity took into account that: (a) prominent stress occurs more commonly on word-initial syllables in English disyllabic (DSW) words; (b) adult-like production of DSWs with weak (unstressed) initial syllables presents a notable challenge in childhood phonological acquisition (Bernhardt & Stemberger, 1998); and (c) DSWs with initial unstressed syllables and MSWs, defined as three or more syllables, are infrequent on common norm-referenced tests (occurring less than 10% of the time; James, 2006). Consequently, a dedicated corpus of MSWs was coded whenever researchers evaluated an independent set of at least ten (the minimum number used in the included studies) DSWs with initial unstressed syllables and/or MSWs[3]. Tests with fewer than 10% complex words, e.g. the often-used *Goldman Fristoe Test of Articulation* (*GFTA*; Goldman & Fristoe, 1969, 1986), were coded for mono- and di-syllabic

[3] e.g. Catts (1986); Dodd et al. (2002)

words; otherwise, a mixture of mono-, di- and multi-syllabic words was coded. MSWs were not coded for connected speech elicitation, because type/token frequencies were unreported.

2.3.2.2.1 Reliability of Coding

Two Canadian certified and registered speech-language pathologists completed the coding of the included studies: (a) the first author, a doctoral student with dedicated coursework in meta-analysis and some 30 years of clinical experience, and (b) an SLP, holding a Master's degree with completed research project relating to PPD, and four years of clinical experience. The available *ES* data and the relevant study moderators were coded as described in sections 2.3.2.1 and 2.3.2.2, respectively. Coding definitions were discussed, and when not agreed upon, the first author's coding stood. The correlation and degree of agreement for multiple coders rating multiple moderators was estimated with the *intraclass correlation (ICC)* (Cohen, Cohen, West & Aiken, 2003). Because all coders coded all moderators, and to allow greater generalizability to other pairs of coders, the *ICC* was tested in a fully crossed random effects design. High reliability on all study moderators was indicated, *ICC(2,1)* = .976 [CI = .974-.978]. Regarding data selection and *ES* calculation reliability, differences were discussed to achieve 100% agreement.

2.3.3 Determination and Analysis of *ESs*

This section outlines the methods for the next three stages of the meta-analysis, beginning with obtaining *ESs* for the selected data, followed by the descriptive and then inferential analyses. The descriptive analysis applies to the distribution and weighting of *ESs*. The inferential process includes multiple regression analyses (MRAs) for identifying significant variability in any literacy construct, and subsequent testing of that variance on the selected moderators.

2.3.3.1 Calculation of *Cohen's d*

To allow examination of the standardized magnitudes (thus independent of sample size) of actual mean differences between groups with and without PPD, *Cohen's d* was the selected *ES* for the meta-analysis. To ensure independence of samples, one *ES* per literacy construct per study was chosen. For calculation of *d,* weighted means and/or pooled standard errors (i.e. proportionate to comparison group sizes) were derived from available raw scores, gains, percentiles, and standard scores. Because the value of *d* tends to be upwardly biased in small samples, by applying conversion formulas (Hunter & Schmidt, 1990), accuracy was adjusted to *unbiased d,* which was used in all subsequent analyses.

2.3.4 Descriptive Analysis

In the first step of the descriptive analysis, distributions of *unbiased d* were assessed through construction of funnel plots, which allow for the greater variance in small samples, reducing the number of outliers. The second step was to adjust *unbiased d* to obtain more precise estimates of the population *ESs,* and adjust study influence relative to sample size. To obtain this *weighted d*, *unbiased d* was weighted by the inverse variance (Hunter & Schmidt, 1990). The adjustment contributed to retention of outliers, and thus, all of the data in the inferential analyses. The magnitude of *weighted d* was interpreted according to Cohen's (1992) conventions (small ≥ 0.2; medium ≥ 0.5; large ≥ 0.8). For the *z-test* of the significance of each study *ES* and the mean for each construct, *95% Confidence Intervals* (*CI*) were constructed.

2.3.5 Inferential Analysis

Moving to the inferential analysis, two hierarchical multiple regression analyses ($p < .05$) were conducted to determine if any of the literacy constructs, *PA*, *DEC*, *NWD*, *FL* or *SP*, had significant variance (beyond that expected from sampling error) to explain by testing on the study and participant moderators. In the first regression, a fixed effects model (assumes all possible studies are included) analyzed *weighted ds.* However, the analyzed studies were likely estimating for different populations because of: (a) a small sample of *ESs* per construct; (b) limitations in attaining sample size power through random sampling of children with PPD; and (c) variation in comparison group equating among studies. The multiple regression was therefore repeated using a random effects model (assumes all possible studies are not included) on the *mixed weighted ESs*, which entailed re-weighting the *weighted ds* by adding the variability across populations of study samples (*Random Effects Variance Component*: *REVC*). The additional weighting yielded a more stringent regression because the resultant 95% *Credibility Interval (CredInt)* for the *mixed weighted d* could be expected to increase (more likely to contain zero), potentially decreasing the number of significant *ESs*.

2.3.5.1 Testing of Moderators

In this last stage of the meta-analysis, for each significant literacy construct, two groups of *mixed weighted ESs* were compared on a selected moderator by a *Q*-ANOVA (based on *Chi-square* for categorical variables, similar to 1-way ANOVA), $p < .05$. The result indicated whether the grouping variable (moderator) accounted for significant variability between the means of the *ESs* and whether there was significant variability (heterogeneity) remaining to be explained by other moderators. This process of testing moderators continued until either: (a) homogeneity of variances (no additional significant variance) was reached; (b) all moderators of interest had been tested; or (c) there were insufficient *ESs* (for this analysis, $n < 3$) to allow further meaningful analysis.

To address more extraneous influences of variables commonly associated with behavioral research, representative study design moderators were distinguished from participant characteristics directly pertinent to the research questions. Study design characteristics comprised: (a) cultural context, attributed to the various countries in which included studies were conducted; (b) gender balance (males comprising 50% to 60% versus ≥ 61% of a sample), because of conflicting information about gender differences in phonological development (Sander, 1972; Shriberg, Austin, Lewis, McSweeny, & Wilson, 1997; Smit, Hand, Freilinger, Bernthal & Bird, 1990); (c) sample size, because groups with TD were frequently larger than those with PPD; and (d) literacy evaluation type, due to variation in specific tests used to measure a given construct.

Directly related to the research questions were participant moderators, including mean age (within one-year intervals) and characterization of PPD at the point of literacy evaluation. The mean age grouping variable compared early (*learning to read*) with later (*reading to learn*) childhood literacy acquisition whereas characterization of PPD encompassed: (a) resolution, i.e. groups with and without a proportion of children with resolved PPD), (b) severity/subtypes, i.e. moderate to severe, with or without mild PPD), and (c) elicitation contexts, i.e. single words versus connected speech, and word length by syllables (MSWs). Further to severity/subtypes, those making reference to consistency (*delay, consistent and inconsistent*) were infrequently coded for. To increase available sample sizes for testing of moderators, the coders compared descriptions of these subtypes (B. Dodd, personal communication, January 7, 2013) with traditional severity ratings (*mild, moderate, severe*) and agreed to collapse: *delay* with *mild;* and *consistent* and *inconsistent* with *moderate to severe* and *severe*[4]. Bias was considered minimized by: (a) the small number of samples in the *mild* category coded for *delay* (*PA, NWD* and *SP* [n=18]; *DEC* [n=19]); and (b) grouping the *moderate to severe* and *severe* codes in any tests of moderators. Small sample sizes of groups coded with more severe or persistent PPD precluded separate analyses.

2.4 Results

The *effect size* data (*Cohen's d*) were examined to address the research questions concerning word level literacy acquisition of children with PPD in comparison to peers with typically developing phonology (TD). In this regard, the descriptive analysis first outlines to what extent children with PPD are disadvantaged in word level literacy. Moderate to large magnitudes of difference were predicted for each literacy construct, with the greatest in *SP*. Findings pertaining to these predictions will be discussed by construct in the following order: *Phonological Awareness (PA), Word Decoding (DEC), Fluency (FL), Nonword Decoding (NWD)* and *Spelling (SP)*.

In the subsequent section concerning inferential results, the results of multiple regression analyses (MRAs) of *ESs* and assessment of study design moderators are first presented. Then, the research questions about literacy acquisition in children with and without PPD are addressed. Regarding stage of literacy acquisition, differences were expected in both the early and later stages, with larger differences in the later stage, assuming a slower developmental trajectory in children with PPD (e.g. Puranik et al., 2008). The last question addressed potential differences

[4] *CAS* was also included in the latter group.

according to the characterization of PPD (e.g. Lewis & Freebairn, 1992). A first prediction was that degree of PPD at time of literacy measurement would explain variation, whether defined in terms of severity and sub-type, or by resolution. Second, it was predicted that speech evaluation method would also moderate word level literacy outcomes and support the need to assess phonologically complex speech contexts (i.e. MSWs and connected speech) in school-aged children.

2.4.1 Descriptive Analysis Results

Relevant to the descriptive results are the visual displays of data in funnel and forest plots for the *unbiased d* and *mixed weighted ds,* respectively. Data for four literacy constructs, *PA*, *DEC, NWD* and *SP,* are displayed in both plots but because the sample for *FL* was small (n = 4), only the related forest plot was constructed. The funnel plots depict the distributions of *unbiased d,* from which outliers can be visualized. The treatment of outliers was to adjust *ES* accuracy through weighting by the inverse variance, enabling retention of all data in the statistical analysis. For the *mixed weighted ds* (weighted additionally by the *REVC*), the forest plots convey the *ES* values and 95% *Confidence Interval*s (*CIs*) of each study and the mean *d*, with associated 95% *Credibility Interval*s (*CredInts*). The forest plots confirmed moderate to large effect sizes between children with and without PPD on the literacy constructs, as follows: moderate for *DEC* and *NWD* and large for *PA, FL* and *SP*. The means for *PA*, *DEC*, *NWD* and *SP* were significant but only that for *PA* was generalizable to other populations. Further details concerning these findings are reported below.

With reference to the funnel plots for *PA*, *DEC, NWD* and *SP* (see Figure 2.1 through Figure 2.4), distributions were normalized using logarithmic lines, setting limits at three standard deviations. In the subsequent assessment of outliers, attention was given to negative values of *unbiased d,* whereby children with PPD outperformed those with TD, a less expected outcome. Negative *ES* outliers were observed in the plots for *DEC* and *NWD* only and were few, specifically: *DEC,* 2 (n = 17); *NWD,* 3 (n = 9). In the raw data for *FL* (n = 4), there was one negative value. However, because the weighting procedures were expected to adjust both negative and all other outliers sufficiently, negative outliers were retained in the subsequent analyses. This procedure avoided selective data removal and consequent reduction of construct sample sizes.

The forest plots for each literacy construct, *PA*, *DEC*, *FL, NWD*, and *SP*, are displayed in Figure 2.5 through Figure 2.9. The respective plots show the range of the *mixed weighted d* for each included study, and the *mean* for the construct, with the interval estimates of significance (95% *CI* and 95% *CredInt*). Interval estimates that contained zero were not significant. For *PA,* (see Figure 2.5) two study *CIs* were without significance; however, the *CI* and *CredInt* indicated a significant large mean d_r^. For *DEC,* (see Figure 2.6), five study d_r^*s* were not significant but the *CI* suggested a significant moderate mean d_r^. For *FL*, Figure 2.9 demonstrated the significant large *ES* for the study by Puranik et al. (2008); however, the mean for the construct was not significant. For *NWD* (see Figure 2.7), three study d_r^*s* were without significance but the *CI* for the moderate mean d_r^ was significant. Regarding *SP* (Figure 2.8), the d_r^*s* of four studies were nonsignificant but the *CI* for the large mean d_r^ was significant.

In summary, for all literacy constructs, predictions of moderate to large differences were substantiated between the mean *mixed weighted ds* for groups with and without PPD. The largest *ES* was predicted for *SP*; however, the largest was found for *FL*, and the *ESs* for *PA* and *SP* were equivalent. In spite of the greater difficulty attributed to *NWD* than *DEC*, a moderate mean effect was obtained for both constructs. Concerning generalization, the mean *mixed weighted d* for the small *FL* sample was without significance. In contrast, the *95% CIs* for all other constructs suggested significant means, but only the *95% CredInt* for *PA* indicated that the *mean ES* was also generalizable to other possible populations.

2.4.2 Inferential Analysis Results

2.4.2.1 Multiple Regression Analysis Results

Two separate multiple regression analyses **(**MRA) were conducted to determine if there was significant variance ($p < .05$) among the *ESs* of any of the constructs to allow testing of moderators. Both fixed and random effects models[5] were implemented for the *weighted* and *mixed weighted ds*, respectively. A theoretical model was constructed that considered the incremental demands on phonological and orthographic knowledge, plus availability of semantic cues. This model was used to establish a hierarchical order for entering a single literacy construct at each step: *PA, DEC, FL, NWD* and *SP.* In the first regression, the R^2 *change* (*ES* for MRA) reached significance once *NWD* was entered, $F(3,60) = 6.17$, $p = .001$. The addition of *SP* also resulted in a significant model $F(4,60) = 4.55$, $p < .003$; however, the R^2 *change* was not significant, indicating that only *NWD* accounted for the significance of the last two models.

In the second regression of the *mixed weighted ds,* the R^2 *change* for all four models was significant (see Table 2.1). The constant, *PA*, was also significant, $t = 5.71$, $p < .001$. Therefore,

[5] Fixed effects model assumes all possible studies were included. Random effects model assumes that the analyzed studies were estimating for different populations.

all of *PA, DEC, NWD* and *SP* had additional variance to explain for the moderator analysis. Because the sample size for *FL* was too small to make meaningful comparisons on the moderators, it was not examined further.

2.4.2.2 Results for the Testing of Moderators

The next analysis concerned the significance of the moderators on the *Q*-tests of between-group variances for the mean *mixed weighted ds* ($p < .05$), first for study design and then for participant characteristics. For the literacy constructs, these results and those for the *Q*-tests of homogeneity of variances within groups ($p < .05$) are shown in Tables 2.2 to 2.6. Although the intent was to test all literacy constructs (*PA, DEC*, *NWD* and *SP*) on all moderators, constructs with insufficient samples of *ESs* were excluded from the analysis.

Study design moderators were initially tested aside from the participant characteristics, in order to address selected factors known to influence developmental behavioral research: cultural range, proportion of male participants, sample size and specific literacy measures used. The subsequent sequence of testing participant characteristics concerned differences in literacy outcomes resulting from earlier versus later childhood literacy ages, and characterization of PPD in terms of: severity/sub-types or resolution, and level of speech complexity evaluated (MSWs and connected speech).

2.4.2.2.1 Results for Study Characteristics

Table 2.2 summarizes the analyses of the study moderators. These moderators appeared to have minimal influence on the outcomes for the literacy constructs, with some exceptions to *SP*. For this construct, two moderators were significant, cultural context and measure used.

With regard to the first moderator, cultural context, the included studies were primarily from the United States and England, with a limited number from New Zealand, Scotland and Canada. All constructs were tested on cultural context, with nonsignificant results, with the exception of *SP*; for this construct, differences were evident between the United States and England, whether or not either country was grouped with New Zealand, Scotland and/or Canada.

The next moderator, gender ratio, was examined to ascertain whether literacy outcomes reflected proportions of girls and boys in the study. Balanced gender was defined as ≤ 60% male to allow latitude for the greater likelihood of protracted speech and language development in boys. The gender moderator was not significant for groups with PPD, and homogeneity was reached on one construct, *SP*.

Regarding groups with TD, gender ratio was balanced in all but one study of *NWD* (n=9), and so was not tested. Of the three remaining constructs (*PA, DEC* and *SP*)*,* only *PA* was significant, comparing respective large and moderate *ESs* for balanced and unbalanced gender. Unexpectedly, the larger *ES* pertained to groups with smaller as opposed to larger proportions of boys. Nevertheless considering the small number of *PA* studies lacking gender balance (4 of 18), it was concluded that gender ratio did not have a substantial influence on group literacy differences in the studies included in the meta-analysis.

With respect to the third study design moderator, sample size, significant results would have confirmed the expectation that larger *ESs* are related to larger samples. In this regard, comparisons of smaller with larger groups of children, with or without typically developing phonology, reached significance for all constructs, with two exceptions: (a) *PA,* for children with PPD, and (b) *DEC,* in the group with TD. With respect to children with PPD, across literacy constructs, the mean *weighted ds* for smaller samples ranged from small to moderate but were compared with consistently large *ESs* in larger samples, as expected. Regarding typically developing children, *ESs* were more various; in smaller samples, they ranged from small to large, and were moderate or large in larger samples. However, larger samples of children with and without PPD were related to the largest mean *ds*, as anticipated. Nevertheless, *ESs* in some small studies were possibly underestimating differences between children with and without PPD.

The last study moderator, literacy measure used, was tested for differences from two perspectives on standardization. The first analysis compared evaluations with and without standardization, and the second, between specific standardized measures. Nonstandardized evaluations were either too small (*DEC, SP*) or too various (*PA*) to be split further. With respect to *NWD*, the *Woodcock Reading Mastery Test* (*WRMT;* Woodcock, 1973, 1987, 1998)*,* was mainly used, leaving no meaningful groupings. However, of differences between standardized and nonstandardized measures for *PA, DEC* and *SP,* only *SP* was significant. Large *ESs* were compared on this construct, with the largest ascribed to methods without standardization. This suggested that the types of measures might have been administered differently and/or placed different amounts of emphasis on various orthographic patterns.

Moving to differences among standardized evaluations, comparisons were available for *DEC* and *SP*, with mixed results. In terms of *DEC*, there was no significant difference between measures used within and outside of North America. Conversely, for *SP*, the result was significant between the frequently used *Test of Written Spelling* (*TWS;* Larsen, Hammill, & Moats, 1999) and a remaining collection of evaluations. In sum, particular standardized tests did not account for significant variability in literacy outcomes; however, when nonstandardized *SP* measures were used, larger differences were found between children with and without PPD.

Overall, literacy outcomes between children with and without PPD were not significantly influenced by the study design parameters investigated: cultural range, gender balance, sample size or evaluation method. Exceptions to this concerned *SP,* for which differences were related to countries where studies were conducted (i.e. North America versus England), and to whether or not evaluation methods were standardized.

2.4.2.2.2 Results for Participant Characteristics

Turning to participant characteristics, Table 2.7 summarizes the results for the testing of earlier versus later childhood literacy mean ages, PPD severity and resolution, and level of speech complexity evaluated. Moderators examined for literacy phase and PPD characterization were associated with significant differences in *NWD* and *SP* outcomes, while speech complexity had a bearing on all constructs that could be tested (*PA, DEC* and *SP*)*,* as outlined below.

Concerning the first participant moderator, mean age at literacy evaluation, results were significant for *NWD* and *SP* (with adolescent groups removed, e.g. Lewis and Freebairn, 1992). The test for *SP* also reached homogeneity, suggesting that there was no further variance to explain. With the caveat of the small sample size for *NWD*, a large *ES* for children in the initial phase was compared to a negligible mean in the later phase, attributed to two of three studies in which children with PPD outperformed those with TD. For *SP* and *PA,* the largest *mean weighted d* was also associated with the earlier phase, but moderate *ESs* for *DEC* were virtually equivalent. Notwithstanding the incongruity of negative *ESs* for *NWD*, the results indicated greater differences between children with and without PPD in the earlier phase of literacy attainment, but that the range remained moderate to large.

The next tests of moderators concerned characterization of PPD in terms of severity (combined with subtypes) and resolution of PPD. Beginning with severity, studies tended to examine ranges from mild to severe, or moderate to severe, making tests on more discrete categories impossible. Comparing these two groups, however, significance was found for *NWD* and *SP* but not *PA* and *DEC*. Corresponding group differences for *NWD* were large and negligible, the latter as a result of children with PPD out-performing those with TD. For *SP*, significance related to large *mean weighted ds*, in contrast to nonsignificant moderate *ESs* for *PA* and *DEC*. An additional comparison for *PA*, traditional severity categories versus consistency subtypes, was also not significant between large *ESs.* Therefore, *SP* appeared set apart from the remaining constructs with regard to moderation by severity category.

The third participant moderator tested, resolution (i.e. resolved and unresolved), yielded a pattern consistent with that of severity: significant for *NWD* and *SP* but without significance for *PA* and *DEC.* However, for *NWD* and *DEC*, the resolved group was equally split as to full or mixed resolution, but few were of mixed resolution in the tests for *PA* and *SP*. For *NWD*, *mean ESs* between groups unresolved and mixed as to resolution, were large and negligible, respectively. Negative *ESs* in the selected studies accounted for the negligible *ES* of the latter group. For *SP*, large *ESs* were compared, the largest in the unresolved group. Nonsignificant group *ESs* for *PA* were also large but for *DEC*, were moderate.

In summary, whether PPD was characterized in terms of traditional severity categories, consistency subtypes or resolution, results were consistently significant for *SP* and *NWD*, with larger effects in groups with the most protracted development. However, *NWD* comparisons were unique in the proportion of negative *mixed weighted ds* attributed to very small samples (n = 3). In contrast, moderate to large nonsignificant differences for *PA* and *DEC* were maintained between children with and without PPD.

The last participant moderator regarding characterization of PPD was complexity of speech evaluation context. This was tested in two ways: on word length by syllables, and on evaluation of connected speech. For word length differences, *PA, DEC AND SP* were tested, and for connected speech evaluation, *PA* and *DEC* were tested, with results suggesting that speech complexity was a common relevant factor in literacy differences. All three constructs tested on word length showed significant results. *PA* was the only construct for which MSW evaluation could be compared in isolation of shorter words. Furthermore, the difference between large and moderate *mean ESs* was homogeneous, suggesting that the moderator of word length accounted for all of the variance. The tests for *DEC* and *SP* compared moderate and large effects, respectively, with the largest differences for studies that included dedicated MSW evaluation in combination with shorter words.

The final moderator tested concerned evaluation of connected speech, which in the studies was always supplemented with single word assessment. Therefore, the combined evaluations (connected speech plus single words) were compared with evaluation of solely single words. Results were contradictory for the constructs tested, *PA* and *DEC,* correspondingly with and without significance. The *ESs* compared for *PA* were large and moderate, and for *DEC*, both moderate. However, in both cases, the larger *mean* was ascribed to single word evaluation, suggesting that word complexity was the context more relevant to word level literacy.

The overarching findings of the tests of participant characteristic moderators were threefold. First, the *mean mixed weighted ds* for subgroups remained moderate to large between children with and without PPD on all constructs except *NWD*. This construct, the smallest in sample size, was more anomalous in its association with negative *ESs*, suggesting greater caution about the related findings. Second, characterization of PPD in terms of literacy acquisition phase, or severity as determined by traditional categories (with subtypes included) or resolution, was significant for *NWD* and *SP*, but not *PA* and *DEC*. Whether or not significant, larger *ESs* were consistently accorded groups with the most protracted phonological development. Third, for constructs tested on complexity of speech evaluation context, *PA, DEC* and *SP,* results were significant for the use of dedicated assessment of MSWs, whether or not in conjunction with shorter words. A significant result for *PA* on connected speech evaluation (combined with single word assessment) also suggested the importance of speech context evaluated. The inferential analysis thus completed, the results of the meta-analysis are discussed in turn below.

2.5 Discussion

The discussion begins with a brief review of the motivation of this meta-analysis of the relationship between protracted phonological development (PDD) and word level literacy. Subsequently, for the descriptive and inferential analyses in turn, methods used and corresponding results are summarized, and then conclusions are drawn. The descriptive analysis deals with the identification of outliers, and in addition, the *ESs* for each study and the means for the constructs. The inferential analysis includes the regression on the literacy constructs, followed by the moderating effects of study design and participant characteristics. From there, the discussion proceeds to clinical implications and limitations of the current meta-analysis, and concludes with suggested directions for future research.

2.5.1 Review of Motivation for the Meta-analysis

With respect to the motivation for this study, there were contradictory findings in the literature concerning the relationship of PPD and literacy acquisition (Catts, 1993; Meredith, 2002; Peterson et al., 2009). Clarifying the basis for differences could contribute to determining effective therapies for spoken phonology that in turn could mitigate effects on literacy. The meta-analysis set out to determine the convergence of evidence about the magnitude of literacy differences, and the moderating effects of central study and participant characteristics. The number of moderators was minimized in order to focus on the questions for the research and to reduce the probability of a Type I error.

The meta-analysis examined two possible major sources of variation in literacy outcomes between studies: (1) evaluation of specific word level literacy constructs of *Phonological Awareness (PA), Decoding (DEC), Fluency (FL), Nonword Decoding (NWD)* and *Spelling (SP)*; and (2) participant characteristics regarding age at literacy evaluation and/or characterization of PPD. The ensuing research questions asked whether differences: (a) were consistent across constructs; (b) varied by age in terms of stage of childhood literacy (earlier, *learning to read*, versus later, *reading to learn*); and (c) varied depending on researchers' characterization of PPD, as defined by severity/subtype, resolution, or the complexity of speech elicitation context (i.e. word length by syllables, or single words versus connected speech). Before answering these questions about participant characteristics, it was necessary to assess whether aspects of study design also explained variance in literacy outcomes; consequently, moderators of cultural context, sample size, gender and specific literacy measures were tested.

2.5.2 Review of Key Methodology for the Meta-analysis

A review of the methodology for the meta-analysis provides a qualifying context for the discussion of results that follows. Both descriptive and inferential analyses were conducted.

A descriptive analysis addressed the first question of the study: " Are children with PPD consistently disadvantaged in word level literacy acquisition compared with TD peers, and if so, to what extent?" Initially, *effect sizes (ESs)* for the mean differences between children with and without PPD (*unbiased Cohen's d*) were calculated for 67 independent samples identified in the 34 included studies. The *ESs* corresponding to all constructs were predicted to be moderate to large, with the largest expected for *SP*. Outliers were inspected visually in funnel plots. To treat the identified outliers, all *ESs* were weighted by two factors: (a) the *inverse variance*, to adjust study influence by sample size, and (b) the *Random Effects Variance Component* (*REVC*), because the included studies were likely estimating for different populations of studies. Then, by constructing forest plots of the *95% Confidence Intervals (CIs)* and the 95% *Credibility Intervals (Cred Ints)*, the significance and generalization of *ESs* were examined. The significance of study differences was judged in relation to the *CI* for the construct mean, which also suggested generalization of the construct mean, to the population for the sample of studies included in the meta-analysis. More stringently, the corresponding *Cred Ints* indicated whether the construct mean was representative of all possible populations of studies, completing the descriptive analysis.

An inferential analysis was subsequently conducted to address the primary research questions about participant characteristics. Initially, multiple regression analysis (MRA) was used to establish that there was sufficient variance among the *ESs* on any construct, to test the moderators. The conflicting literacy outcomes identified in the research suggested that all variances would be significant, predictions confirmed by the MRA ($p < .05$). This allowed the moderators of age at literacy measurement and characterization of PPD to be evaluated on the Q statistic (similar to *Chi-square* for categorical variables) whenever constructs had sufficient studies to group on the variables. Following from this requirement, *FL* was excluded from the moderator analyses because of its small sample size.

Before assessing the moderators for the research questions, the study design moderators were tested. Significance was considered possible for cultural context, gender, and specific literacy measures, but less likely for sample size. In terms of cultural context, potential significance related to the range of countries where studies were conducted. Second, concerning gender ratio, a major imbalance was expected to significantly bias *ES* magnitudes in favor of girls. Third, specific literacy measures were expected to be moderated by standardized as opposed to nonstandardized measures, because of the psychometric properties of the former. Further to standardized measures, any differences were predicted to relate to cultural contexts. Finally, it was expected that the wide ranges in sample sizes across studies, and between groups with and without PPD, were adjusted for adequately by the weighting of the *ESs.*

Following the testing of the study design characteristics, participant characteristics were evaluated in the inferential analysis. Concerning the first of these, age at literacy evaluation, it was anticipated that differences between children with and without PPD would have widened by the later childhood literacy phase. Second, the various methods for defining PPD, whether in terms of severity, subtypes or resolution, were all predicted to be significant for two reasons. First, assuming the connection between spoken phonology and word level literacy (Perfetti et al., 2005; Wood et al., 2009), more protracted phonological development could reasonably be expected to slow literacy acquisition. Second, it was possible that severity, subtypes and resolution essentially described equivalent constructs of PPD. Last, regarding speech elicitation context, significance was expected because of the added length and complexity of word structure and phoneme sequences in multisyllabic words or connected speech.

2.5.3 *Effect Size* Results

The first question for the meta-analysis concerned whether differences between children with and without PPD in word level literacy skills were consistent. As predicted, uniformity was demonstrated by moderate to large mean *ESs* between children with and without PPD, for all examined constructs: *PA, DEC, FL, NWD* and *SP*. However, in the examination of outliers, negative *ESs* were unexpectedly identified, indicating that children with PPD had outperformed their typically developing peers. This circumstance bears discussion before further considering the magnitude of the ESs overall.

There were two important statistical observations about negative *ESs*: (a) magnitudes were typically negligible (e.g. Catts, 1993; Sutherland & Gillon, 2007), and (b) all were without significance because their *confidence intervals* contained zero. The magnitudes therefore caused little impact on the mean *ES* for a given construct and were not generalizable beyond the primary study. Explanations for the negative direction might relate to several aspects of meta-analytic and primary methodologies, depending on the construct evaluated. Concerning *PA,* for selected sub-groups with PPD, provision of phonological awareness intervention (Sutherland & Gillon, 2007) could have enhanced performance on dependent *PA* variables. With respect to *DEC, FL* and *NWD*, the meta-analytic calculation of *ESs* from participants' gain on selected data, often resulted in a negative value. Consideration of this data, however, showed gain from below average to average-range scores for children with PPD, but from two average range scores for children with typically developing phonology. In some studies (e.g. Bird et al., 1995; Catts, 1991, 1993), both groups of children lost ground across measurement points, but children with PPD fell behind to a lesser extent. From these observations, it appeared that children with PPD had lower literacy at the outset, and as such, had less to lose.

There were also sample and construct biases in primary studies, unanalyzed in the meta-analytic process, which seemed to contribute to negative *ES* findings. Sample bias was probable in terms of: (a) heterogeneous samples stemming from wide age ranges (e.g. Burrus, 2007; Lewis & Freebairn, 1992); and (b) children with PPD having higher performance than comparison groups on constructs highly correlated with literacy, e.g. verbal language comprehension and/or production, or nonverbal IQ (Bishop & Clarkson, 2003). With respect to the latter, children with PPD have shown equivalent *DEC* and *NWD* attainment in earlier childhood literacy, but later demonstrated moderate differences in comprehension (e.g. Ekelman, 1993; Wellman et al., 2011). This is consistent with gaps that have widened between children with and without PPD in studies of *FL* (Puranik et al., 2008).

Having considered the influence of negative *ES* outliers, the moderate to large mean *ESs* between children with and without PPD are considered next. Such magnitudes were borne out for all constructs examined in word level literacy; nevertheless, construct-specific magnitudes were not altogether as expected. First, the largest differences were predicted for *SP (mixed weighted d = .84)* because it was theoretically the most complex skill of the included constructs. However, means were also large for *PA (mixed weighted d = .85)* and *FL (mixed weighted d = 3.04)*. Next, finding moderate *ESs* for both *DEC (mixed weighted d = .60)* and *NWD (mixed weighted d = .50)* was unanticipated because of suppositions that *NWD* reflects automaticity of sound-symbol relationships. Presumably, weak connections within the phonological system would in turn impact negatively on access to related orthographic patterns. Overall, the *ES* findings suggested that as a group, children with PPD could be expected to score about one standard deviation below peers with TD in *PA* and *SP,* and up to three standard deviations lower in *FL*. For *DEC* and *NWD,* performance of children with PPD appeared approximately one-half of a standard deviation below those with typically developing phonology. Because most studies accounted for SES factors and nonverbal intelligence, the *ESs* derived from the meta-analysis suggested that children with PPD were consistently underachieving in literacy. These differences apparently held even when children classified as having mild, moderate and severe PPD were examined together (due to inability to isolate sufficient sub-samples in the included studies).

From a theoretical perspective, explanations for the magnitude of *ESs* found in the current meta-analysis might make reference to accessing phonological and orthographic (i.e. letters) representations. Both types of knowledge are likely utilized to varying degrees from the earliest stages of reading and spelling with reciprocal influences (Lehtonen, 2009). Other researchers who shared this perspective (e.g. Blachman, 2000; McDougall & Hulme, 1994; Muter, 1994; Stanovich, 1992; Tunmer & Hoover, 1992) considered phonological representation to be a necessary but insufficient foundation for word decoding. Evidence of early orthographic knowledge comes from Apel, Wolter and Masterson (2006). These researchers reported that orthographic processing explained unique variance in young children's spelling detection and written output of novel words, given only minimal exposure to their forms in print. With respect to later literacy, Duncan (2009) defended the notion that both the phonological and orthographic representations have ongoing development in order to read and spell complex words.

In terms of accounting for the *ESs* specific to the included constructs, the likely complexity of interactions between phonological and orthographic representations leads to relative dependencies that are not uniform. Presumably, speech production is strongly connected to phonological representations that are involved in reading and spelling. Therefore, in tasks of *PA,* primary reliance on phonological organization up until *cracking the code* of print (decoding), could account for the large difference between children with and without PPD. As the ability to decode words progresses, reciprocal contributions of the phonological and orthographic systems advance *PA* ability. However, children with history of PPD might continue to be disadvantaged. For example, children might experience confusion over less transparent sound-symbol correspondences, that is, letters or letter patterns that are associated with more than one speech sound, and the converse, where a single speech sound relates to more than one letter pattern; or in another case, have difficulty understanding the relationship of a base word to related morphologically complex words (Apel & Lawrence, 2011).

Mutual dependence on phonological and orthographic representations is more probable for *DEC, FL, NWD* and *SP,* from the outset of their development. Between children with and without PPD, the large *ESs* for *FL* and *SP* could logically be precipitated by the added factor of skill automaticity. Automaticity is required for success in *FL* by definition (Puranik*,* 2008) but also is relevant to *SP* (Ehri, 1992) because of the additional cognitive capacities needed for writing the orthography (i.e. insofar as the evaluation methods used in the included studies). Moderate *ESs* for both *DEC* and *NWD* are more difficult to account for. In both of these processes, speed of connections between phonological and orthographic systems are likely determined by accessing phonological components of assorted sizes, including those related to lexical frequency and neighborhoods, and phonotactics, e.g. onset and rime patterns (Duncan, 2009; Lehtonen, 2009). The lexical restructuring theory suggests that stronger connections to lexical information would facilitate accuracy of *DEC* over *NWD*, in the same way that vocabulary growth is proposed to create ongoing refinement of the spoken phonological system. One consideration is that children with PPD *catch up* to some extent in *NWD* because there is a limited number of orthographic relationships to be learned in early literacy. With respect to *DEC*, however, the number of MSWs to be read increases during the same period (Cunningham, 1998; Duncan, 2009), words that are predictably more difficult to decode because of their length and complexity. Possibly the equivalent *ESs* for *DEC* and *NWD* in the meta-analysis reflect the growing demand on skills of *DEC*. Additional orthographic chunks will need to be learned, for example, corresponding to affixes, and connected to prior orthographic knowledge.

A final point of discussion about the mean *ESs* for the literacy constructs concerns generalization. The *95% CIs* were significant for all but *FL*, suggesting that the results for *PA, DEC, NWD* and *SP* were generalizable to other samples of studies described by equivalent moderators. As noted earlier, the very small sample of studies of *FL* (n = 4) and the unique large database sampled (n = 10,221) by Puranik et al. (2008) made it difficult to draw further conclusions beyond the current included studies. However, the large *ES* calculated from the *FL* data selected for the meta-analysis predicted that *ESs* of similar magnitude could also be expected given similar database samples and *FL* measures. In terms of generalization to populations of studies described by other moderators, only the *95% CredInt* for *PA* was significant, perhaps owing to the breadth of tasks (contact the author for supplemental information) making up the composite scores that were tested. The discussion now turns to the assessment of moderators included in the current meta-analysis, first with regard to study characteristics, and then participant characteristics of direct relevance to the remaining research questions.

2.5.4 Study Characteristic Moderators

To explain *ES* differences between children with and without PPD, the literacy outcomes were tested for study moderator influence before addressing the participant characteristics for the main research questions. With the exception of *NWD,* all constructs were tested on cultural context, sample size, gender, and specific literacy measures. *NWD* was excluded from testing on the latter two grouping variables because there were no meaningful splits available. An overall assessment of the results for the study characteristics suggested that effects were not even; however, certain patterns surfaced, including that: (a) gender ratio was primarily not a moderator, (b) sample size was a moderator for several constructs and, (c) effects were more frequent for *SP* than for any other construct. From this general assessment, a more in-depth discussion of the study characteristics selected for the current meta-analysis will first consider cultural context and measure used because of their associations with *SP,* followed by gender and sample size effects.

Concerning *SP,* cultural context and standardization of measures were significant moderators specific to this construct. Cultural context differences were pertinent to the two countries in which the majority of studies took place, the United States and England, but appeared unrelated to differences in particular standardized evaluations. Sub-groups of specific standardized measures compared a more recently published test, the *Test of Written Spelling* (*TWS*; Larsen, Hamill, & Moats, 1999), to older publications used outside of the United States (e.g. *Schonell Graded Word Spelling Test*, Schonell & Goodacre, 1974; *Graded Spelling Test*, Vernon, 1977), also without effect. Therefore, other explanations for cultural differences were considered with regard to spoken English dialect.

Researchers have suggested that it could be useful to confirm whether or not dialect influences skills related to word level literacy, e.g. morphological awareness (Apel and Lawrence, 2011). From the perspective of the included studies, differences in American and British spoken English dialects might account for the observed effect of cultural context on *SP*. That is, with regard to sonorous phonemes (vowels and consonants), spelling challenges might vary because of different pronunciations represented by the same orthography. For example, using tasks of phoneme identification in CVCC and CCVC spellings, Lehtonen (2009) has demonstrated the influence of sonority on *SP*. When sonorous consonants (e.g. *r* and *l*) appeared in consonants clusters in the words, both child and adult speakers of American English more often chunked the consonant with the preceding or following vowel. How speakers of American and British English might treat the same contexts would be of interest, because the dialects contrast as to whether spelled sonorant consonants are pronounced, e.g. /ɚ/ (schwar) or /l̩/ (syllabic "l") in words such as *water* or *bottle*, respectively. School instructional practices in *SP* might also vary accordingly.

Another study characteristic concerning *SP* was the significant effect of non-standardized evaluations, which differed from traditional standardization with basal and ceiling effects. For example, Nathan et al. (2004) analyzed the results of British Statutory Assessment Tests that evaluated the spelling of 30 words not normally used in the children's written compositions. Words included both monosyllables and MSWs with and without regular orthographic patterns. Based on whole word scores, children were assigned to one of three academic attainment levels expected for grade in school. However, some studies used methods other than whole word scoring, accounting for spelling of smaller word elements. Snowling and Stackhouse (1983) asked four children identified with *Childhood Apraxia of Speech* to spell 52 familiar CVC words comprised of stop and nasal consonants (i.e. *m, b, p, t, d, n, g,* and *k*). Consonant spelling was studied in relation to change across three places of articulation; therefore, vowel spelling was not considered. Apel and Lawrence (2011) scored phonemes and morphemes on a 15-word dictation test. Of the large *mean ESs* compared on the moderator of test standardization, the difference between children with and without PPD was larger in the group of studies that used nonstandardized evaluations. The *ESs* for the larger samples studied by Snowling and Stackhouse (n = 74), and Apel and Lawrence (n = 88) specifically, were also significant. This suggests that more information is gained from having children spell a representative sample of word structure and orthographic components, than from applying basal and ceiling effects.

The study characteristic concerning gender ratio was tested independently within groups with and without PPD. Both groups were tested on the gender moderator for *PA, DEC* and *SP,* and in addition for *NWD,* in the group with PPD. Gender ratio in the group with TD was a moderator for *PA* and *SP,* but was without significance in any of the other tests. Normative information about childhood phoneme acquisition has suggested gender differences (Smit, Hand, Freilinger, Bernthal & Bird, 1990) but whenever reported, has favored girls (McLeod, 2009). Therefore, in relation to children with PPD, an effect due to higher male proportions in the group with TD would predictably have reduced *ES* differences. Because the largest differences, however, were found for *PA* and *SP*, the suggestion was that gender proportions did not account for the findings for the TD group.

Sample size was also independently evaluated in groups with and without PPD. Sample sizes were often between non-equivalent groups, the largest mainly ascribed to children with TD (e.g. McDowell, 2004; Overby et al., 2012; Silva et al., 1984) because recruiting large samples of participants with PPD is frequently difficult. In the tests of the sample size moderator, only two tests were not significant: children with PPD in studies of *PA*, and those with TD in studies of *DEC*. In the significant comparisons, the largest *mean ES* was attributed to the group of studies with larger samples. Therefore, it is possible that *ESs* were underestimated in the studies with smaller samples of either group of children.

2.5.5 Participant Characteristics

The discussion moves on to the moderators of central interest in the current meta-analysis, the participant characteristics. These were defined in terms of age range at literacy evaluation, i.e. early (*learning to read*) and later childhood literacy phases (*reading to learn*), plus parameters characterizing PPD, i.e. severity/sub-types, resolution and complexity of speech context evaluated. In the following section, the results for the tests of each moderator will be discussed in turn and conclude with the support they lend to theoretical views about the course of literacy acquisition for children with PPD.

Regarding age range or phase of literacy development, synopses in the literature offered mixed conclusions about whether there are persistent differences in the literacy of children with and without PPD (e.g. Apel & Lawrence, 2011; Lewis & Freebairn, 1992; Peterson et al., 2009; Puranik et al., 2008; Raitano et al., 2004). In response to this dilemma, the current meta-analysis results suggested ongoing large or moderate differences in *PA, DEC* and *SP*, but not in *NWD*. Also relevant to these differences was whether magnitudes were smaller in the later childhood phase, with variation in results among the constructs. For *PA* and *DEC,* moderate differences were maintained but in *SP*, differences declined from large to moderate in the later age groups. The *ES* for *NWD* also decreased, but from large to negligible, an unexpected finding given that performance of this skill theoretically is highly dependent on the phonological to orthographic relationships of literacy. This unique outcome for *NWD* in the later childhood literacy phase must be considered with caution because the *mean mixed weighted d* was obtained from small samples of children (n = 100) and studies (n = 3), and in two of these, children with PPD outperformed those with TD. Nevertheless, if the difference in *NWD* between children with and without PPD does close by the later phase of childhood literacy, the gain might in turn explain the magnitudes of difference for the other literacy constructs. Researchers have expressed that proficient non-word decoding will enable fluent word decoding (Puranik et al., 2008) and spelling (Ehri, 2009), with reciprocal effects on *PA* (Apel & Lawrence, 2011). Stronger phoneme-to-grapheme correspondences demonstrated in *NWD* would in turn spill over to *SP* (Duncan, 2009). Thus, *SP* proficiency might align with *PA* and *DEC* abilities, as was the case in the current tests of the moderator (i.e. all three constructs with moderate *ESs* in the later phase of childhood literacy). Facility with *NWD* might also have been sufficient to prevent *PA* and *DEC* differences from increasing, and for some children with PPD to attain word level literacy in the average range. The discussion next considers whether parameters specific to describing PPD have moderating effects comparable to those for age range.

Characterization of PPD was also determined to be relevant. The first characteristic to be examined was the children's status in terms of resolution of PPD. Coding on this parameter generated studies of children whose PPD was unresolved, resolved or mixed (a combination of both). In order to avoid confounding the group of studies of resolved PPD, studies of mixed samples were added to those that examined children with unresolved PPD exclusively. This strategy potentially entailed a decrease in the *ES* for the group mixed as to resolution of PPD. However, for each literacy construct, the proportion of studies of unresolved PPD was always greater (range = 0.67 to 0.80), with relative increases in the *ES*. This situation helped to minimize the confound.

The tests of the resolution moderator resembled those for age range in two respects: (a) the magnitudes of the sub-group mean *ESs,* and (b) the significance of results. For *PA* and *SP*, the mean *ESs* were large, and for *DEC* were moderate, whereas large and negligible *ESs* were tested for *NWD*. Results were significant for *SP* and *NWD,* but not for *PA* and *DEC*. Whether or not significant, differences between children with PPD and TD were not confined to groups of children with solely unresolved PPD. The results for *NWD* were again exceptional, suggesting that children with resolved PPD history did not differ from children with TD. However, as described above, the small sample of studies of *NWD*, including those in which children with PPD outperformed those with TD, suggested caution in interpreting this outcome.

The differing results among the literacy constructs on the tests of resolution were possibly related to the numerous methods that determined status of PPD, used singly and in combinations. Frequently, percent consonants and/or vowels correct (PCC/PVC) were determined for published elicited single word tests or conversational samples. Criterion-referenced cut-offs were established for these measures in terms of percentile rank, standard score, number of phonological processes, or whole word production inconsistency percentage. A summary of these various evaluations included: (a) SDCS categorical metric for Normalized Speech Sound Disorders (75% of children by 6 years of age; Shriberg, 1994), (b) conversational PCC, (c) *GFTA* PCC ranking cut-off (range: 33rd to 15th PR), (d) mean *GFTA* PCC, (e) *Photo Articulation Test* (*PAT:* Lippke, Dickey, Selmar, & Soder, 1997): PCC+PVC SD cut-off (1SD from the mean *z*-score for age); (f) *Khan-Lewis Phonological Analysis* (*KLPA-2*: Khan & Lewis, 2002) number of phonological processes (3 = moderate; 4 = excessive), (g) *Computerized*

Articulation and Phonology Evaluation System (CAPES: Masterson & Bernhardt, 2001) PCC on combinations of Levels from 1 to 4; (h) *Diagnostic Evaluation of Articulation and Phonology* (*DEAP*: Dodd, Hua, Holm, Crosbie, & Ozanne, 2002) inconsistency percentage cut-off (< 40%, of whole word productions). Puranik et al. (2008) described varied Florida Department of Education therapy eligibility criteria that referred to either: number and years of delayed consonants in relation to normative acquisition data, intelligibility judgment, disordered error patterns, or rating of moderate to severe impairment on an articulation rating scale. Nevertheless, the similarity of results for the tests of resolution compared with age at literacy measurement, suggested that both moderators indexed growth in phonology and literacy over time. The study by Puranik et al. (2008) presented such a trajectory for children with PPD in terms of *FL*, from Grade 1 through Grade 3, with marginal differences indicated between children with persistent as opposed to resolved PPD. Resolution of PPD has also been characterized in terms of traditional categories or sub-types, the discussion of which follows.

Severity/sub-types relative to characterization of PPD, were tested in two ways; by comparing traditional severity categories with inconsistency sub-types and by comparing rankings within severity categories. Sufficient studies existed for testing the inconsistency subtypes only for *PA*, with nonsignificant results. Although limited to *PA*, this outcome supported the validity of combining the subtypes with the traditional severity categories, affording a larger sample size for the second comparison. Within severity categories, tests of moderate to severe and mild PPD were conducted. The magnitudes of sub-group *ESs* for each construct replicated those used in the tests of mean age and resolution, in addition to iterating the patterns of significance, i.e. significant results for *NWD* and *SP* but nonsignificant for *PA* and *DEC*. For all literacy constructs, the analogous results for resolution and severity/subtype categories suggested that the moderators defined equivalent constructs of PPD, representing the course of the children's phonological development. For some children not identified with PPD until the early phase in which literacy was also evaluated, severity/subtype and resolution coincided. Where children were identified prior to the later phase in which literacy was measured, severity/subtype and resolution measured status of PPD at a different point on the continuum.

Further to evaluation of word length, evaluation of MSWs was an important moderator, with significant effects on all of the tested constructs, *PA*, *DEC* and *SP.* Significance was found between evaluations of mono- and disyllabic words, and assessment of MSWs with or without mono- and disyllabic words. These findings suggested that the addition of MSWs provided supplemental information about the children's phonological development. For *PA,* the effect of MSW evaluation was further highlighted when studies that had amalgamated mono-, di- and multisyllabic words were removed, resulting in a significant homogeneous comparison of mono- and di-syllables with MSWs. Regarding *ES* magnitudes, the studies that evaluated MSWs showed a large difference between children with and without PPD, whereas the difference was moderate for studies that assessed principally mono- and di-syllables. With respect to *DEC*, moderate *ESs* were tested and for *SP*, *ESs* were large, as opposed to moderate, when more MSWs were evaluated.

The final examination of speech context compared studies with only single words versus both single words with connected speech. This comparison was considered legitimate because assessment of phonology in connected speech generally also entails single word elicitations (Skahan, Watson, & Lof, 2007). Of the literacy constructs, only *PA* and *DEC* could be tested on the moderator of connected speech assessment, with *PA* significant and *DEC* not. Concerning

PA, the magnitudes of the *ESs* in the significant comparison for single words were larger than were tested in the combination of single words and connected speech. This is consistent with Shriberg and Morrison (1992), who reported that in comparison with single-word elicitation tests, words in conversational samples were simpler in structure, and included a relatively lower proportion of multisyllabic words.

In summary, for the participant characteristics, three patterns were identified. First, *NWD* and *SP* were moderated by all of the participant characteristics, whereas only the evaluation method for PPD had effects on *PA* and *DEC*. These results suggest that *NWD* and *SP* may be more sensitive to strength of, or access to underlying phonological representations, possibly because these skills rely more on systematic connections between phonological and orthographic representations (Ehri, 1992, 2005). A second pattern was the significance of PPD assessment methods for all constructs tested (i.e. *PA*, *DEC* and *SP*). Phonological evaluation methods appear to be a crucial variable in differentiating literacy performance of children with and without PPD. This result also supported the relevance of evaluating child phonology on a continuous scale of measurement (Peterson et al., 2009) rather than in terms of discrete categories related to severity/sub-type or resolution. The third pattern concerned specific contexts of phonological assessment, especially MSW assessment. Studies which included more complex phonological contexts during assessment, showed more systematic relationships with literacy acquisition. This has clinical implications, as discussed below in the conclusions.

2.5.6 Theoretical Views Concerning Lower Literacy in Children with PPD

Final comments concern theoretical views about the relationship of PPD and lower literacy skills. Both one-dimensional and multi-dimensional interpretations have appeared in the literature to explain this relationship. In the first group, explanations refer either to an underlying cognitive process or to the developmental timeline of child phonology. For example, one view is that a *core phonological deficit* (e.g. Stanovich, 1992) contributes to poor phonological representations that slow down the acquisition of literacy skills. In terms of the pace of phonological development, supporters of *the critical age hypothesis* (e.g. Nathan et al., 2004) suggest that the impact of PPD on literacy is mitigated significantly by resolution prior to formal literacy instruction. Other recent perspectives have favored the interaction of multiple components in word level literacy, in terms of: (a) multiple linguistic and nonverbal cognitive deficits (e.g. Peterson et al., 2009; Raitano et al., 2004); (b) synergies between language-based skills (e.g. Lewis & Freebairn, 1992), and (c) overlaps of multiple levels of linguistic awareness, including phonemic, morpho-syntactic and orthographic components (Apel & Lawrence, 2011; Kirk & Gillon, 2007). Because of the various moderating effects of the participant characteristics, the current meta-analysis substantiates that a version of multiple component interaction is better able to explain the relationship between PPD and word level literacy acquisition. Consistently significant moderating effects on all of the literacy constructs that would have been needed to support a core deficit in phonological representations in children with PPD were not found. Nor was there support for a critical age by which phonology must be *normalized* in order for children with and without PPD to have equivalent literacy outcomes. Differences in literacy were moderate to large whether evaluated in the pre- early literacy or post-early literacy phase. Spelling differences were larger for children whose PPD was considered more severe or unresolved at the point of literacy evaluation, but moderate to large differences were also found for spelling, phonological awareness and decoding when PPD was considered less severe or resolved. Conceptions about the interaction of multiple linguistic components is more satisfying in terms of both processing models described by connectionism (Stemberger, 1992) and nonlinear linguistic hierarchies of spoken phonology (Bernhardt & Stemberger, 1998). These explanations were also more consistent with the significant effects of speech production context, particularly with regard to the complexity of MSWs. Connectionist language processing theories and theories of nonlinear phonology will be considered again in the concluding discussion in Chapter 5. In the next section, the chapter concludes with limitations to the meta-analysis and the implications for future research. Implications for clinical practice are also considered.

2.6 Limitations and Suggestions for Future Research

The current meta-analysis was the first known to systematically examine the relationship of PPD and literacy. Moderate to large differences were found between children with and without PPD on all of the word level literacy constructs, and were larger in the earlier than in the later childhood literacy phase. Cultural context (i.e. United States versus England) was a moderator of *SP,* likely as a result of differences in English dialect. Among the moderators characterizing PPD, all were significant for *NWD* and *SP*, whereas only evaluation method had effects on *PA* and *DEC*. The assessment of phonology in complex linguistic contexts moderated all literacy constructs tested (i.e. *PA*, *DEC* and *SP*), mainly regarding evaluation of MSWs, but assessment of phonology in connected speech was also significant for *PA*.

The meta-analysis was limited by the restricted number of studies that could be included, and the small samples generated for each of the literacy constructs (*PA, DEC, FL, NWD* and *SP*), especially for *FL* and *NWD*. Fewer than 25 samples were identified for any of the examined constructs, and fewer than 10 for *NWD* and *FL.* The findings for *PA* were likely the most robust because of the larger sample of studies. The large differences for *FL* between children with and without PPD were not significant, and thus the construct had to be excluded from the moderator analyses. However, differences for other constructs were significant, and hence generalizable to the population of the included studies (large differences for *PA* and *SP,* and moderate differences for *DEC* and *NWD*). The differences for *PA, NWD* and *SP* were also generalizable to other possible populations. In the future, additional studies of all chosen literacy constructs are needed, particularly of *NWD* and *FL*. More studies could enhance generalizability of findings through a replication of the meta-analysis that evaluates all constructs on all chosen moderators.

Another limitation of the study relative to the restricted set of studies examined was the number of moderators tested, done in order to reduce the possibility of Type I errors in the significant *Q-tests*. Testing additional moderators in the future might help account for the heterogeneity within sub-groups, particularly because few tests of the moderators had a homogeneous result. Homogeneity was limited to the moderators of mean age at literacy evaluation for *PA*, and for *SP*, gender in the PPD group and characterization of PPD in terms of word length. In future research, evaluating language development as a moderator could inform the relative effects of language and phonology on literacy (Lewis & Freebairn, 1992; Peterson et al., 2009), because it is difficult to isolate PPD from generally protracted language development. Testing narrower age ranges may uncover more incremental differences in literacy acquisition. For example, for *FL,* Puranik et al. (2008) showed developmental trajectories in age levels from one school grade to another, using growth curve modeling.

Methodologies in the studies included also limit the impact of the meta-analysis. There were, overall, relatively low numbers of children with PPD in the selected studies. Researchers have commonly commented on the challenge of acquiring large enough samples with power to detect differences between children with and without PPD. Within the included studies, two-thirds sampled fewer than 25 children in either of the compared groups and in the remaining one-third, rarely more than 100 children comprised either group. To address the issue of sample size, researchers have recently turned to secondary data from databases, for retrospective study designs (e.g. Overby et al., 2012; Puranik et al., 2008). Primary data might also be retrieved (e.g. CHILDES; MacWhinney, 2000) to design larger prospective studies. A caution about these alternate methodologies is the inability to ensure equivalence of procedures for selecting participants and gathering data. Therefore, comparing findings from database samples to those in primary studies would be a valuable future endeavor.

Another issue regarding sampling methodology in the selected studies concerns participant characteristics. For the current meta-analysis, studies had insufficient groups of children with moderate to severe PPD, confirming that more studies are needed concerning the literacy of these children (e.g. Lewis, 2007). To increase the validity of categorical characterization of PPD and the ability to compare research results, researchers may need to standardize the evaluations and criteria for sub-group assignment. Alternatively, according to researchers such as Peterson et al. (2009), relationships between PPD and literacy could be evaluated on a continuous scale. In this way, power can also be addressed, while avoiding semantic confusion of category labels and the arbitrary nature of participant assignment. Continuous scales could be incorporated in longitudinal designs to better follow the course of PPD and draw conclusions about the strength and direction of the relationships to literacy. Longitudinal designs might further include children with history of PPD whose early childhood literacy seems satisfactory, but whose later literacy is compromised in some way, especially in reading comprehension and written composition (Lipka et al., 2006).

2.7 Clinical Implications

Another important consideration for evaluating PPD status in future studies, the assessment of competence in complex linguistic contexts, is also relevant to clinical practice. Evaluation of linguistic complexity pertains not only to phonological production in connected speech (Peterson et al., 2009; Rvachew, 2007), but also in multisyllabic words (Ekelman, 1993; James, 2006; Lewis & Freebairn, 1992). However, there is no theoretically sound norm-referenced multisyllabic word evaluation method that is easily replicable for single word and connected speech elicitation. In research, words tested have varied (e.g. Carroll & Snowling, 2004; Core, 2004; Preston & Edwards, 2007) and in some cases, variability of whole word production has been evaluated (e.g. Harris et al., 2011; Sutherland & Gillon, 2006). Creation of a valid MSW corpus necessarily will entail inclusion of representative word structure and consonant sequences, and consideration of word familiarity in terms of cultural and academic contexts, and imageability.

The connection between spoken MSW production and literacy demonstrated by the meta-analysis is of clinical relevance in terms of planning intervention for children with PPD. Researchers in literacy have stressed the need for ongoing elaboration of the underlying structure needed for support of reading and spelling more complex words (e.g. Duncan, 2009). This implies that information gained from parallel evaluations of speaking, phonological awareness, reading and writing of MSWs might allow speech-language pathologists to better tailor therapy programs. For example, beyond whole word scoring of *SP*, more detailed rubrics have been applied to children's spellings, in order to account for the phonological components of words that have not been represented orthographically (e.g. Burrus, 2007). Even without benefit of comparable measures for other word level literacy constructs (i.e. *PA*, *DEC* and *SP*), MSWs selected for speech therapy could also be practiced in meaningful word level literacy to reinforce connections between phonological, orthographic and semantic components.

A final clinical implication of the meta-analysis is the additional support for early intervention. Larger differences between children with and without PPD were observed in the early childhood as opposed to later childhood literacy phase. Potentially, earlier therapy that contributes to progress in the early phase would catalyze reduction of later differences. Increases in children's performance in *PA* to average or better levels following early intervention (Kirk & Gillon, 2007) suggest that similar effects are possible for other word level literacy constructs.

The studies examined here enabled the current meta-analysis by contributing a body of data concerning the relationship of PPD to word level literacy. The meta-analysis confirmed that children with PPD are at risk for literacy development. The findings underline the importance of evaluating PPD in complex linguistic contexts, including MSWs, in order not to miss children who appear to have mastered the basics of spoken phonology but continue to be challenged by more complex phonology. Future meta-analyses of larger samples of studies are needed that relate literacy attainment to PPD measured on a continuous metric, and that also examine outcomes of different phonological and literacy interventions in the preschool and early school years.

In order to further quantify and examine the phonology of MSW productions of school age children, a measure was needed. In Chapter 3, therefore, the pilot study of a whole word MSW metric is presented.

Table 2.1 Model summary for Multiple Regression of *mixed weighted d* for the literacy constructs

Model	R	R^2	$Adj\ R^2$	Se	ΔR^2	ΔF	df1	df2	Sig ΔF
*PA*DEC*	.345	0.119	0.105	1.87	0.12	8.35	1	62	0.005
*PA*DEC*FL*	.475	0.226	0.2	1.77	0.11	8.44	1	61	0.005
*PA*DEC*FL*NWD*	.534	0.285	0.249	1.71	0.06	4.96	1	60	0.03
*PA*DEC*FL*NWD*SP*	.842	0.709	0.689	1.10	0.42	85.96	1	59	0

Note. PA = Phonological awareness. DEC =Word decoding. NWD = Nonword decoding. FL = Fluency. SP = Spelling.
Result for the constant, *PA*: $t = .571$, $p < .001$.
$^*p < .05$

Table 2.2 Study moderators tested and significance across literacy constructs

Moderator	Groups tested	*PA*	*DEC*	*NWD*	*SP*
Cultural context	US vs England (with/without other[a])	✓	✓	✓	*
Proportion male participants					
PPD	Near 50% to 60% vs 61 to 70%+	✓	✓	✓	✓
TD	Near 50% to 60% vs 61 to 70%+	*	✓		*
Sample size					
n_{PPD}	4 to 25 vs 30 to 98	✓	*	*	*
n_{TD}	4 to 44 vs 98 to 815	*	✓	*	*
Measures of construct[b]	Standardized & nonstandardized	✓	✓		*
	Specific standardized		✓		✓

Note. PA = Phonological awareness. DEC = Word decoding. NWD = Nonword decoding. SP = Spelling. ✓ = construct tested on the moderator. PPD = Protracted phonological development. TD = Typical phonological development.
[a]Other = New Zealand, Scotland &/or Canada. [b]Because *NWD* was typically evaluated with the *Woodcock Reading Mastery Tests* (*WRMT; Woodcock, 1987, 1998*), no test of the moderator was available.
*p < .05

Table 2.3 Participant moderators tested and significance across literacy constructs

Moderator	Groups tested	*PA*	*DEC*	*NWD*	*SP*
Range of mean age[a]	4;0 to 6;11 vs 7;0 to 8:11	✓			
	5;0 to 7;11 vs 8;0 to 10;11		✓	*	*
Resolution	Resolved vs Unresolved	✓	✓	*	*
Severity	Mild & moderate with/without severe vs Moderate &/or severe	✓	✓	*	*
Measures of PPD at literacy					
Word lengths	Mono- & di-syllabic vs Multi- with/without Mono- & di- syllabic	*	*		*
Speech context	Single words vs Single words & connected speech	*	✓		

Note. PA = Phonological awareness. DEC = Word decoding. NWD = Nonword decoding. SP = Spelling. ✓ = construct tested on the moderator. PPD = Protracted phonological development. TD = Typical phonological development.

[a]Age at which literacy was evaluated, within closest approximations of earlier and later phases of childhood literacy (Chall,1983).

**p* < .05

Table 2.4 Results of tests of the moderators for *Phonological awareness* (n = 20)

Study characteristics						
Moderator	Groups tested	*mixed weighted ES* ($\hat{d_r}$)	n	Test for HOV	*Q* between groups	*Q* within groups
Cultural context	US vs England & other[a]	0.83 vs 0.73	13 7	*Q*(19) = 41.12*	2.67	*Q*(18) = 38.45*
Proportion male participants						
PPD	Near 50% to 60% vs 61% to 70%+	0.78 vs 0.90	9 9	*Q*(17) = 39.96*	3.14	*Q*(16) = 36.82*
TD	Near 50% to 60% vs 61% to 70%+	0.87 vs 0.45	14 4	*Q*(17) = 40.51*	5.85*	*Q*(16) = 34.67*
Sample size						
N_{PPD}	9 to 19 vs 32 to 86	0.91 vs 0.75	13 7	*Q*(19) = 41.12*	3.45	*Q*(18) = 37.67*
n_{TD}	6 to 17 vs 32 to 98	0.92 vs 0.72	14 6	*Q*(19) = 41.12*	4.31*	*Q*(18) = 36.81*
Measures						
Standardized & nonstandardized	*CTOPP, PAT-2, PIPA* vs Nonstandardized	0.78 vs 0.84	10 10	*Q*(19) = 41.12*	2.48	*Q*(18) = 38.64*
Range of mean age[b]	*4;0 to 6;11* *vs* *7;0 to 8;11*	0.754 0.751	14 4	*Q(17) =* *26.66**	.41	*Q(16) =* *26.25*
Resolution	*Unresolved* *vs* *Resolved with/without unresolved*	0.77 0.83	14 5	*Q(18) =* *41.12**	2.72	*Q(17) =* *38.39**
Severity	*Mild, moderate with/without severe* *vs Delay, consistent &/or inconsistent*	0.77 0.95	15 3	*Q(17) =* *39.59**	2.72	*Q(16) =* *39.86**

Table 2.4 Results of tests of moderators for *Phonological awareness* (continued)

Participants characteristics (continued)						
Moderator	Groups tested	*mixed weighted ES* ($\hat{d_r}$)	n	Q for HOV	Q between Groups	Q within Groups
Complexity of speech evaluation						
Word lengths	Mono- & disyllabic	0.84	7	$Q(16) = 40.97$*	3.22	$Q(15) = 37.74$
	vs					
	Multi-syllabic with/without mono- & disyllabic	0.77	10			
	Mono- & disyllabic with/without multi-syllabic	0.74	13	$Q(16) = 40.97$*	3.44	$Q(15) = 35.73$
	vs	0.95	4			
	Multi-syllabic					
Contexts	Single words	0.91	14	$Q(18) = 40.95$*	4.85*	$Q(17) = 36.10$*
	vs					
	Single words & connected speech	0.68	5			

Note. HOV = Homogeneity of variance within groups. PPD = Protracted phonological development. TD = Typical development. *CTOPP = Comprehensive Test of Phonological Processing* (Wagner et al., 1999). *PAT-2 = Phonological Awareness Test-2* (Robertson & Salter, 2007). *PIPA = Preschool & Primary Inventory of Phonological Awareness* (Dodd et al., 2003).
[a]Other = New Zealand. [b]Age at which literacy was evaluated, within closest approximations of earlier and later phases of childhood literacy (Chall,1983).
*$p = .05$

Table 2.5 Results of tests of the moderators for *Word decoding* (n = 17)

Moderator	Groups tested	*mixed weighted ES* (d_r^)	n	Test for HOV	Q between groups	Q within groups
Study characteristics						
Cultural context	US	0.65	8	$Q(16)$ = 50.01*	0.81	$Q(15)$ = 49.20*
	vs	vs				
	England & other[a]	0.67	9			
Proportion male participants						
PPD	Near 50% to 60%	0.66	6	$Q(13)$ = 39.18*	0.74	$Q(12)$ = 38.44*
	vs	vs				
	61% to 70%+	0.76	8			
TD	Near 50% to 60%	0.69	10	$Q(13)$ = 39.18*	0.63*	$Q(12)$ = 38.56*
	vs	vs				
	61% to 70%+	0.83	4			
Sample size						
NPPD	4 to 19	0.45	12	*Q(16) = 50.01**	5.53*	*Q(13) = 44.48**
	vs	vs				
	32 to 86	0.78	5			
nTD	4 to 44	0.62	14	*Q(16) = 50.01**	1.42*	*Q(15) = 48.59**
	vs	vs				
	98 to 815	0.73	3			
Measures						
Standardized & nonstandardized	WIAT, WRMT	0.63	7	*Q(16) = 50.01**	1.22	*Q(15) = 48.80**
	vs	vs				
	Other standardized & nonstandardized	0.73	10			
Standardized	WIAT, WRMT	0.63	7	*Q(13) = 44.92**	0.96	*Q(12) = 43.96**
	vs	vs				
	Other standardized	0.69	7			
Participant characteristics						
Range of mean age[b]	5;0 to 7;11	0.70	10	$Q(16)$ = 50.01*	1.71	$Q(15)$ = 48.30*
	vs					
	8;0 to 10;11	0.51	7			
Resolution	Unresolved	0.73	11	$Q(16)$ = 50.01*	2.84	$Q(15)$ = 47.17*
	vs					
	Resolved with/without unresolved	0.51	6			

Table 2.5 Results of tests of the moderators for *Word decoding* (continued)

Participant Characteristics (continued)						
Moderator	Groups tested	*mixed weighted ES* ($\hat{d_r}$)	n	*Q* for HOV	*Q* between groups	*Q* within groups
Severity	Mild, moderate with/without severe	0.69	10	*Q*(14) = 27.47*	.27	*Q*(13) = 27.20*
	vs					
	Moderate &/or severe	0.62	5			
Complexity of speech evaluation						
Word lengths	Mono- & disyllabic	0.59	8	*Q*(13) = 30.24*	4.14*	*Q*(12) = 26.10*
	vs					
	Multi-syllabic with/without monosyllabic	0.73	6			
Contexts	Single words	0.68	10	*Q*(13) = 30.24*	3.31*	*Q*(12) = 26.93*
	vs					
	Single words & connected speech	0.63	4			

Note. HOV = Homogeneity of variance within groups. PPD = Protracted phonological development. TD = Typical development. *WIAT = Wechsler Individual Achievement Tests* (Wechsler, 1992a). *WRMT = Woodcock Reading Mastery Test* (Woodcock, 1973, 1987, 1998). *TWS = Test of Word Spelling* (Larsen et al., 1999).
[a] Other = New Zealand. [b]Age at which literacy was evaluated, within closest approximations of earlier and later phases of childhood literacy (Chall,1983).
*p = .05

Table 2.6 Results of tests of the moderators for *Nonword decoding* (n = 9)

Moderator	Groups tested	*mixed weighted ES* ($d_r^{\wedge}$)	n	Test for HOV	*Q* between groups	*Q* within groups
		Study characteristics				
Cultural context	US	0.61	5	*Q*(7) =	3.38	*Q*(6) =
	vs	vs		34.71*		31.33*
	England & other[a]	0.28	3			
Proportion male participants						
PPD	Near 50% to 60%	0.75	3	*Q*(6) =	2.77	*Q*(5) =
	vs	vs		27.93*		25.16*
	61% to 70%+	0.38	4			
Sample size						
N_{PPD}	9 to 19	0.13	6	*Q*(8) =	14.43*	*Q*(7) =
	vs	vs		35.20*		20.77*
	33 to 44	0.89	3			
n_{TD}	6 to 24	0.23	5	*Q*(8) =	6.48*	*Q*(7) =
	vs	vs		35.20*		28.73*
	30 to 98	0.75	4			
		Participant characteristics				
Range of mean age[b]	4;0 to 7;11	0.88	5	*Q*(7) =	14.21*	*Q*(6) =
	vs			28.88*		14.67*
	8;0 to 10;11	-0.03	3			
Resolution	Unresolved	0.88	5	*Q*(8) =	20.35*	*Q*(7) =
	vs			35.20*		14.85*
	Resolved	-0.08	4			
Severity	Mild, Moderate with/without severe	0.78	3	*Q*(8) = 35.20*	15.00*	*Q*(7) = 20.20*
	vs	-0.16	6			
	Moderate &/or severe					
Complexity of speech evaluation						
Word lengths	Mono- & disyllabic	0.63	5	*Q(7) =*	.07	*Q(6) =*
	vs			*26.90**		*26.83**
	Mono-, di- & multi-syllabic	0.68	3			

Note. HOV = Homogeneity of variance within groups. PPD = Protracted phonological development. TD = Typical phonological development.
[a]Other = New Zealand. [b]Age at which literacy was evaluated, within closest approximations of earlier and later phases of childhood literacy (Chall,1983).
*p = .05

Table 2.6 Results of tests of the moderators for *Nonword decoding* (continued)

Moderator	Groups tested	*mixed weighted ES* (d_r^)	n	*Q* for HOV	*Q* between groups	*Q* within groups
Participant characteristics (continued)						
Complexity of speech evaluation						
Word lengths	Mono- & disyllabic vs	0.63	5	*Q*(7) = 26.90*	.07	*Q*(6) = 26.83*
	Mono-, di- & multi-syllabic	0.68	3			

Note. HOV = Homogeneity of variance within groups. PPD = Protracted phonological development. TD = Typical phonological development.
[a]Other = New Zealand. [b]Age at which literacy was evaluated, within closest approximations of earlier and later phases of childhood literacy (Chall,1983).
**p* = .05

Table 2.7 Results of tests of the moderators for *Spelling* (n = 14)

Moderator	Groups tested	*mixed weighted ES* ($d_r^{\wedge}$)	n	Test for HOV	*Q* between groups	*Q* within groups
Study characteristics						
Cultural context	US	0.90	7	*Q*(12) = 41.66*	6.98*	*Q*(12) = 34.71*
	vs	vs				
	England & other[a]	0.77	7			
Proportion male participants						
PPD	Near 50% to 60%	0.99	3	*Q*(10) = 18.97*	2.16	*Q*(9) = 16.81
	vs	vs				
	61% to 70%+	0.85	8			
TD	Near 50% to 60%	0.90	6	*Q*(11) = 22.92*	2.07	*Q*(10) = 20.86*
	vs	vs				
	61% to 70%+	0.82	6			
Sample size						
NPPD	4 to 23	0.68	10	*Q(13) = 41.69**	9.50*	*Q(12) = 32.19**
	vs	vs				
	39 to 86	0.97	4			
nTD	4 to 24	0.68	10	*Q(13) = 41.69**	9.50*	*Q(12) = 32.19**
	vs	vs				
	35 to 219	0.97	4			
Measures						
Standardized & nonstandardized	WIAT, TWS & otherb	0.81	10	*Q(13) = 41.69**	7.47*	*Q(12) = 34.22**
	vs	vs				
	Nonstandardized	0.99	4			
Participant characteristics						
Range of mean age[c]	5;0 to 7;11	0.94	5	*Q*(12) = 41.97*	8.00*	*Q*(11) = 34.00*
	vs					
	8;0 to10;11	0.77	8			
Resolution	Unresolved	0.94	7	*Q*(12) = 41.91*	7.65*	*Q*(11) = 34.26*
	vs					
	Resolved	0.77	6			
Severity	Mild, moderate	0.86	3	*Q(9) = 25.16**	5.31*	*Q(8) = 19.84**
	with/without severe	0.84	7			
	vs					
	Moderate &/or severe					

Table 2.7 Results of tests of the moderators for *Spelling* (continued)

Participants characteristics (continued)						
Complexity of speech evaluation						
Word lengths	Monosyllabic vs	0.79	7	*Q*(11) = 22.66*	5.66*	*Q*(10) = 17.00
	Multi-syllabic with/without mono- & disyllabic	0.87	5			

Note. HOV = Homogeneity of variance within groups. PPD = Protracted phonological development. TD = Typical phonological development. *WIAT = Wechsler Individual Achievement Tests* (Wechsler, 1992). *TWS = Test of Word Spelling* (Larsen et al., 1999).
[a]Other = New Zealand. [b]Other = *Graded Word Spelling Test* (Vernon, 1997), *Schonell Graded Spelling Test* (Schonell & Schonell, 1956) & *Metropolitan Achievement Test* (Durost et al., 1959). [c]Age at which literacy was evaluated, within closest approximations of earlier and later phases of childhood literacy (Chall,1983).
**p* = .05

Figure 2.1 Funnel plot of *unbiased d* for *Phonological awareness* composites (n = 20)

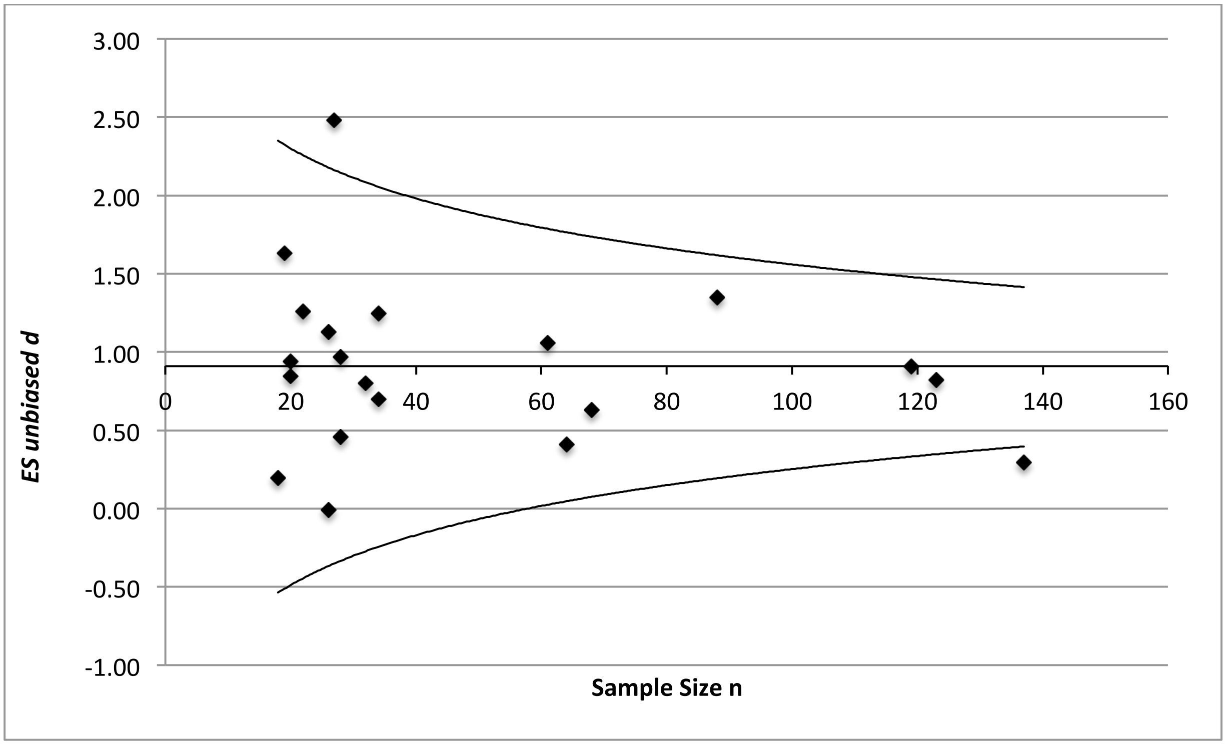


Figure 2.2 Funnel plot of *unbiased d* for *Word decoding* (n = 17)

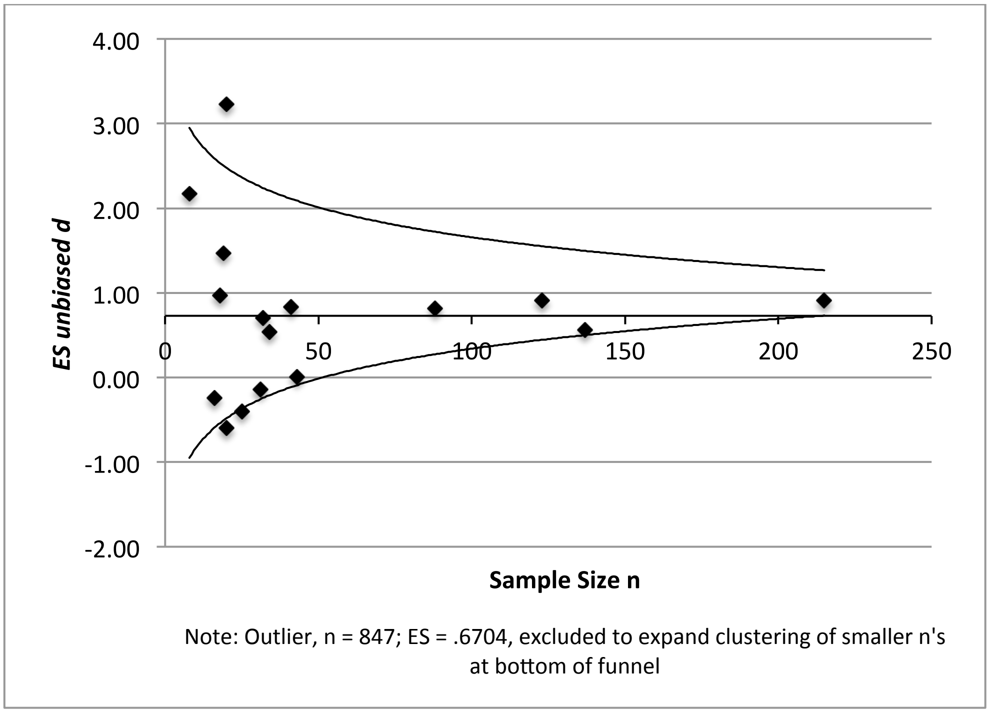


Figure 2.3 Funnel plot of *unbiased d* for *Nonword decoding* (n = 9)

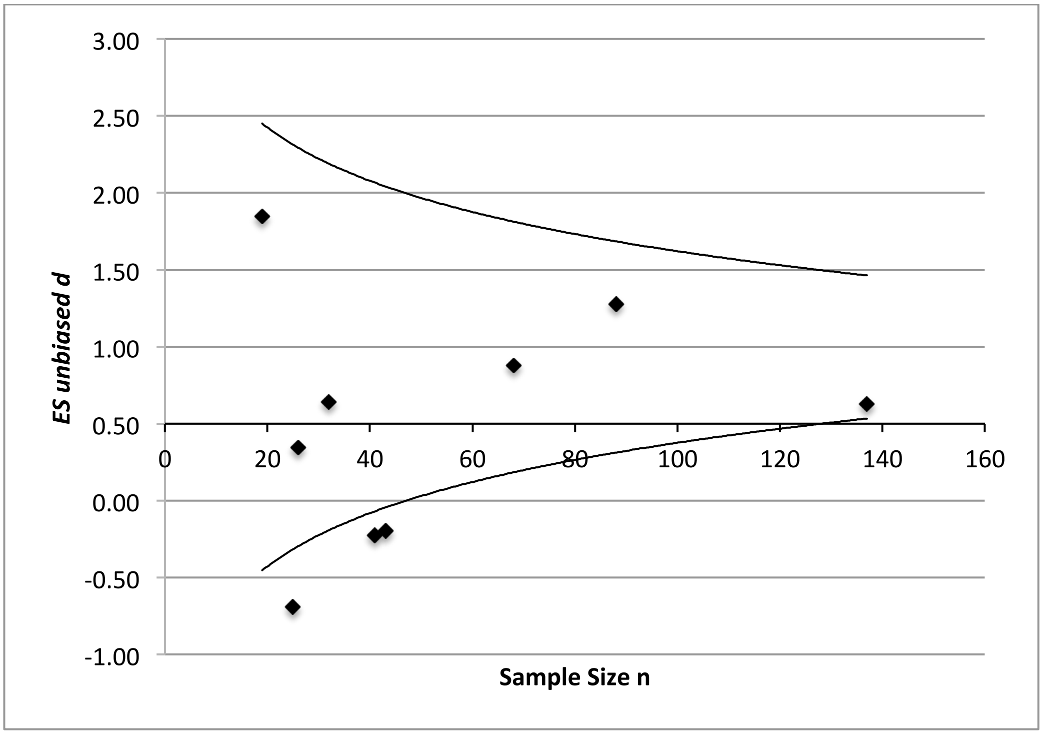


Figure 2.4 Funnel plot of *unbiased d* for *Spelling* (n =14)

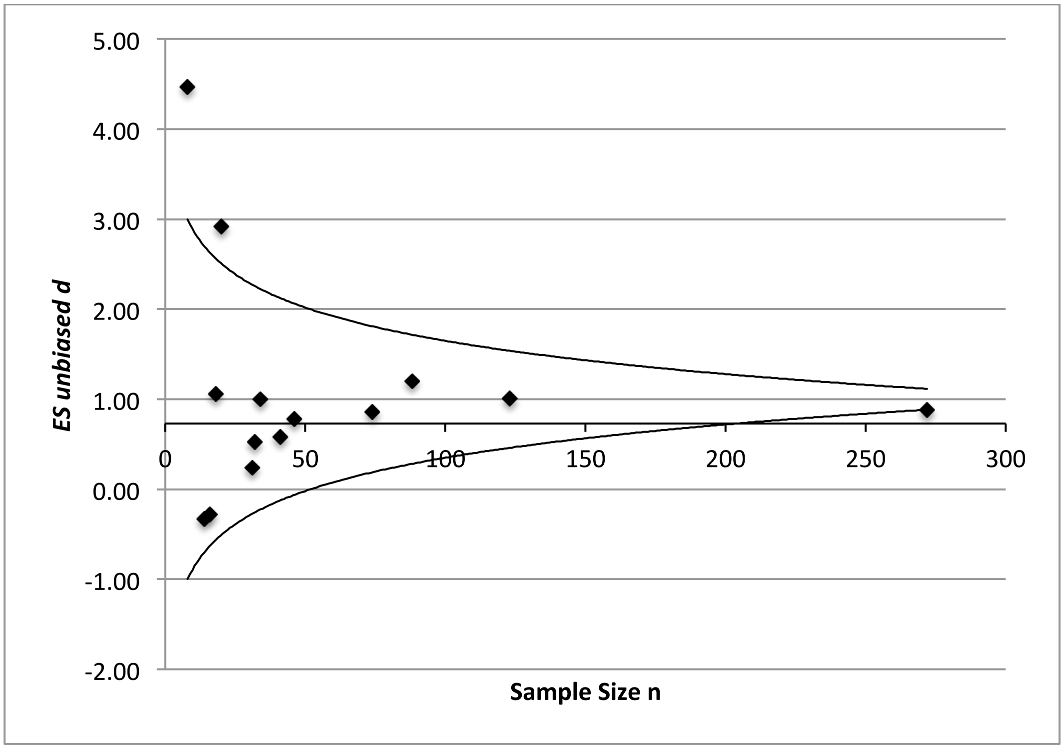
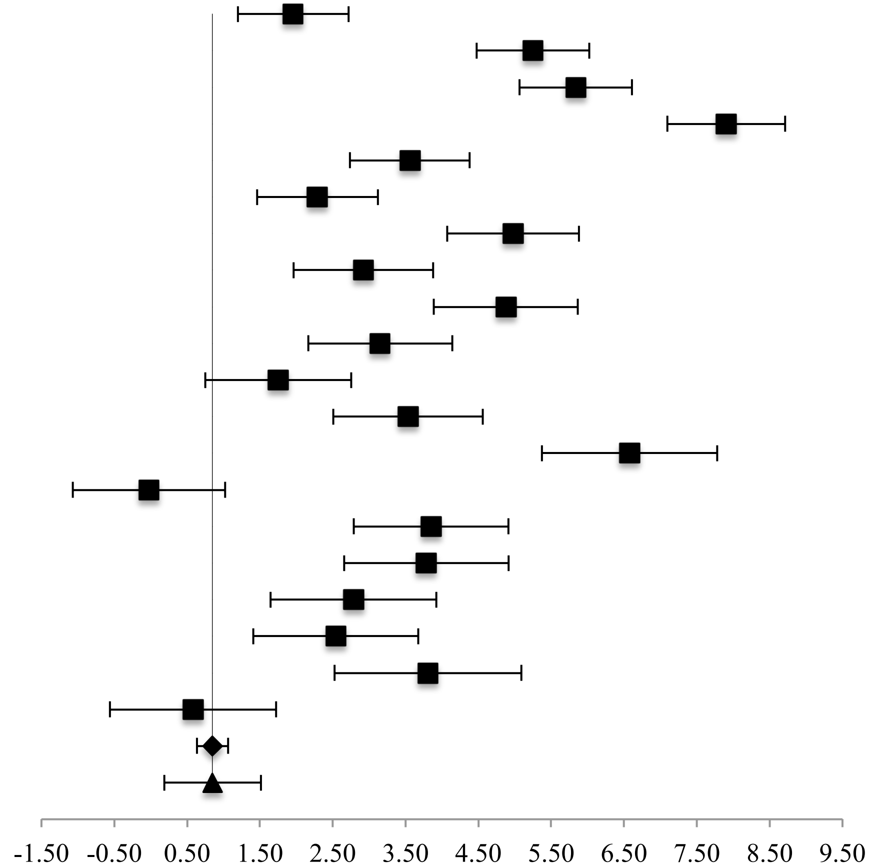


Figure 2.5 Forest plot of *Effect size* d_r^ for *Phonological awareness*

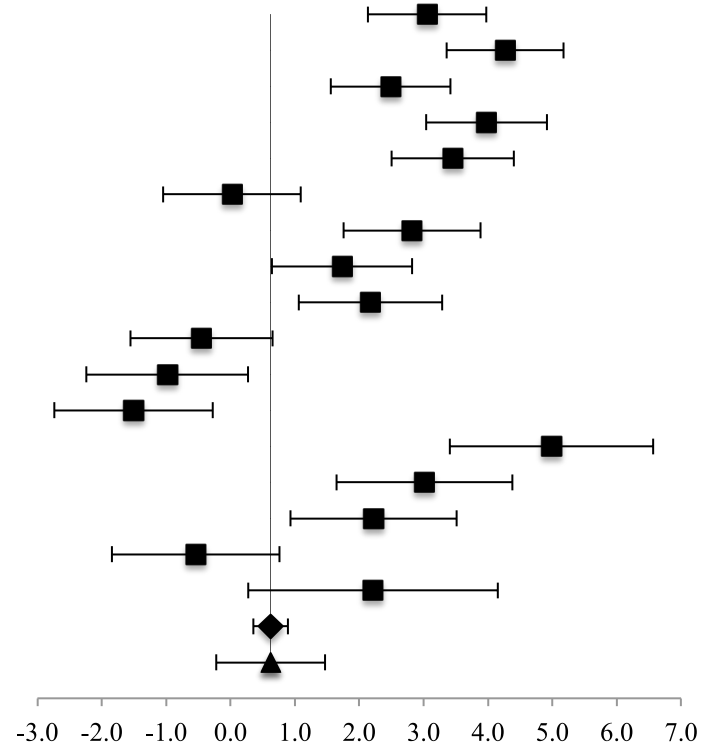
	n	d_r^	(95% CI)
McDowell (2004)	137	1.96	(1.19 - 2.72)
Peterson et al., (2009)	123	5.25	(4.47 - 6.02)
Raitano et al. (2004)	119	5.84	(5.06 - 6.61)
Apel & Lawrence (2011)	88	7.90	(7.09 - 8.71)
Rvachew (2007)	68	3.56	(2.74 - 4.38)
Chabon (1980)	64	2.29	(1.46 - 3.12)
Holm et al. (2008[1])	61	4.98	(4.07 - 5.88)
Carroll & Snowling (2004)	34	2.92	(1.96 - 3.88)
Lewis & Freebairn (1992[1])	34	4.88	(3.89 - 5.87)
Meredith (2002)	32	3.15	(2.16 - 4.14)
Webster et al. (1997)	28	1.75	(0.75 - 2.76)
Cowan & Moran (1997)	28	3.53	(2.51 - 4.56)
Preston & Edwards (2007)	27	6.58	(5.37 - 7.78)
Sutherland & Gillon (2007)	26	-0.02	(-1.07 - 1.02)
Rvachew et al. (2003)	26	3.85	(2.79 - 4.91)
Webster & Plante (1992)	22	3.78	(2.65 - 4.92)
Core (2004)	20	2.79	(1.65 - 3.93)
Ekelman (1993)	20	2.54	(1.41 - 3.67)
Harris et al. (2011)	19	3.81	(2.52 - 5.09)
Holm et al. (2008[2])	18	0.58	(-0.56 - 1.72)
Mean		0.85	(0.64 - 1.06)
Credibility Interval			(.19 - 1.51)



Note: Random effects model. d_r^ = *unbiased d* weighted by the *inverse variance* and the *Random Effects Variance Component (REVC)*.

Figure 2.6 Forest plot of *Effect size* d_r^ for *Word decoding*

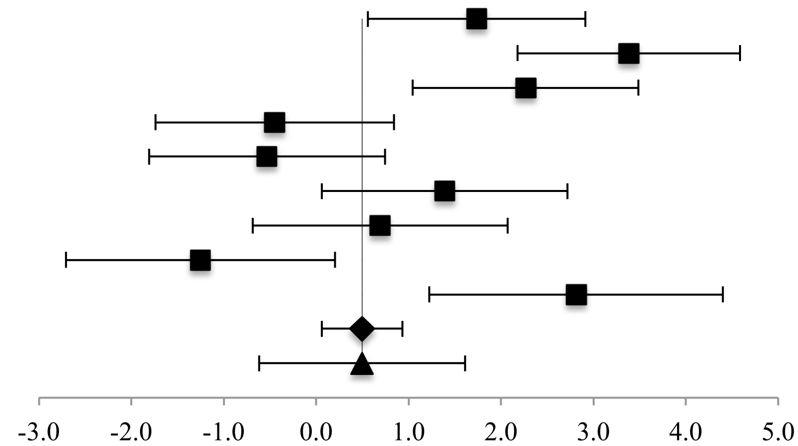
	n	d_r^	(95% CI)
Silva et al. (1984)	847	3.06	(2.14 - 3.97)
Overby et al. (2012)	215	4.27	(3.36 - 5.17)
McDowell (2004)	137	2.49	(1.56 - 3.42)
Peterson et al. (2009)	123	3.98	(3.04 - 4.92)
Apel & Lawrence (2011)	88	3.45	(2.50 - 4.40)
Catts (1993)	43	0.02	(-1.04 - 1.09)
Kirk & Gillon (2003)	41	2.82	(1.75 - 3.88)
Carroll & Snowling (2004)	34	1.74	(0.65 - 2.82)
Meredith (2002)	32	2.18	(1.06 - 3.29)
Bird et al. (1995)	31	-0.45	(-1.56 - 0.66)
Wellman et al. (2011)	25	-0.98	(-2.24 - 0.27)
Ekelman (1993)	20	-1.51	(-2.74 - -0.28)
Stackhouse (1982)	20	4.99	(3.41 - 6.57)
Harris et al. (2011)	19	3.01	(1.64 - 4.38)
Holm et al. (2008[2])	18	2.22	(0.93 - 3.51)
Bishop & Clarkson (2003)	16	-0.54	(-1.85 - 0.76)
Snowling & Stackhouse (1983)	8	2.21	(0.27 - 4.15)
Mean		0.63	(0.36 - 0.90)
Credibility Interval			(-.22 - 1.47)



Note: Random effects model. d_r^ = *unbiased d* weighted by the *inverse variance* and the *Random Effects Variance Component (REVC)*.

Figure 2.7 Forest plot of *Effect size* d_r^ for *Nonword decoding*

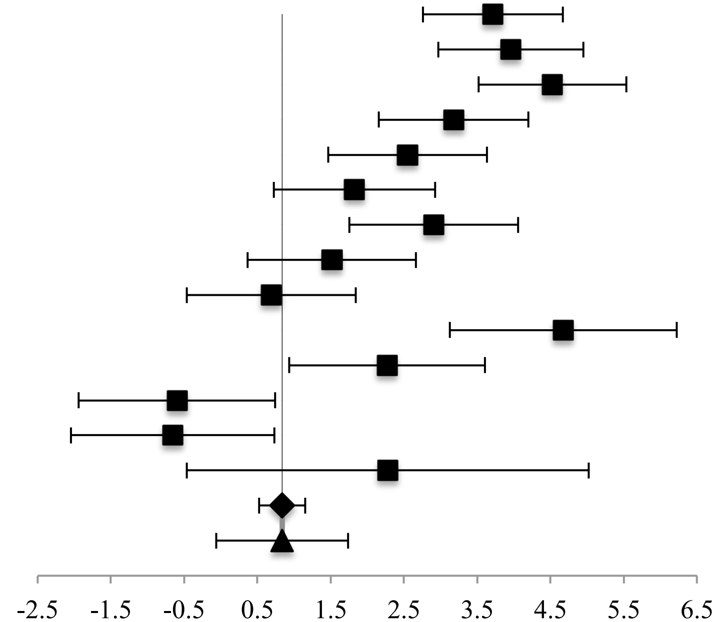
	n	d_r^	(95% CI)
McDowell (2004)	137	1.74	(0.56 - 2.92)
Apel & Lawrence (2011)	88	3.38	(2.18 - 4.57)
Rvachew (2007)	68	2.27	(1.05 - 3.5)
Catts (1993)	43	-0.45	(-1.74 - 0.84)
Kirk & Gillon (2003)	41	-0.53	(-1.81 - 0.75)
Meredith (2002)	32	1.39	(0.06 - 2.72)
Sutherland & Gillon (2007)	26	0.69	(-0.70 - 2.07)
Wellman et al. (2011)	25	-1.25	(-2.71 - 0.21)
Harris et al. (2011)	19	2.81	(1.23 - 4.40)
Mean		0.50	(0.06 - 0.93)
Credibility Interval			(-.61 - .61)



Note: Random effects model. d_r^ = *unbiased d* weighted by the *inverse variance* and the *Random Effects Variance Component (REVC)*.

Figure 2.8 Forest plot of *Effect size* $\hat{d_r}$ for *Spelling*

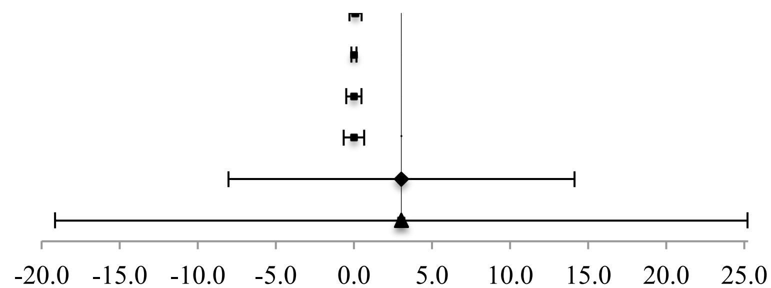
	n	d_r^	(95% CI)
Overby et al. (2012)	272	3.71	(2.76 - 4.67)
Peterson et al. (2009)	123	3.96	(2.97 - 4.95)
Apel & Lawrence (2011)	88	4.52	(3.52 - 5.53)
Nathan et al. (2004)	74	3.18	(2.16 - 4.20)
Lewis & Freebairn (1992[1])	46	2.55	(1.47 - 3.63)
Kirk & Gillon (2003)	41	1.82	(0.72 - 2.93)
Lewis & Freebairn (1992[2])	34	2.91	(1.76 - 4.06)
Meredith (2002)	32	1.51	(0.36 - 2.67)
Bird et al. (1995)	31	0.69	(-0.46 - 1.84)
Stackhouse (1982)	20	4.67	(3.12 - 6.22)
Holm et al. (2008[2])	18	2.27	(0.93 - 3.60)
Bishop & Clarkson (2003)	16	-0.60	(-1.94 - 0.74)
Burrus (2007)	14	-0.66	(-2.04 - 0.73)
Snowling & Stackhouse (1983)	8	2.28	(-0.46 - 5.02)
Mean		0.84	(0.53 - 1.15)
Credibility Interval			(-.06 - 1.74)



Note: Random effects model. d_r^ = *unbiased d* weighted by the *inverse variance* and the *Random Effects Variance Component (REVC)*.

Figure 2.9 Forest plot for *Effect size* d_r^ for *Fluency*

	n	d_r^	(95% CI)
Puranik et al. (2008)	10221	.09	(.31 - .50)
Peterson et al. (2009)	123	.006	(-.16 - .18)
Rvachew (2007)	68	.002	-.48 - .48)
Catts (1991, 1993)	43	-.005	(-.67 - .66)
Mean		3.04	(-8.05 - 14.13)
Credibility Interval			(-19.13 - 25.21)



Note: Random effects model. d_r^ = *unbiased d* weighted by the *inverse variance* and the *Random Effects Variance Component (REVC)*.

Chapter 3: Evaluation of Multisyllabic Word Production in Canadian English- or French-speaking Children within a Nonlinear Framework

3.1 Synopsis of Study

In the previous chapter, the meta-analysis suggested the relationship of multisyllabic words (MSWs) to literacy, and to the evaluation of PPD. For word level literacy, differences between children with and without PPD were moderate to large, in particular for phonological awareness, word decoding and spelling. The differences held for literacy evaluation pre- or post-early literacy, and for PPD characterized as less severity or resolved by various methodologies. Notably, the latter included assessment in complex speech contexts, i.e. connected speech and/or MSWs. Currently, there is no theoretically justified, evidence-based metric for evaluating segmental and prosodic components of multisyllabic words. The current chapter, therefore, reports a pilot study that evaluated a MSW metric embedded in nonlinear phonological and language processing frameworks. The aim was to describe application of the metric to more than one language spoken in Canada, English and French. The English data are central to the thesis but the French data were retained because the integrity of interpretation depends on the study as a whole. Six representative MSWs were analyzed for ten English-speaking 5-year-olds with typically developing speech, and eight French-speaking children, ages 3 to 4 years, with protracted phonological development (PPD). Mismatches were tallied (with and without vowels), with totals ranked by word and participant, then compared with ranks from Phonological Mean Length of Utterance (PMLU; Ingram, 2002) and Percent Consonants Correct (PCC; Shriberg, Austin, Lewis, McSweeny, & Wilson, 1997) tallies. For both groups, the number of different ranks was significant in comparisons of MSW metrics with PMLU and PCC. Rank orderings were systematically higher for English-speaking children using the MSW metric, with/without vowels, and for French-speaking children using the MSW metric with vowels. Overall, the MSW metric was particularly suitable for fine-grained differentiation of phonological accuracy in MSW production.

3.2 Introduction

Recent studies suggest that multisyllabic words (MSWs) are important contexts for evaluation of a child's speech production and literacy capacity, particularly beyond age 5 years (e.g. James, 2006; Mason & Bernhardt, 2014). Although researchers have indicated the need to evaluate a wide range of segmental and prosodic components in developmental phonological analysis (Arias & Lleo, 2013; Ingram, 2002; Stoel-Gammon, 2010) a theoretically justified and evidence-based metric for MSW analysis is currently unavailable. The current exploratory study thus introduces a quantitative whole word MSW metric that tallies phonological mismatches, examining its reliability, convergent validity with two existing measures, Percent Consonants Correct (PCC: Shriberg, Austin, Lewis, McSweeny, & Wilson, 1997) and Phonological Mean Length of Utterance (PMLU; Ingram, 2002), and sensitivity to variation in accuracy of phonological output. Face validity of the metric is grounded in theories of nonlinear phonology (e.g. Bernhardt & Stemberger, 1998) and language processing (e.g. Presson & MacWhinney, 2010; Stemberger, 1992; Wheeler & Touretzky, 1997). Utility of the metric was examined in two languages with contrasting dominant stress patterns: English (word-initial stress) and French (word/phrase-final stress). The following sections discuss first theoretical and developmental support for evaluating MSWs, and then current analysis methods, as a background for the research questions for the study.

3.2.1 Theoretical Support for Multisyllabic Word Evaluation

Both linguistic theories and language processing models are relevant for evaluation of MSW production. Current linguistic theories include nonlinear phonological frameworks that delineate the hierarchical structure of words (e.g. features and segments, syllables, stress, feet), and constraint-based theories, Optimality Theory (OT), which portray phonological output as a result of competition between markedness (complexity/cue cost) and faithfulness to the underlying form (cue validity). The two linguistic theories posit that phonological acquisition is a gradual re-ranking of competing output constraints at various levels of the phonological hierarchy (Stoel-Gammon & Bernhardt, 2013). Current language processing models can provide accounts for both the multi-tiered construction of an intended word, and the interaction effects of output constraints, i.e. lexical and phonological cue cost and validity (Presson & MacWhinney, 2010), and connectionism (Bernhardt, Stemberger, & Charest, 2010). These processing models posit that parallel, weighted and interactive activation cascades and spreads over word-relevant information in morphosyntactic, semantic and phonological cognitive processing networks. Inhibition processes filter redundant phonological information and compete with excitatory processes that facilitate interactions across large cortical networks (Presson & MacWhinney, 2010).

Child phonological representations may be constrained by insufficient linguistic content or specification. Such constraints are reflected in non-robust chunking of speech perception experiences, i.e. inadequate strength of connections among levels of representation within a certain network radius (Pierrehumbert, 2006), and slowed rate and distribution of network growth (Elman, 1999). Trade-offs may occur between reduced availability and reliability of phonological input and the cognitive load associated with using the cue to achieve adult-like output (e.g. cue cost or amount and type of processing). Such trade-offs may result in insufficient activation of more marked phonological units, and/or inaccurate activation and inhibition rates among competing units. This in turn may result in disorganization or incorrect pacing of network assembly processes needed to produce adult-like output (Presson & MacWhinney, 2010).

The output of MSWs, with their complex form, requires multiple interactions of semantic, morphosyntactic, and phonological processes. Because of buffer (short-term memory capacities) and licensing (allowable and correct assignment of phonological components) constraints, there is potential for varied phonological mismatches across repeated productions of the same word. If there is insufficient inhibition of irrelevant semantic information, for example, lexically based mismatches may occur. For example, during production of *hippopotamus* /ˌhɪpəˈpʰɑːɾəməs/, other short words might be activated at the same time, e.g. *hip, hit.* If *hit* is not inhibited, a lexically based mismatch may result, e.g. [ˌhɪt.təˈpʰɑːtəməs]. Alternatively, prosodic effects might occur. For instance, words with marked stress patterns (e.g. weak-Strong-weak patterns in English) may show weak syllable deletion with consequent inhibition/deletion of segments (consonants, vowels) that do occur in words with unmarked stress patterns: e.g. *computer* /kəmˈpʰjuːɾɚ/ as [ˈpʰjuːɾɚ] versus *come IN* /ˌkʰəmˈĩn/ as [ˌkʰəmˈĩn]. In another case, segments might be produced, but not in the target location; i.e. they are not linked sequentially to their respective timing units (production time allotted to a segment; Bernhardt & Stemberger, 1998), e.g. *explodes* /ɪkˈsploʊdz/ as [kəˈsploʊdz]. Thus, interactions of multiple levels of form can result in various types of mismatches for MSWs, which may or may not occur in monosyllabic words, i.e. lexically based mismatches or phonological deletions, insertions, and in particular, assimilations, reduplications, and transpositions. The following section describes acquisition of MSWs.

3.2.2 Acquisition of Multisyllabic Words

English and French have a rich variety of MSWs, although, as noted above, differences in their prosodic structure may affect developmental patterns. English, a stress-timed language, has a preponderance of words with initial stress, although two-footed words often have stress towards the end of the word (Shockey, 2003); French, a syllable-timed language, has almost exclusively phrase-final stress (Fikkert, Freitas, Grijzenhout, Levelt, & Wauquier, 2004). Developmental studies concerning MSWs are discussed for the two languages in turn below.

For English, the only in-depth study of MSWs to date is that by James (2006). She examined monomorphemic MSWs in 264 Australian English-speaking children aged 4 to 7 years, observing two principal developmental stages: between ages 4;0 and 6;11, children mastered word length and phoneme sequences, but showed stress pattern mismatches; after age 7;0, a variety of stress patterns was acquired. Segmental accuracy also changed over the developmental periods. Although the percentage of vowels correct (PVC) exceeded 90% across ages and word types, vowels were mastered first in disyllabic words (DSWs) at age 4 years, and then in MSWs at age 7 years. PCC did not exceed 90% for either DSWs or MSWs until age 7, although was significantly higher in stress-initial DSWs at both ages 4 and 7 years.

James (2006) also described frequencies and types of phonological processes. The median phonological process frequency was generally below 2% for DSWs at age 4, and for

MSWs, at age 7. A few processes were as frequent as 3% to 6%: (a) *consonant deletion* in DSWs and MSWs (both age groups); (b) *syllable deletion* and *assimilation* in MSWs (4-year-olds); and (c) *schwa tensing/lengthening* in MSWs (both age groups). *Syllable* or *schwa insertion* was rare. *Consonant insertion, metathesis, and transposition* were rare in DSWs, but present in MSWs for about half of the children. Most processes decreased in frequency from ages 4 to 5. The exception was *schwa lengthening*, which increased up to age 7; a small sacrifice was perhaps made in terms of syllable timing in order to preserve the segments of the word. Process occurrence was most often related to presence of initial weak syllables or complexity of consonant sequences. If initial weak syllables were deleted, onsets in word-initial (WI) position were sometimes transposed elsewhere in the word. In consonant sequences with shared place or manner, e.g. nasal-nasal sequences, *animals* /ˈæːnəmɫ̩z/, or transitions between place, e.g. dorsal-coronal-labial-coronal, *octopus* /ˈɑːktəˌpʊs/, mismatches increased. Overall, the vulnerability of a word was dependent on its prosodic, segmental and sequential characteristics, explainable in terms of cascading activations during licensing (Wheeler & Touretzky, 1997), and cue validity/faithfulness (Presson & MacWhinney, 2010; Bernhardt & Stemberger, 1998, respectively).

The James (2006) study, although a landmark study for acquisition of English MSWs, has reduced generalizability for other languages, and even for other dialects of English, because of dialect and lexical characteristics of the elicitation word list, including: (a) Australian lexical items, with variant pronunciations, i.e. lack of syllabic and post-vocalic /ɹ/, and presence of schwa in certain words, e.g. *zucchini* /zuˈkini/ as [zəˈkini]; (b) frequent multimorphemic words;

and (c) restricted numbers of DSWs with iambic (weak-Strong) stress. Further investigation of English MSW development is clearly warranted.

As for English, relatively few studies describe MSW development for French. In early development, children tend to produce words limited to one binary foot; thus, three- and four-syllable words are generally truncated to DSWs with word-final (WF) stress maintained, and early DSWs are often reduplications of the WF syllable (Rose & dos Santos, 2008; Rvachew, Marquis, Brosseau-Lapré, Paul, Royle, & Gonnerman, 2013; Wauquier & Yamaguchi, 2013). Wauquier and Yamaguchi attributed the consistent, non-random preservation of WF syllables to French prosodic influences. Relative to segments, the Rose and dos Santos (2008) study of European French-speaking 1- and 2-year-olds showed that consonants with different manner features might appear in monosyllabic words, but not in MSWs, due to deletion of the first syllable. For Quebec French-speaking 1- to 4-year-olds, Rvachew et al. (2013) also observed higher frequency of consonant substitutions or deletions in unstressed (non-final) syllables. In older children (ages 5 to 7 years), Rvachew et al. noted consonants were least accurate in unstressed syllables of MSWs. Taken together, the findings suggest that segments are more likely to show substitutions and deletions in MSWs, because of more components that are marked/ have higher cue cost (Bernhardt & Stemberger, 1998; Presson & MacWhinney, 2010). There is therefore justification for further exploration of the development of French MSWs.

3.2.3 Evaluation of Multisyllabic Words

The preceding discussion indicates the relevance of including a representative sample of MSWs in phonological assessment protocols for school-age children. However, analysis metrics to date have not specifically focused on MSWs. Advantages and disadvantages of existing metrics for MSW analysis are outlined below as motivation for the proposed metric.

Phonological evaluation has traditionally focused on segmental evaluation, e.g. with traditional articulation tests, or PCC/PVC measures of connected speech (Shriberg et al., 1997). Such methodologies often tally binary judgments of segmental correctness, and therefore, do not capture multiple differences among mismatched productions. To increase sensitivity to this variability, researchers have designed extended metrics to account for: (a) the relative impact of various phonological processes on intelligibility (Preston, Ramsdell, Oller, Edwards, & Tobin, 2011); (b) differences described by specific phonological features (Hall, Adams, Hesketh, & Nightingale, 1998); and (c) the increasing phonological length and complexity of words accompanied by vocabulary growth (Ingram, 2002). The trend in these evaluation metrics, therefore, has been to examine a number of phonological components in addition to consonants or vowels, i.e. syllables, and/or consonant/vowel timing units. Examples of such metrics include modified PCC/PVC (James, 2006); and Ingram's (2002) Phonological Mean Length of Utterance (PMLU) and Proportion of Whole Word Proximity (PWP), and their various adaptations (Arias & Lleo, 2013; MacLeod, Laukys, & Rvachew 2011; Newbold, Stackhouse, & Wells, 2013). Stoel-Gammon's (2010) Word Complexity Measure (WCM) for children with small vocabularies also evaluates stress pattern and syllable number but disregards earlier-established word structures and segments. For MSWs, with their great variety of lengths, stress patterns, complexity of syllables, and feature sequences, the multiple levels and potential lexical interference for various sub-components of the words are paramount. However, the current metrics do not evaluate potential lexical processing effects, nor all levels of the phonological hierarchy (Ingram, 2002), and may be difficult to replicate for research or clinical purposes, because of cumbersome and often arbitrary component weighting, such as in Hall et al. (1998) or Preston et al. (2011).

The various current measures have also differed in the focus on accuracy or inaccuracy and the relative quantification of various word components. For example, PCC, PMLU, PWP, and WCM evaluate accuracy. Refined on monosyllables and disyllables, these measures may be less applicable for analysis of longer words. Few researchers have exclusively quantified inaccuracy (e.g. Hall et al., 1998). More often, mismatch scores have been combined with match scores, i.e. deduction of mismatches after match proportions have been calculated, making results more difficult to interpret, for example: Weighted Speech Sound Accuracy (WSSA) (Preston et al., 2011), modified PCC (James, 2006), and rule-added PMLU (MacLeod et al., 2011). Some researchers have attempted to capture assumed differences in relevance of various components through assigning different weights (e.g. Hall et al., 1998; Preston et al., 2011; Stoel-Gammon, 2010). However, components that are possibly difficult for children to produce, such as schwa vowels in unstressed syllables (James, 2006), or affricates (Stoel-Gammon, 2010), have been accorded less weight. Such assignations, while acknowledging possible relative differences in mismatch types, remain necessarily arbitrary, without external validation, such as impact on intelligibility. Furthermore, evaluation is less comprehensive when certain components are ignored, e.g. epenthesised segments (MacLeod et al*.,* 2011), or specific sound classes (Stoel-Gammon, 2010), because the interaction of segmental accuracy with word structure is lost. For example, various productions of a long word such as *hippopotamus* /hɪpəˈpʰɑːɾəməs/, with a number of early acquired consonants, i.e. /h, p, m/, might be equal using PMLU tallies and PCC (see Table 3.3), whereas taking into account more variables (e.g. foot, prosodic word, lexical effects, transposition of segments) might generate further gradation across different productions (Ingram, 2002).

In summary, while addressing certain aspects of the phonological system, the available metrics do not address the complexities and interactions between lexical selection, segmental feature analyses and all prosodic levels of the phonological hierarchy (e.g. word length, complexity of syllables, and feature sequences). Many also focus on accuracy only or use weighted measures. Thus, an alternative measure for MSW analysis appears motivated, as discussed below.

3.2.4 The Current Study

The current study piloted a new analysis measure of MSWs in school-aged children in both English and French. The metric builds on previous methodologies for measuring whole word production, such as PCC (Shriberg, et al., 1994) and PMLU (Ingram, 2002), and early versions of the MSW metric for English (Mason, Bernhardt, & Masterson, 2010) and Spanish (Schretlen, 2013). The latter two studies utilized a more elaborate tally metric for word structure than for PMLU and PCC (i.e. stress, length by syllables, and consonant and vowel timing units), but did not include tallies for potential lexical effects or feature mismatches. The current metric was designed to be comprehensive, incorporating lexical selection and segmental feature analyses in addition to all prosodic levels of the phonological hierarchy. The proposed metric thus evaluates the following:

1. Lexical selection effects: components of phonological word neighbors and other phonologically similar words or syllables that surface in output as a result of interactive processing of lexical and phonological representations (Bernhardt & Stemberger, 1998).

2. Prosodic word levels of the nonlinear phonological hierarchy: foot (syllables grouped by patterns of a syllable carrying prominent stress and one or more nonprominent syllables; Bernhardt & Stemberger, 1998); syllable number; consonant and vowel timing units; and consonant and vowel features.

3. Phonological component interactions with respect to sequence effects (to which parallel licensing applies; Shattuck-Hufnagel, 1992; Wheeler & Touretzky, 1997), i.e. transposition, metathesis, assimilation, and in addition, deletion and insertion.

While acknowledging that accuracy measures provide important general information about phonological competence, the mismatch tallies were selected as the measure for the metric. Information about mismatches can help explain variability with regard to specific hierarchical interactions and competing constraints of the phonological system, and in terms of clinical application, provide specific targets for remediation. Simple mismatch tallies were chosen over weighted measures, in order to avoid arbitrary decisions about relative importance of elements, and to increase the reliability of judgments made across more word components.

In order to evaluate the new metric, three comparisons were made with two major previous measures (PCC, PMLU): (a) a MSW metric including vowel feature tallies, versus one without such tallies; (b) a MSW metric without vowel tallies, versus PCC; and (c) a MSW metric with vowel tallies, versus tallies for PMLU.

The following predictions were made:

1. Convergent validity:
    a. Moderate correlations between the proposed MSW metric and the PCC tally procedure were predicted because both measures evaluate phonology, but the current metric includes more levels of the phonological hierarchy;
    b. Larger but still moderate correlations between the MSW metric and the PMLU tally procedure were predicted because both methods include prosodic elements relevant to syllable structure; however, the current metric includes more prosodic levels.
2. Sensitivity: The MSW metric was predicted to produce significantly fewer ties in rank order within words, and across children, in the following comparisons, because the inclusion of more prosodic levels expands the potential range of the data:
    a. the MSW metric without vowels and tallies for deriving PCC;
    b. the MSW metric with vowels and tallies for deriving PMLU;
    c. the MSW metric with and without vowels.
3. Reliability: Reliability of the MSW metric was predicted to be high because of the relative simplicity of tallying phonological mismatches.

3.3 Method

The data for the current study were drawn from a larger study (for English, see Chapter 4). Relevant methodologies from the larger study are first discussed, and then specific methods for the current study.

3.3.1 The Larger English Study

English-speaking participants meeting study eligibility criteria were 64 TD monolingual kindergarten children (ages 4;9 to 6;5: 34 girls, 30 boys) from a random sample of parent/caregiver consents. (At the time of the study, only TD children were participating.) The sample comprised 18% of the total Kindergarten enrollment and was representative of the demographics of an urban/rural school district of south-central British Columbia, Canada. The children were typically developing, as documented by birth, medical and developmental histories in parent/caregiver questionnaires, and verified by kindergarten teachers and school-based speech-language pathologists. All participants were reported to have passed a bilateral Kindergarten hearing screening from 500Hz to 4000Hz in the school district.

Phonological samples were collected using the Computerized Articulation and Phonology Evaluation System (CAPES: Masterson & Bernhardt, 2001) Phonemic Profile (Profile) and Individualized Phonological Evaluation 4 (IPE4) single word elicitations. The Profile samples 27 monosyllables, 16 disyllables and 3 trisyllables, totaling 46 words. IPE4 samples a wide range of MSWs: 13 disyllables, 29 trisyllables, 8 four-syllable, 3 five-syllable and 2 six-syllable words, totaling 55 words. For the single word elicitations, the researchers designed cloze-style prompts, in order to reduce or eliminate the possibility of priming participants for target stress patterns and segments. If the scripted prompt did not elicit the target word, delayed imitation (the prompt or a short direct request to say the word, interjected between the examiner's model and the child's word production)/interrupted, and then immediate imitation was used.

Testing occurred over two 1-hour sessions at the child's school. The Profile and IPE4 (Masterson & Bernhardt, 2001) were administered during the first and second sessions, respectively. Because the children were also evaluated on a variety of language and literacy-related measures for the larger study, presentation order of the CAPES tasks was balanced with the comprehension and production demands of the other tasks. The samples were recorded using a Marantz PMD660 digital recorder (Kleinburg, Canada) with a built-in microphone set on automatic level.

After preliminary transcription of a portion of the data, a narrow transcription conventions document was developed following Bernhardt and Stemberger (2012). Inter-rater transcription reliability was calculated on 15% of the data, randomly selected by gender (five boys, five girls). From independent transcriptions of the first author and a graduate SLP student, point-to-point agreements were calculated for primary stress, consonants, and vowels. Stress agreement was not calculated for Profile words because they vary little for stress pattern, but for IPE4, agreement was 96%. Consonant and vowel agreements, respectively, were 90% and 92% for the Profile, and 85% and 94% for IPE4.

Twenty MSWs from the Profile and IPE4 were selected for analysis, based on phonological complexity and word familiarity. Word complexity was addressed by including: (a) variety in length and stress patterns, i.e. one to two feet comprised of two to five syllables; (b) early and later-developing segments; and (c) challenging adjacent and non-adjacent consonant sequences of manner (e.g. [±continuant], [nasal], tap and syllabic) and place (e.g. coronal-dorsal, dorsal-coronal and coronal-labial). Word familiarity took into account word frequency and Age of Acquisition (AoA) effects, variables that have been shown to affect accuracy (Gierut & Morrisette, 2012). Word frequency was the chosen indicator of familiarity in order to avoid biases in typical parental survey data for AoA. Because word frequencies may be comparable in print and spoken language (Gierut & Dale, 2007), estimates were sourced from a large sample (Lee, 2001) of objective child speech data (Morrison, Chappell & Ellis, 1997). That is, frequencies were calculated using ChildFreq (Baath, 2010), for which the basis for counts is on actual occurrence in child speech samples, as opposed to phonotactic probability and neighbourhood density. Age ranges relevant to the current study are also included in the data base.

3.3.2 The Larger French Study

The French data were from eight monolingual Manitoba French-speaking children with PPD (the larger study did not include children with typically developing phonology), including six boys and two girls (3;1 to 4;6, 3;11 ±0;5). All children were living in primarily Francophone communities with all mothers speaking to their child in French; however, children were most likely exposed to some spoken English in their communities. All children attended a French preschool program. Caregivers reported no developmental concerns other than a mild-to-severe delay in phonological development. Each participant scored within normal limits on a measure of vocabulary comprehension. Due to the children's lack of English vocabulary, their English phonological skills were not tested. Data collection and audio recording followed the same procedures as the English study.

Data from the first two participants were transcribed by the native French-speaking co-author (Bérubé) with the collaboration of the other co-authors (Bernhardt & Stemberger) trained in transcription (speakers of French as an additional language). The pronunciation of the French co-author (Bérubé) was considered the adult target, unless the child used another acceptable form for the dialect area. Because the first two participants' data were the first to be transcribed for the overall study in French, each token was then transcribed by consensus in a group setting. Reliability was not calculated, but the final token required agreement of all listeners. This process led to creation of a document to be used for the project, outlining transcription conventions. The transcriptions were entered into CAPES (Masterson & Bernhardt, 2001), the Phon program (Rose & MacWhinney, 2014) and spreadsheets for analysis.

3.3.3 Procedures for the Current Study

For the current exploratory study, ten English-speaking children, five boys and five girls, (5;4 to 6;0, 5;10 ± 2.67), and eight French-speaking children, 2 boys and 6 girls, (3;1 to 4;6) were selected. All children had notable mismatches on at least four of six words chosen (described below) on which to evaluate the proposed whole word MSW metric. Because adult-like word productions were also included in the analysis, the statistical analysis of the metric was conducted on a breadth of accuracy within words and across children, while testing ease of replication on the more mismatched productions.

3.3.3.1 Word Selection in English and French

For English, the current study examined 6 of 20 MSWs selected for the larger study from the CAPES IPE4 (Masterson & Bernhardt, 2001): *explodes* /ɪkˈsploʊdz/, *computer* /kʰəmˈpjuːɾɚ/, *hospital* /ˈhɑːspɪɾl̩/, *thermometer* /θɚˈmɑːməɾɚ/, *cash register* /ˈkʰæːʃˌɹɛdʒɪstɚ/, and *hippopotamus* /ˌhɪpəˈpʰɑːɾəməs/. A wide range of phonological characteristics was represented (see Table 3.1): one and two feet; two to five syllables; WI, WF and medial (WM) unstressed syllables, including sequences of WF weak syllables; consonant clusters across word positions, including coda-onset sequences; challenging adjacent and non-adjacent manner and place sequences; syllabic consonants; later acquired sibilants, liquids and taps; and diphthongs and lax vowels (see Table 3.1).

To determine whether the six words for the pilot study were equally familiar to 5-year-olds, their frequencies as calculated by ChildFreq (Baath, 2010) were compared with the mean frequency (e.g. Gierut & Dale, 2007) for 19 of the 20 words for the larger study (*balloons* was excluded as a high frequency outlier). Relative to research criterion cut-offs for high frequency, (e.g. 100 per million, Morissette, 1999; 291 per million, Gierut & Morissette, 2012), all six words were deemed low frequency. Furthermore, because no word was more than 1SD from the corpus mean (32.09 ±28.97), all were judged comparable with respect to familiarity.

From the larger French word list, 6 of 17 MSWs were selected, including *balançoire* /balɑ̃ˈswɑʁ/, *crocodile* /kʀɔkɔˈdzɪl/, *hôpital* /ʔɔpiˈtal/, *arc-en-ciel* /ʔaʀkɑ̃ˈsjɛl/, *dentifrice* /dɑ̃tsiˈfʀɪs/, and *hippopotame* /ʔipɔpɔˈtam/. As for English, words were chosen to represent a balance of place and manner sequences, such that a variety of consonants were sampled (see Table 3.2).

3.3.3.2 The Multisyllabic Word Mismatch Tally Method

As noted previously, the MSW metric combined mismatches in lexical selection with phonological parameters of the prosodic word hierarchy, providing a whole word value that increased relative to discrepancy from the adult target. A tally was assigned for a lexical selection mismatch within a foot, and for all mismatched components within each phonological tier: prosodic word and foot (stress), syllables (word length), consonant and vowel timing units (word shape), and segments (features). Independent judgments at each level provided a means to attribute importance to the pervasive effects on lower tiers of mismatches at higher levels, without otherwise scaling component values. For instance, for syllable deletion in a one-footed word, a mismatch tally was assigned for the stress pattern mismatch, in addition to absences of the syllable, consonant and vowel timing units, and segmental features. Tallies of mismatches were summed to obtain the whole word total, but it was also possible to compare sub-totals for lexical selection, or other prosodic levels, singly or in combination (e.g. stress, word structure and segmental features). In the next sections, the MSW procedure for assigning tallies relevant to lexical and phonological mismatches is discussed further, and its application demonstrated.

To begin, transcripts were aligned with a target row of a spreadsheet, in which columns contained relevant nonlinear phonological components, as outlined in the preceding paragraph. Mismatches in lexical selection, and stress of each foot, were tallied in separate columns, with an additional column for stress on the prosodic word tier, for words of two feet. Transcript components were aligned in absolute order of production, such that columns were added for syllable, vowel or consonant insertions, avoiding arbitrary non-analogous positioning. To establish number of nuclei, vowels produced were aligned with target vowels in the foot, syllable, vowel timing unit and feature columns. Similarly, consonants produced were aligned with target consonants (including syllabic consonants and taps) in the foot, syllable, consonant timing unit and feature columns. Mismatched segment alignment was according to the least discrepancy with features of a target. When a mismatch was tallied, it was also color coded for insertion, deletion, transposition or assimilatory phonological processes, for later analyses.

Mismatch judgments first considered each foot for lexical selection influences, i.e. where mismatched syllables were members of (usually) monosyllabic word neighbourhoods. From the larger study, a survey of the data for words of two feet, such as /ˌhɪpəˈpʰɑːɾəməs/, indicated lexical selection effects in one or both feet. Each foot was therefore analyzed independently, so that the susceptibility of particular feet to lexical selection mismatches could be examined. No foot was tallied for more than one lexical selection mismatch in order to avoid bias as a result of similarity to words of more than one syllable, e.g. *magician* with *musician*, *electric* with *electricity,* or *(hippo)potam(us)* with *bottom*. When a lexical mismatch was tallied, no further phonological mismatch tallies were assigned unless word structure was affected. For example, during production of *hippopotamus* /ˌhɪpəˈpʰɑːɾəməs/, *hip* and its word neighbours, such as *hit,* were possibly activated. The added competition from activation of coronal segmental place features in /ɾ/ and /s/ could contribute to lack of inhibition of *hit,* facilitating the appearance of [hɪt] in the first syllable of a production like [ˌhɪt.təˈpʰ ɑːtəməs]. In the first foot, ˌ/hɪpə/, the syllable, [hɪt], would be tallied as a lexical selection mismatch, with no further tallies, because of matching syllable structure with [hɪp]; the syllable, [tə], would then be tallied for the consonant timing unit and feature insertions of [t] (see Table 3.3).

Phonological components were next evaluated in turn by tier: stress for the prosodic word and each foot, syllables, consonant and vowel timing units, and segmental features. For each foot, stress mismatches were tallied for shifts of primary stress, or for altered stress patterns as a result of syllable deletion or insertion. For a two-footed word, a maximum of one tally was assigned for prosodic word stress, whether a mismatch occurred in either or both feet. Timing unit tallies were assigned if: (a) segments were metathesized, inserted, or deleted without compensatory lengthening; (b) a full vowel was inserted before a syllabic consonant; (c) lax vowels in unstressed syllables appeared as full vowels; and (d) taps were produced as (longer) stops. Segmental feature mismatches were tallied for consonants in terms of place, manner and laryngeal feature categories (one tally per category maximum); and for vowels, with respect to height, backness, and tenseness, in addition to nasalization and roundness, when relevant. Non-allophonic and non-assimilatory diacritics, for consonants and vowels, respectively, were also tallied.

The MSWCV metric (i.e. vowels included with consonants) is next demonstrated for English, *hippopotamus* /ˌhɪpəˈpʰɑːɾəməs/, and French, *hippopotame* /ʔipɔpɔtam/ (see Table 3.3 and Table 3.4, respectively). Transcripts appear in ascending order of the corresponding MSWCV total and rank used for the statistical analyses. In cases of tied totals, the average of the potential ranks was assigned (Field, 2009). Because the tables were intended to present the data for all of the children, correct targets were also included, e.g. in Table 3.3, for Child 1 and Child 2. Further explanation of mismatch tally assignment follows, with reference to Table 3.3.

Table 3.3 compares various productions of /ˌhɪpəˈpʰɑːɾəməs/, a word with two feet, /ˌhɪpə/ and /ˈpʰɑːɾəməs/. Beginning with *lexical selection*, only [ˌhɪb̥əˈnɑːnənoʊs], received a mismatch tally. The diphthong [oʊ] in the WF syllable was considered influenced by activation of the common word abbreviation, /ˌhɪpoʊ/. As a result of [oʊ] transposition, a vowel timing unit and its associated segmental features were inserted, all of which received tallies. The [o] was tallied for height, backness, tenseness and rounding, whereas the lax vowel, [ʊ], was considered an acceptable schwa variant, and therefore, not tallied. Stress mismatches were relevant to [ˌhɪˌpəˈpʰɑːˌtʰəˌməs] and [ˌhɪpəˈpʰɑːməs]. First, [ˌhɪˌpəˈpʰɑːˌtʰəˌməs] was tallied for equal syllable stress that altered first foot syllable prominence, for the second foot weak syllable sequence, and therefore, for stress of the prosodic word, totalling three tallies. For [ˌhɪpəˈpʰɑːməs], the WF weak syllable sequence was altered through syllable deletion of /ɾə/, mismatching both second foot and prosodic word stress, for two tallies. The interaction of stress with word length was reflected in the syllable deletion tally that compounded in tallies for corresponding consonant and vowel timing unit deletions, along with segmental features. By comparison, for [ˌhɪpəˈpʰɑːməs], with compensatory vowel lengthening, only consonant timing unit deletion was tallied, but the features of /ɾ/ and /ə/ were tallied nevertheless.

Mismatched features of the tap, /ɾ/, were also central to other productions of /ˌhɪpəˈpʰɑːɾəməs/, shown in Table 3.3. Manner features considered for /ɾ/ included: [–nasal, –lateral, +sonorant, +continuant] (Bernhardt & Stemberger, 1998); however, no tally was assigned for [–continuant] (redundant to [+sonorant], which was included). Concerning [ˌhɪpəˈpʰɑːdəməθ] and [ˌhɪˌpəˈpʰɑːˌtʰəˌməs], /ɾ/ appeared as [d] or [tʰ], respectively, so was tallied for [–sonorant], in addition to [-voice] for the latter. For [ˌhɪpəˈpʰɑːdəməθ], the place mismatch for /s/, realized as [θ], was also tallied. Next, in [ˌhɪpəˈpʰɑːnəməs] and [ˌhɪpəˈpʰɑːməmɪs̪], nasal assimilation was tallied for /ɾ/ produced as [n] and [m], respectively; and in addition, the latter for reduplicated [labial]. The production [ˌhɪpəˈpʰɑːməmɪs̪] was also tallied for dentalization of /s/. The final production [ˌhɪb̥əˈnɑːnənoʊs] was tallied for several assimilatory processes: (a) in the non-prominent foot, [ˌhɪb̥ə], voicing of /p/; and (b) in the prominent foot, [ˈnɑːnənoʊs], nasalization of /pʰ/ and /ɾ/, and additionally for coronal assimilation for /pʰ/ and /m/.

3.3.3.3 Methods for the Comparison Metrics

The procedures for the comparison tallies derived from the PMLU and PCC metrics, were in accordance with Ingram (2002) and Shriberg (1994), respectively. For the PMLU, one tally was assigned each vowel (two for long vowels and diphthongs, a departure from typical PMLU calculations, but matching the current MSW metric), consonant and syllabic consonant (e.g. /l/ or /ɹ/) produced, except for any segments in excess of the adult target. Each correct consonant was allotted one additional tally. For the PCC tally, only consonants (including syllabic consonants and taps) that matched the adult target completely were counted as correct; hence, non-target diacritics were counted as consonant mismatches. Tally totals and ranks derived from the PMLU and PCC methodologies are included in Table 3.3 and Table 3.4.

3.3.3.4 Reliability

For each of the MSW tally metrics, the first and second authors (Mason & Bernhardt) established intra-rater reliability of mismatch tallies on all ten productions of two English words and all eight productions of two French words, first through a joint training session, building consensus for each of the words. The raters then re-tallied the same words independently until each reached 100% agreement with the consensus tallies twice.

For inter-rater reliability, an independent coder, a graduate of an SLP Master's thesis program, tallied all of the words for a randomly selected proportion of the samples: for English, four children balanced for gender (40%); and for French, two females (25%). The greater word complexity of English accounted for the higher proportion relative to French. Reliability was tested on the two-way random intraclass correlation (ICC), across all of the tallies for the MSWCV metric, and for each of the MSWCV, PMLU tally and PCC totals. The results corresponding to the comparisons for English were: ICC(2,1) = .94, .90, .96, .883, and the 95% CIs [.93, .95], [.79, .95], [.99, .98], and [.75, .95], respectively. The corresponding results for French were: ICC(2,1) = .83, .96, .93, .90, and the 95% CIs [.71, .89], [.95, .99], [.79, .98], and [.77, .97], respectively. The correlations suggested high consistency of the MSWCV metric with each of the comparison metrics, MSWCV, PMLU tally and PCC totals. Testing of the MSWC total was considered redundant because of its derivation from spreadsheet formula subtraction of vowels from the MSWCV total.

Disagreements were examined across all of the tallies for the MSWCV metric and were centered on alignment related to assimilatory processes. For instance, for /ˈkʰæːʃˌɹɛʤɪstɚ/ realized as [ˈkʰæːθdɚˌɹɛðdɚ], there was possibly reduplication of the WF sequence, [dɚ], in the first foot, [ˈkʰæːθdɚ], and deletion of the syllable, /ʤɪ/, from the second foot, [ɹɛðdɚ]. Alternatively, the CV timing units corresponding to /ʤɪ/ might have transposed and subsequently been filled with segmental features of /dɚ/. Another example concerned the medial consonant, [ð], in the second foot. On the one hand, there was [+continuant] correspondence of [ð] with target cluster onset /s/, and potentially concomitant sibilant [+anterior] place mismatch

(/ʃ/ realized as [θ]). The [+voice] feature of [ð] might then have been assimilated from any or all other segments within the foot. On the other hand, there was [+voice], and to some extent, [+continuant] correspondence of [ð] with /ʤ/, potentially co-occurring with sibilant [+anterior] place mismatch. If so, [ð] would appear in the coda of [ɹɛð], a default of vowel timing unit deletion associated with /ɪ/. This scenario would have prevented an arguably disallowed [ðd] English consonant onset sequence. Together, these examples demonstrate the many tally options associated with the complexities of MSWs and potential processing interactions, and supports the relatively greater utility of a whole word metric.

3.3.3.5 Data Analysis

In order to compare the different analysis procedures, rank ordering of words and participants by tally total was done according to each metric: the current metric, PMLU, PCC. In cases of tied totals, the average of the potential ranks was assigned (Field, 2009). Convergent validity with respect to consistency of the relationships of the tally procedures with PMLU and PCC, was evaluated for each word, using the Spearman rank order correlation (r_s). Sensitivity was considered in terms of whether the new metric produced significantly fewer ties than the other measures, within various word productions and across children. Friedman's χ^2 ($p < .05$) was used to examine: (a) the number of ties derived by each of the metrics, in comparisons of the number of different ranks for each word, and for the children; and (b) whether any metric systematically ranked the data higher or lower than another. Wilcoxon's Signed-rank test was used for the post hoc analyses, but without Bonferroni correction for the number of tests, because the small sample reduced the power to detect significance. The procedure in turn increased the probability of a Type I error, however, suggesting cautious interpretation of the results. Effect size, *r*, was also calculated for each of the post hocs. For English and French, three comparisons were made relative to convergent validity and sensitivity: (a) a MSW metric with vowel features (MSWCV) versus one without such features (MSWC), because of potential low vowel transcription reliability (Ingram, 2002); (b) a MSW metric with vowel timing units (MSWCV) versus tallies for PMLU, because the latter excluded vowel features; (c) a MSW metric without vowels (MSWC), versus tallies for PCC, because the latter included consonants only.

3.4 Results

The results are presented as follows. First is convergent validity for both languages, and next, sensitivity for English, and subsequently, French.

3.4.1 Convergent Validity

The correlations for the whole word total tallies of the MSW metric with PMLU and PCC are presented for English, in Table 3.5, and for French, in Table 3.6. For both languages, there were consistent, very high negative relationships between tallies for MSWCV and PMLU, the expected direction given the match to mismatch comparisons. The strength of the relationships was as predicted, supporting similar theoretical basis of the metrics. In contrast, significance for the MSWC and PCC tallies was mixed as to words, with significant strong negative correlations found for three to four of six words, in both languages. Predictions were upheld that correlations of tallies for MSWC with PCC would generally be weaker than those for MSWCV and PMLU, because the new metric included more levels of the nonlinear hierarchy. The strong correlations of the MSWCV with PMLU for all words suggested the current metric was an appropriate and preferable whole word measure of MSW phonology.

3.4.2 Sensitivity

For English, the number of different ranks obtained for the productions of each word on the metrics, is displayed in Figure 3.1, and for the children, in Table 3.7. In Figure 3.1, there were fewer tied ranks for the productions of each word in the comparisons of MSWCV with PMLU tallies, and of MSWC with PCC tallies. Comparing MSWCV with MSWC, there were fewer tied ranks for the productions of three words, *explodes*, *hospital,* and *hippopotamus*; and equal numbers for the three others, *computer*, *thermometer*, and *cash register*. With the exception of participant F2, MSWCV and MSWC sequenced the children similarly in terms of accuracy, but there was less agreement with the other metrics, the least with PCC tallies, which appeared more random in comparison. The mean number of different ranks was as follows: MSWCV = 7.33 $\pm$1.03 (range: 6-9); MSWC = 6.67 $\pm$1.21 (range: 6-9); PMLU tally = 4.83 $\pm$1.17 (range: 3-6); PCC tally = 4.0 $\pm$0.89 (range: 3-5), decreasing as fewer levels of the phonological hierarchy were accounted for.

Significance for the statistical tests was set at $p < .05$. Friedman's test suggested significant differences in the number of different rankings for the English words, χ^2 (3, N = 20) = 15.87, p = .001. Two of the post hoc Wilcoxon Signed-rank tests were also significant: MSWCV compared with PMLU tallies, Z = 2.21, p = .026; and MSWC tally compared with PCC tallies, Z = 2.23, p = .027, with large effect sizes (r = .64). Friedman values were also examined as to whether any metric systematicity ordered the total tallies higher or lower than any other, with significance obtained for all words: for *explodes, computer, hospital, thermometer*, *cash register* and *hippopotamus,* respectively, χ^2 (3, N = 20) = 19.77, 10.85, 12.6, 15.9, 11.33, and 20.3. One or more post hoc comparisons using the Wilcoxon test were also significant for each word, with large effect sizes ($r > .50$): MSWCV with MSWC, for *hospital*, *cash register* and *hippopotamus*; MSWCV compared with PMLU tallies, for *explodes, computer, hospital* and *hippopotamus*; and MSWC compared with PCC tallies, for *explodes, thermometer*, and *cash register*. Concerning the test of the total tallies for the children, significance was again found, χ^2 (3, N = 20) = 26.20, $p = .001$. The pertinent post hocs were also significant: MSWCV compared with MSWC, $Z = 2.52$, $p = .01$, $r = .56$; MSWCV compared with PMLU tallies, $Z = 2.50$, $p = .01$, $r = .56$; and MSWC compared with PCC tallies, $Z = 2.80$, $p = .005$, $r = .63$; all with large effect sizes.

For French, the number of different ranks obtained for each metric for each word is displayed in Figure 3.2, and for the children, in Table 3.8. Similar to English, there were more ranks for each word using the MSW metric (MSWCV and MSWC), than for the PMLU or PCC tallies. The MSWCV metric further dissociated the rankings from the MSWC metric for three words: *balançoire* (*swing*), *crocodile* (*crocodile*), and *hôpital* (*hospital*). Paralleling English, there was more ranking dispersion for the children across the MSWCV, MSWC and PMLU tally metrics than for the PCC tally.

Friedman's test was used to evaluate the number of different rankings derived by the four metrics for French, with significant differences, χ^2 (3, N = 16) = 15.94, $p = .001$. Post-hoc tests using the Wilcoxon Signed-rank were also significant for two of the analyses: the MSWCV tally compared with the PMLU tally, $Z = 2.14$, $p = .02$ and with the PCC tally, $Z = 2.21$, $p = .02$; and the MSWC tally compared with the PMLU tally, $Z = 2.12$, $p = .02$ and with the PCC tally, $Z = 2.33$, $p = .01$, with large effect sizes ($r > .50$). Friedman values were also examined as to whether any of the four metrics rank ordered the data systematically higher or lower than another. Results were significant for five of the six words, including *balançoire* /balɑ̃ˈswaʁ/, *crocodile* /kʀɔkɔˈdzɪl/, *hôpital* /ʔɔpiˈtal/, arc-en-ciel /ʔaʀkɑ̃ˈsjɛl/, and *dentifrice* /dɑ̃tsiˈfʀɪs/. $p < .05$). Post-hoc analyses using the Wilcoxon test showed that MSWCV was the most sensitive measure in differentiating productions of /ʔɔpiˈtal/, /ʔaʀkɑ̃ˈsjɛl/, and /dɑ̃tsiˈfʀɪs/.

3.5 Discussion

The current pilot study examined the application of a whole word MSW metric to analysis of the productions of six vulnerable MSWs in each of Canadian French and English, for ten typical-speaking 5-year-olds, and eight 3- and 4-year-olds with PPD, respectively. Mismatches for the various words were tallied for lexical effects and across the phonological hierarchy from the prosodic word to the feature. The aims were to acquire preliminary evidence concerning convergent validity with two previous metrics, PMLU and PCC, and to assess the sensitivity of the new metric to gradations of phonological variability, in addition to its reliability. Questions were: (a) whether the MSWCV and MSWC tallies would correlate with tallies for deriving PMLU and PCC across words and children; (b) whether the new metrics would be more sensitive, i.e. distinguish among the variable productions in terms of fewer tied (distinct) rankings of productions or speakers than PMLU and PCC; and finally (c) whether MSWCV better accounted for the data than MSWC. The results for English and French are discussed below, for the words (with a note on lexical effects), and then for participants.

Concerning the question of convergence, the number of significant correlations of the MSW metric with the comparison PMLU and PCC tallies, for the words in English and in French, suggested that the MSW metric provided a valid indication of MSW phonological status.

The stronger and more consistent correlations with PMLU indicated that inclusion of more levels of nonlinear phonology was applicable to a whole word MSW measure. Having also included lexical selection influences, finding these correlations suggested value in uniting theoretical aspects of language processing with those of phonology. Studies of large samples with more opportunities for lexical selection mismatches are needed, however, to establish the relative importance of such a level within a MSW metric. For example, lexical selection might be relevant to better understanding PPD. That is, lexical selection influences might be more frequent because the cue cost(s) of the multiple phonological components during licensing leaves insufficient cognitive capacities to inhibit incorrect selection. Lexically related components, with stronger connections in the semantic network, will have lower resting levels and therefore be activated more quickly as chunks. As for explaining the mixed results between MSWC and PCC, patterns of production differences among words were unclear, possibly because MSWs have unique complexities. Judged across words, however, the MSW metric appeared to have promising application as a whole word phonological measure, but more words will require examination to confirm this.

Relative to sensitivity, there were significant findings relative to the number of rankings in English and French, for the MSWCV and MSWC, i.e. in the tests of the tallies for MSWCV compared with PMLU, and the tallies for MSWC compared with PCC. Significance was associated consistently with arriving at fewer tied rankings with the MSW tally metrics across productions of all of the words (see Figure 3.1 and Figure 3.2). The increased range in the number of ranks provided further gradations of production inaccuracy. Visual inspection of the English data identified tied productions resulting from PCC tallies, which had very different phonological characteristics. For example, ranks were compared regarding two productions of *cash register* /ˈkʰæːʃˌɹɛdʒɪstɚ/, the first matched for target stress pattern, and the second, unmatched: [ˈkʰæːs̺ˌwɛdʒ̺ɪs̺t̺ə̣] versus [ˈkʰæːθdɚˌɹɛðdʊ]. On one hand, the former showed matches for foot and syllable structure, and features, except for small place mismatches for /ʃ, dʒ, s/, and rhotics /ɹ, ɚ/. On the other hand, the latter had more pervasive stress, foot and syllable structure mismatches because of syllable deletion ([dʒɪ]), and reduplication ([θdɚ, ðdʊ]). The PCC tallies ranked both productions 6.5th, whereas [ˈkʰæːθdɚˌɹɛðdʊ] ranked 9th or 10th by the other metrics, and [ˈkʰæːs̺ˌwɛdʒ̺ɪs̺t̺ə̣], 3rd to 6th. Thus, as expected, by analysing all levels of the phonological hierarchy plus possible lexical influences on word production, the new metric appears more sensitive than previous metrics.

Relative to the question comparing MSWC and MSWCV, there was no significant difference in tally rankings, although the metric that included vowels produced fewer ties for half of the words, i.e. *explodes, hospital,* and *hippopotamus*. A larger data set of vowels, particularly in more words longer than a foot, might reveal the greater utility of the MSWCV tally over the MSWC metric, but the predominance of consonant sequences in MSWs might also overshadow this. In any case, tallying vowel features points to the importance of the interaction of vowels with consonants or the preservation of syllable nuclei. The advantage of tallying vowel features would be the opportunity afforded to identify difficulties when they do exist.

Relative to lexical effects, few tallies were observed across the small word sets, one for *explodes* /ɪkˈsploʊdz/, as [ˌkʰəs̺pəˈləʊzˡ]; and two for *thermometer* /θɚˈmɑːmərɚ/: [hɚˈmɑːnərʊ], [θʊ̥ɚ̥ˈmɑməθtʊ]. The pronunciation [ˌkʰəs̺pəˈləʊzˡ] may have been influenced by a 'word' such as *kapow* or *kablam* in action adventures. The first production for *thermometer* may have shown interference of the pronoun, *her*, and the second, priming from a preceding word in the elicitation, *cash register* /ˈkʰæːʃˌɹɛdʒɪstɚ/. Although lexical effects were relatively few, in the cited instances, they did contribute to rank-order distinctiveness. Given a larger data set, such tallies might assist in delineating small differences among productions (Hall et al., 1998; McLeod et al., 2011) that arise as a result of lexical selection because of the competition in parallel interactive cognitive processes.

Concerning sensitivity relative to the participants, the three comparisons of the MSW tally metric also differed systematically with respect to higher or lower ranking of the productions across children in English and in French. Similar to what was described for words above, there were notable differences in the sequential order in the ranking of specific children, in comparisons of the MSW tally metrics with the tallies for PMLU and PCC. In Table 3.7, for example, F5 was ranked highest for the new metric (most mismatches), whereas F2 was lowest-ranked (least matches) using the tallies for deriving PMLU and PCC. Apart from developmentally appropriate segmental mismatches, inspection of the children's data revealed that M4 preserved foot structure in terms of stress patterns and syllable number, but had syllable structure mismatches for abutting consonant sequences (clusters and coda-onset). The data for F5 also included such consonant sequence mismatches, but in addition, showed mismatches of lexical selection, and several syllable insertions that compounded to disrupt the stress patterns, syllable structures and features of the target words. The breadth of lexical and phonological components included in the MSW metrics provided more detailed identification of the complexity differences between the phonological systems for these children than the PMLU and

PCC tallies. Table 3.7 shows, however, that the *number* of different rankings for the children was virtually equivalent for all measures except PCC, for which there were one-third fewer. The lack of correspondence in *sequential order of specific children,* however, reflected the variations in the scope of the theoretical underpinnings of the metrics.

3.6 Conclusion

Overall, the MSWC and MSWCV metrics appeared to be useful for evaluation of MSWs through analysis of multiple levels of the phonological hierarchy (and their interactions) plus possible lexical effects. The metrics can provide subtle distinctions between various productions of a word without the need for arbitrary weighting systems in at least two languages, English and French (and potentially, Spanish, Schretlen, 2013). Further evaluation of the metric is needed to determine: (a) the relative importance of vowel tallies; and (b) potential lexical influences, with more refined assessment of word familiarity, priming effects and range of neighbourhoods for syllables of MSWs. The current pilot study is limited by the small samples of words and children. Further research is also needed for languages other than English. In Chapter 4, for English, analysis for the main study for the thesis is presented, in which the metric is also discussed in further detail.

Table 3.1 Phonological characteristics for English words

Word	Target	Feet	Syllables	Stress	Word shape	*Extended-10* consonants and complexities
explodes	/ɪkˈsploʊdz/	1	2	wS	VCCCCVVCC	/s, l, z/ Initial weak syllable; diphthong; coda-onset cluster; onset cluster within; final syllable coda cluster Sequences: [+/- continuant]; Dorsal/Coronal; Labial/Coronal; Coronal/Coronal
computer	/kʰəmˈpjuːɾɚ/	1	3	wSw	CVCCCVCC̩	/ɚ/ Initial weak syllable; schwa; coda-onset cluster; onset cluster within; tap; final syllabic consonant coda Sequences: [+/-sonorant]; Dorsal/Labial; Labial/Labial/Coronal
hospital	/ˈhɑːspɪɾɫ̩/	1	3	Sww	CVCCVCC̩	/s, l/ Final weak syllable sequence; onset cluster within; tap; final syllabic consonant coda Sequences: [+/- continuant]; Coronal/Labial; Labial/Coronal
thermometer	/θɚˈmɑːməɾɚ/	1	4	wSww	CC̩CVCVCC̩	/θ, ɚ/ Initial weak syllable; schwa; within and final syllabic consonant; reduplicated consonant; tap Sequences: nasal; Coronal/Labial; Labial/Labial

Table 3.1 Phonological characteristics for English words (continued)

Word	Target	Feet	Syllables	Stress	Word shape	*Extended-10* consonants and complexities
cash register	/ˈkʰæːʃˌɹɛdʒɪstɚ/	2	4	Ssww	CVCCVCVCCC̩	/ʃ, ɹ, dʒ, s, ɚ/ Two feet; final weak syllable sequence; coda-onset cluster; onset cluster within; final syllabic consonant coda Sequences: [+/- continuant]; Dorsal/Coronal; Coronal/Coronal
hippopotamus	/ˌhɪpəˈpɑːɾəmɪs/	2	5	swSww	CVCVCVCVCVC	/s/ Two feet; weak syllable within; final weak syllable sequence; reduplicated consonant; tap Sequences: Labial/Labial; Labial/Coronal; Coronal/Labial

Note. *Extended-10* consonants include /tʃ/ and /dʒ/ with the *Late 8* (Shriberg, 1983), i.e. [ʃ, ʒ, s, z, l, ɹ/ɝ/ɚ, θ, ð]. Place sequences may be in both directions.

Table 3.2 Phonological characteristics for French words

Word	Target	Feet	Syllables	Stress	Word Shape	Complexities[a]
balançoire	ba.lɑ̃.ˈswɑʁ	1	3	wwS	CVCVCCVC	Onset cluster final; unstressed nasal vowel; Sequences: [-/+continuant]; Labial/Coronal; Coronal/Labial; Dorsal
crocodile	kʀɔ.kɔ.ˈdzɪl	1	3	wwS	CCVCVCfVC	Onset cluster initial; Sequences: [-/+continuant]; Dorsal/Coronal
hôpital	ʔɔ.pi.ˈtal	1	3	wwS	VCVCVC	Sequences: [-/+continuant]; Labial/Coronal
arc-en-ciel	ʔaʀ.kɑ̃.ˈsjɛl	1	3	wwS	VCCVCCVC	Onset cluster within; Onset cluster final; unstressed nasal vowel; Sequences: [+/-continuant]; Dorsal/Coronal; Palatal/Coronal
dentifrice	dɑ̃.tsi.ˈfʀɪs	1	3	wwS	CVCVCCVC	Onset cluster medial; Onset cluster final; unstressed nasal vowel, Sequences: [-/+continuant] within cluster; Coronal/Labial; Dorsal/Coronal
hipopotame	ʔi.pɔ.pɔ.ˈtam	1	4	wwwS	VCVCVCVC	4+ syllable, Sequences: [-/+continuant]; Labial/Labial; Coronal/Labial

Note. There are no French studies that have systematically determined benchmark ages at which phonemes are acquired. Later acquired consonants, beyond 48-53 months, include [ʃ, ʒ, s, j] (MacLeod, Sutton, Trudeau, & Thordardottir, 2011). Place sequences may be in both directions.

Table 3.3 Metric comparison for English: *hippopotamus*

Child	ˌ/hɪpəˈpʰɑːrəmɪs/	Lexical	Stress	Syllables	TUs	Features	MSWCV		PMLU Tally		MSWC		PCC Tally	
							Total	Rank	Total	Rank	Total	Rank	Total	Rank
1	[ˌhɪpəˈpʰɑːrəməs]	0	0	0	0	0	0	1.5	17	1.5	0	1.5	6	1.5
2	[ˌhɪpəˈpʰɑːrəməs]	0	0	0	0	0	0	1.5	17	1.5	0	1.5	6	1.5
3	[ˌhɪpəˈpʰɑːnəməs]	0	0	0	1	1	2	3	16	3.5	2	3	5	7
4	[ˌhɪpəˈpʰɑːdəməθ]	0	0	0	1	2	3	4	15	6	3	4	4	3.5
5	[ˌhɪpəˈpʰɑːməmɪs̥]	0	0	0	1	3	4	5	15	6	4	5	4	3.5
6	[ˌhɪˌpəˈpʰɑːˌtʰəˌməs]	0	3	0	1	2	6	6	16	3.5	6	6	5	7
7	[ˌhɪpəˈpʰɑːməs]	0	2	1	1	6	10	7	15	6	7	7.5	5	7
8	[ˌhɪpəˈpʰɑːməs]	0	2	1	2	6	11	9	14	9.5	8	9.5	5	7
9	[ˌhɪpəˈpʰɑːməs]	0	2	1	2	6	11	9	14	9.5	8	9.5	5	7
10	[ˌhɪb̥əˈnɑːnənoʊs]	1	0	0	1	9	11	9	13	10	7	7.5	2	10

Note. TUs = Timing units. MSWCV = Multisyllabic word consonant and vowel tally. MSWC = Multisyllabic word consonant tally. PMLU = Phonological mean length of utterance (Ingram, 2002). PCC = Percent consonants correct (Shriberg, 1994).

Table 3.4 Metric comparison for French: *hippopotame*

Child	/(ʔ)ipɔpɔˈtam/	Lexical	Stress	Syllables	TUs	Features	MSWCV		PMLU Tally		MSWC		PCC Tally	
							Total	Rank	Total	Rank	Total	Rank	Total	Rank
1	[(ʔ)ipɔpɔtam]	0	0	0	0	0	0	1	12	1	0	1	4	1
2	[(ʔ)ipɔpɔtam]	0	0	0	0	0	0	1	12	1	0	1	4	1
3	[(ʔ)ipɔpɔtam]	0	0	0	0	0	0	1	12	1	0	1	4	1
4	[biləpɔtan]	0	0	0	0	8	8	4	10	4	8	4	2	6.5
5	[jimpɔtam]	0	1	1	1	9	12	5.5	9	5.5	15	8	3	4.5
6	[kupɔtam]	0	1	1	1	11	14	5.5	9	5.5	11	5	3	4.5
7	[(ʔ)ipɔpa]	0	1	1	3	11	16	7	6	7.5	13	6.5	1	8
8	[ɔtam]	0	1	2	5	12	20	8	6	7.5	13	6.5	2	6.5

Note. TUs = Timing units. MSWCV = Multisyllabic word consonant and vowel tally. MSWC = Multisyllabic word consonant tally. PMLU = Phonological mean length of utterance (Ingram, 2002). PCC = Percent consonants correct (Shriberg, 1994).

Table 3.5 Correlations for the comparison metrics for English

Word	Target	Metrics compared	r_s
hippopotamus	/ˌhɪpəˈpʰɑːɾəmɪs/	MSWCV:MSWC	0.978*
		MSWCV:PMLU	-0.943*
		MSWC:PCC	-0.667*
cash register	/ˈkʰæːʃˌɹɛdʒɪstɚ/	MSWCV:MSWC	0.967*
		MSWCV:PMLU	-0.827*
		MSWC:PCC	-0.45*
thermometer	/θɚˈmɑːməɾɚ/	MSWCV:MSWC	1*
		MSWCV:PMLU	-0.683*
		MSWC:PCC	0.569
hospital	/ˈhɑːspɪɾɫ̩/	MSWCV:MSWC	0.941*
		MSWCV:PMLU	-0.67*
		MSWC:PCC	-0.452
computer	/kʰəmˈpjuːɾɚ/	MSWCV:MSWC	1*
		MSWCV:PMLU	-0.99*
		MSWC:PCC	-0.977*
explodes	/ɪk.ˈsploʊdz/	MSWCV:MSWC	0.956*
		MSWCV:PMLU	-0.681*
		MSWC:PCC	-0.13

Note. r_s = Spearman rank order correlation. MSWCV = Multisyllabic word tally of consonants and vowels. MSWC = Multisyllabic word tally of consonants only. PMLU = Phonological mean length of utterance (Ingram, 2002). PCC = Percent consonants correct (Shriberg, 1994).
*$p < .05$

Table 3.6 Correlations for the comparison metrics for French

Word	Target	Metrics compared	r_s
balançoire	/ba.lɑ̃.ˈswɑʁ/	MSWCV:MSWC	0.952*
		MSWCV:PMLU	-0.864*
		MSWC:PCC	-0.543
crocodile	/kʀɔ.kɔ.ˈdzɪl/	MSWCV:MSWC	0.940*
		MSWCV:PMLU	-0.982*
		MSWC:PCC	-0.584
hôpital	/ʔɔ.pi.ˈtal/	MSWCV:MSWC	0.901*
		MSWCV:PMLU	-0.926*
		MSWC:PCC	-0.956*
arc-en-ciel	/ʔaʀ.kɑ̃.ˈsjɛl/	MSWCV:MSWC	0.976*
		MSWCV:PMLU	-0.981*
		MSWC:PCC	-0.673*
dentifrice	/dɑ̃.tsi. ˈfʀɪs/	MSWCV:MSWC	0.957*
		MSWCV:PMLU	-0.925*
		MSWC:PCC	0.975*
hippopotame	/ʔi.pɔ.pɔ.ˈtam/	MSWCV:MSWC	0.956*
		MSWCV:PMLU	-0.892*
		MSWC:PCC	-0.517

Note. r_s = Spearman rank order correlation. MSWCV = Multisyllabic word tally of consonants and vowels. MSWC = Multisyllabic word tally of consonants only. PMLU = Phonological mean length of utterance (Ingram, 2002). PCC = Percent consonants correct (Shriberg, 1994).
*$p < .05$

Table 3.7 Ranking of English-speaking children across metrics

Child	Rank			
	MSWCV	MSWC	PMLU Tallies	PCC Tallies
M1	1	1	2	8
M2	2	2	1	1.5
M3	3	3	3	1.5
F1	4	4	6	8
F2	5	8	10	10
M4	6	5	4.5	8
M5	7	6.5	4.5	3
F3	8	6.5	9	5.5
F4	9	9	8	4
F5	10	10	7	5.5
Ranks[a]	10	9	9	6

Note. Ranking is from most to least accurate in the MSCV column. MSW = Multisyllabic words. MSWCV = Multisyllabic word tally of consonants and vowels. MSWC = Multisyllabic word tally of consonants only. PMLU = Phonological mean length of utterance (Ingram, 2002). PCC = Percent consonants correct (Shriberg, 1994).
[a]Number of different ranks.

Table 3.8 Ranking of French-speaking children across metrics

	Rank			
Child	MSWCV	MSWC	PMLU Tally	PCC
M3	1	2.5	3	1.5
M5	2	2.5	1	3
F2	3	1	2	1.5
M6	4	4	5	4.5
M4	5	6	4	4.5
F2	6	5	6	6
M2	7	7	7	7
M1	8	8	8	8
Ranks[a]	10	9	9	6

Note. . Ranking is from most to least accurate in the MSCV column. MSW = Multisyllabic words. MSWCV = Multisyllabic word tally of consonants and vowels. MSWC = Multisyllabic word tally of consonants only. PMLU = Phonological mean length of utterance (Ingram, 2002). PCC = Percent consonants correct (Shriberg, 1994).
[a]Number of different ranks.

Figure 3.1 Number of different ranks for the English words

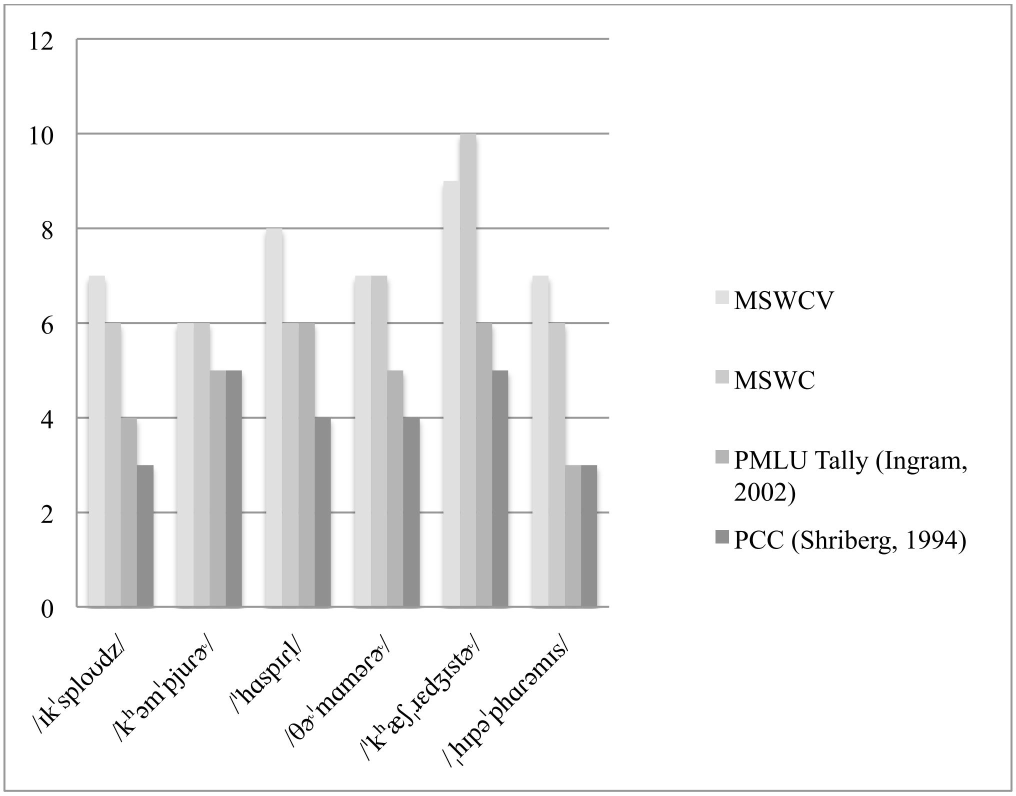
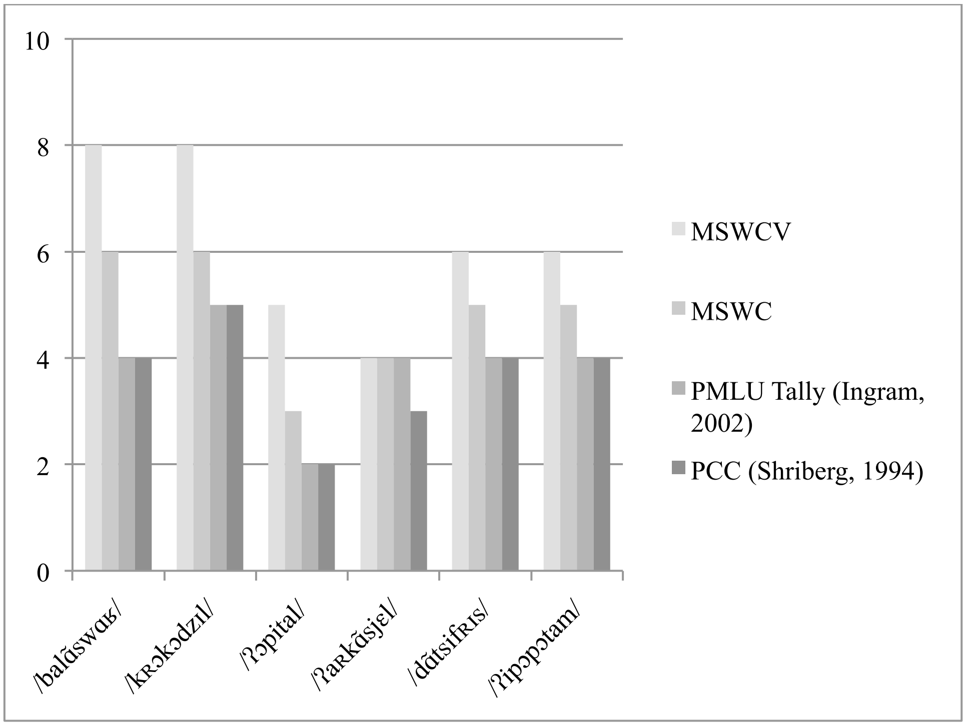


Figure 3.2 Number of different ranks for the French words



Chapter 4: MSW Production in School-Aged Children

4.1 Synopsis of Study

MSW production was examined in a sample of children pre- and post-early literacy instruction in three sub-studies: (a) longitudinal: children with typical speech (22 boys and 22 girls), i.e. 5-year-olds, and 8- to 10-year-olds; and (b) cross-sectional: 8- to 10-year-olds with PPD (7 boys and 5 girls) and those without, i.e. age-matched (7 boys and 5 girls), and a developmental comparison with 5-year-olds (28 boys and 34 girls). Twenty one- and two-footed words were evaluated (various syllable lengths, stress patterns, syllable structures and feature sequences). Lexical, word structure and feature mismatches decreased significantly between ages 5 and 8 years of age. Structure and feature mismatch frequencies were significantly higher for age-matched 8- to 10-year-olds with PPD than without, and were equivalent to 5-year-olds. The metric reliably classified children in their original groupings. Analyses also were conducted for word types in six groups organized by length and prominence patterns. Mismatch patterns were more unique than similar across words, but trends were similar, in that multiple interactions were observed between different hierarchical levels, emphasizing the importance of viewing the words from a nonlinear processing perspective. Future research using the measure for MSW phonological development appears worthwhile.

4.2 Introduction

The study presented in the sections below expands upon the pilot study in Chapter 3, by examining the use of the MSW whole word metric with a larger sample of Canadian English-speaking elementary school children pre- and post-early literacy instruction in their production of a larger number of MSWs (20 as opposed to 10). The sources of motivation for the current study of childhood MSW phonology were threefold: (a) James' (2006) cross-sectional evidence of a developmental progression in Australian English-speaking children; (b) the relationship to literacy suggested by the meta-analysis in Chapter 2; and (c) the utility of a whole word metric for quantifying mismatches, indicated by the pilot study in Chapter 3. Developmental information about MSWs is next briefly reviewed as background, before concluding the section with the proposed comparisons for the current study and the related research questions.

In Chapter 2, a summary of James' (2006) research outlined the growth in MSW accuracy between ages 4 and 7 years of age; i.e. children under the age of 7 mastered word length and phoneme sequences but showed stress pattern mismatches, which were subsequently accurate after age 7. Segmental sequencing accuracy appeared to interact with stress pattern and word length in two regards: (a) in terms of stress, earlier mastery occurred in words with initial syllable prominence; and (b) with respect to length, acquisition in disyllabic words preceded that in longer MSWs (at ages 4 and 7 years, respectively). Relative to the number of words elicited, the frequency of most processes was below 2%, e.g. consonant insertion, metathesis and transposition; whereas the most frequent were from 3% to 6%, e.g. syllable and consonant deletion, assimilation, and schwa tensing or lengthening. Rarely were syllable or schwa insertion observed.

Also in Chapter 3, the pilot study provided preliminary support for the utility of a whole word MSW mismatch tally measure, grounded in theories of nonlinear phonology (e.g. Bernhardt & Stemberger, 1998) and language processing (e.g. Presson & MacWhinney, 2010; Wheeler & Touretzky, 1997). The metric appeared suitable for addressing the complexities and interactions between lexical selection, and all prosodic levels of the phonological hierarchy (e.g.

word length, complexity of syllables, and segmental features, with their sequences). In Chapter 4, therefore, MSW phonological development is examined in three sub-studies, in which the metric was applied to productions of 20 MSWs. The first sub-study was a longitudinal investigation of children with typical phonological development (n = 22; 11 boys and 11 girls), evaluated at age 5 years, and re-evaluated between 8 and 10 years of age. The second and third sub-studies were cross-sectional, and compared children with and without PPD on the basis of age, and stage of literacy instruction. Respectively, the sub-studies examined: (a) post-early literacy age-matched 8- to 10-year-olds (N = 24; 14 boys and 10 girls); and (b) post-early literacy 8- to 10-year-olds with PPD (n = 12; 7 boys and 5 girls) and pre-early literacy 5-year-olds with typical phonological development (n = 62; 28 boys and 34 girls). In addition to statistical comparisons of the participant groups' metric data, mismatch patterns by word type were evaluated in a descriptive linguistic analysis.

The research questions and predictions for the longitudinal and cross-sectional sub-studies of MSW mismatch tallies are next posed in turn. Two questions are addressed for each sub-study, one to do with the quantitative analyses and one concerning the phonological mismatch patterns.

Beginning with the longitudinal research questions for children with typical phonological development, pre- and post-early literacy instruction, were:

1. For children with typical phonological development, evaluated pre-early literacy instruction at 5-years old, and again post-early literacy instruction at 8- to 10 years of age, what is the difference in mismatch frequency with respect to: (a) lexical selection, (b) word structure, (c) segmental features, and (d) a whole word total?

In theory, systems of phonology (Bernhardt & Stemberger, 1998) and higher levels of language processing (e.g. Presson & MacWhinney, 2010; Shattuck-Hufnagel, 1992; Wheeler & Touretzky, 1997) would interact, such that lexical selection mismatches would result. A significant and large between-groups difference was predicted for lexically related mismatches; more were expected for 5-year-olds, because longer lower-frequency words would possibly be unfamiliar. Concerning the MSWs examined in the current study, by ages 8 to 10 years, the influence of word familiarity would presumably be less important, such that lexical selection mismatches would be few. The developmental progressions identified for MSW production accuracy (James, 2006; Kehoe, 2001) suggested that mismatch frequencies would be higher at age 5 years than at ages 8 to 10 years. Because of the wide age range between-groups, significant and large mismatch differences were predicted for the word structure and segmental feature mismatches, and in addition, for the whole word total.

2. For children with typical phonological development, evaluated pre-early literacy instruction at 5-years-old, and re-evaluated post-early literacy instruction at 8- to 10-years-old, what is the difference in types of phonological mismatch patterns, with respect to lexical selection, word structure and segmental features?

In theory, patterns would reflect the interaction of lexical selection with the levels of the phonological hierarchy (Bernhardt & Stemberger, 1998; Presson & MacWhinney, 2010; Shattuck-Hufnagel, 1992). The evidence also suggested that patterns would be related more often to weak word-initial and within-word syllables (James, 2006; Kehoe, 2001). Word structure and feature sequencing patterns of typically developing Canadian and Australian English-speaking 5-year-olds were expected to be analogous, i.e.: (a) primary stress shifts; (b) syllable or consonant deletions; (c) schwa tensing or lengthening; (d) metathesis, transposition or reduplication; (e) feature assimilation; and (f) mismatched manner or place for later developing continuants and affricates, e.g. (s, z, ʃ, tʃ, dʒ, ʒ, v, θ, ð), and also for rhotics (James, 2006). For typically developing 8- to 10-year-olds, patterns limited to slight place mismatches were predicted, i.e. primarily fronting or backing of later developing continuants, e.g. (s, z, θ, ð), and derhotacization (James, 2006).

Second, the research questions for the cross-sectional sub-study of age-matched children with and without PPD were:

1. Concerning age-matched 8- to 10-year-olds with and without PPD, what is the difference in mismatch frequency in terms of: (a) lexical selection, (b) word structure, (c) segmental features, and (d) a whole word total?

Also relevant to this comparison was the prediction that in theory, lexical selection mismatches would result from interactions between systems of phonology (Bernhardt & Stemberger, 1998) and higher levels of language processing (e.g. Presson & MacWhinney, 2010; Shattuck-Hufnagel, 1992; Wheeler & Touretzky, 1997). Because the meta-analysis (Chapter 2) had suggested the relevance of MSW evaluation to describing PPD, children with PPD were expected to have higher frequencies for each sub-total (i.e. lexical, word structure and segmental features), and the whole word total. Significant between-groups differences were also predicted (James, 2006; Kehoe, 2001) but smaller than those between typically developing 8-year-olds and 5-year-olds (the longitudinal sub-study). First, because the children with PPD were several years older than the typically developing 5-year-olds, their phonological systems were expected to be more adult-like (Bernhardt & Stemberger, 1998). Second, word familiarity of the MSWs examined in the current study would presumably be less important for 8-year-olds, with or without PPD. For the lexical sub-total, therefore, a small between-group difference was predicted. Moderate, as opposed to large differences were expected for the word structure and feature subtotals, and for the whole word total. Finally, the expectation was that mismatches of typically developing 8- to 10-year-olds would be limited to features (James, 2006).

> 2. Concerning age-matched 8- to 10-year-olds with and without PPD, what is the difference in types of phonological mismatch patterns, with respect to lexical selection, word structure and features?

In theory, patterns would again reflect the interaction of lexical selection with the levels of the phonological hierarchy (Bernhardt & Stemberger, 1998; Presson & MacWhinney, 2010; Wheeler & Touretzky, 1997). The evidence also suggested that patterns would be related more often to weak word-initial and within-word syllables (James, 2006; Kehoe, 2001). Possibly, MSW phonological patterns of 8- to 10-year olds with PPD would resemble younger typically developing children, e.g. Australian English-speaking 5- or 6-year-olds (James, 2006). Being older, however, 8- to 10-year-olds with PPD were expected to produce more elements of word structure (Bernhardt & Stemberger, 1998; James, 2006), such that deletions would be rare. Compared with older typically developing Australian English-speaking children, however, e.g. 7- to 8-year-olds (James, 2006), assimilatory patterns of 8- to 10-year-olds with PPD were predicted to be more various, i.e. include metathesis, transposition, and reduplication, in addition to feature sequencing mismatches. Specifically, the following patterns were predicted: (a) primary stress shifts; (b) schwa tensing or lengthening; (c) metathesis, transposition, and reduplication; (d) feature assimilation; and (e) mismatched manner or place for later developing continuants and affricates, e.g. (s, z, ʃ, tʃ, dʒ, ʒ, v, θ, ð), and also for rhotics. For typically developing 8- to 10-year-olds, patterns limited to slight place mismatches were predicted, primarily fronting or backing of later developing continuants, e.g. (s, z, θ, ð), and derhotacization (James, 2006).

Third, the research questions for the cross-sectional sub-study of 5-year-olds with typical phonological development pre-literacy instruction, and 8- to 10-year-olds with PPD post-early literacy were:

1. For 8- to 10-year-olds with PPD post-early literacy instruction, versus typically developing 5-year-olds pre-early literacy instruction, what is the difference in mismatch frequency with respect to: (a) lexical selection, (b) word structure, (c) segmental features, and (d) a whole word total?

In theory, interactions between systems of phonology (Bernhardt & Stemberger, 1998) and language processing (e.g. Presson & MacWhinney, 2010; Shattuck-Hufnagel, 1992; Wheeler & Touretzky, 1997) implied that lexical selection mismatches would occur. A significant and moderate between-groups difference was predicted for the number of lexically related mismatches, the most for the 5-year-olds because longer words with their low frequencies possibly would be unfamiliar. Concerning the MSWs examined in the current study, by 8 to 10 years of age, the influence of word familiarity presumably would be less important for children with PPD, such that the number of lexical selection mismatches would be few. Because of the wide age range between groups and the developmental progressions identified for MSW production accuracy (James, 2006; Kehoe, 2001), 8- to 10-year-olds with PPD were expected to have lower word structure and segmental feature subtotals. With respect to these mismatch subtotals, and in addition, the whole word total, significant but small between-group differences were predicted.

2. For 8- to 10-year-olds with PPD compared with typically developing 5-year-olds, what is the difference in types of phonological mismatch patterns, with respect to lexical selection, word structure and features?

In theory, lexical selection in interaction with the levels of the phonological hierarchy would again result in lexical mismatches. Studies (James, 2006; Kehoe, 2001) suggested that mismatch patterns would also relate more often to weak word-initial and within-word syllables. Word structure and feature sequencing patterns of typically developing Canadian and Australian English-speaking 5-year-olds were expected to be analogous, i.e.: (a) primary stress shift; (b) syllable or consonant deletion; (c) schwa tensing or lengthening; (d) metathesis, transposition and reduplication; (e) feature assimilation; and (f) mismatched manner or place for later developing continuants and affricates, e.g. (s, z, ʃ, tʃ, dʒ, ʒ, v, θ, ð), and also for rhotics (James, 2006). Possibly, the MSW phonology of 8- to 10-year olds with PPD would resemble typically developing 5-year-olds. Being older, however, 8- to 10-year-olds with PPD were expected to produce more elements of word structure (Bernhardt & Stemberger, 1998; James, 2006), such that deletions would be rare. Similar to typically developing 5-year-olds, however, assimilatory patterns of 8- to 10-year-olds with PPD were predicted to include metathesis, transposition and reduplication, in addition to feature sequencing mismatches (James, 2006). More specifically, the following processes were expected: (a) primary stress shifts; (b) schwa tensing or lengthening; (c) metathesis, transposition and reduplication; (d) feature assimilation; and (e) mismatched manner or place for later developing continuants and affricates, e.g. (s, z, ʃ, tʃ, dʒ, ʒ, v, θ, ð), and also for rhotics.

4.3 Method

In this section the methodology is described for the three quantitative sub-studies: (a) a longitudinal investigation of children with typical phonological development, evaluated pre-literacy at age 5 years (5TD), and re-evaluated post-literacy between 8 and 10 years of age (8TD); (b) a cross-sectional study of age-matched post-literate 8- to 10-year-olds, one group with typical phonological development (8TD) and the other with PPD (8PPD); and (c) a cross-sectional comparison of post-literate 8- to 10-year-olds with PPD (8PPD) and pre-literate 5-year-olds (5TD) with typically developing phonology. For the second part of the chapter on phonological mismatch patterns, details on the descriptive linguistic analysis of the word types is included in the preamble to that section (4.5.1).

The current section first describes the sample selection of the four groups of children for the sub-studies, one 5TD group, two 8TD groups, and one 8PPD group. Next, the testing procedures for evaluating the children on language-related skills, literacy and MSW production are outlined. Third, criteria for sample selection of the 20 MSWs are reviewed, and in addition, the transcription procedures for the word elicitations. Reliability procedures for the transcriptions and for the assignment of the MSW tallies are then described. The section concludes with a summary of the statistical analysis procedures used to examine the groups, starting with equivalence on two language related variables, vocabulary and working memory. Finally, the multivariate analyses (MANOVA) used to compare the children on the MSW tally sub-totals and total tallies are summarized.

4.3.1 Method for the Participant Samples

The methodology for sample selection and phonological evaluation of the 5-year-olds with typical phonological development (n = 62) was detailed in section 3.3.1. Procedures were replicated to sample and evaluate an equivalent sample size of 8- to 10-year-olds with typical phonological development, for future comparisons. A random sample of 64 children (28 boys and 25 girls) met the eligibility requirements, and was representative of the demographics of the school district. Concerning the current longitudinal study, 8- to 10-year-olds who were first evaluated at age 5 years were identified, such that each age group sub-sample included 22 children (11 boys and 11 girls).

In order to sample children with history of PPD pre- and post-early literacy instruction, all four school-based speech-language pathologists (SLPs) were asked to distribute study information and consent forms to caregivers of children on their caseloads, enrolled in Kindergarten through Grade 5 (the final elementary school year in the school district). In addition to meeting the eligibility criteria for birth, medical and developmental histories, by verbal report of the school-based SLP, the children were required to have had average-range composite standard scores in language comprehension and expression when originally assessed. One clinician declined to participate, and therefore, approximately 25% of schools (also representative of the district demographics) were not sampled. A convenience sample of 21 children resulted, with the youngest seven enrolled in Kindergarten to Grade 2 (age range = 6;1 to 7;4); these students were excluded from the study because the age range of the remainder (7;8

to 10;0) better approximated that of the 8- to 10-year-olds with typical phonological development. For all of these 12 children, who had received therapy in the past, SLPs considered them to demonstrate PPD in comparison with their peers. Two more children were excluded because of at-risk birth history, such that the final included group numbered 12 (7 boys and 5 girls). Subsequently, an age-matched (within six months) comparison sub-group of 8- to 10-year-olds with typical phonological development was selected from the larger sample. For an age tie, the next matching criterion was the period of the school year when evaluation took place, either Fall or Spring.

Testing occurred over two 1-hour sessions, evaluating the children at their schools on a wide variety of language- and literacy-related tasks (the latter included for future study). In order to demonstrate that the comparison groups had average and equivalent language-related skills that might confound phonological production of long words, two tasks were administered: (a) vocabulary comprehension (*Peabody Picture Vocabulary Test-4, PPVT-4;* Dunn & Dunn, 2007); and (b) working memory (*Number Repetition* subtests of the *Clinical Evaluation of Language Functions-4, CELF-4;* Semel, Wiig, & Secord, 2003). Regarding the *CELF-4,* to challenge working memory, all 8- to 10-year-olds were administered *Number Repetition Backward*, but to avoid ceiling effects, the 5-year-olds were tested on *Number Repetition Forward* instead. Concerning MSW phonology, the *Profile* and *IPE4* of the *Computerized Articulation and Phonology Evaluation System* (*CAPES*; Masterson & Bernhardt, 2001) were administered to the 5-year-olds with typical phonological development and children with PPD during the first and second sessions, respectively. Solely *IPE4* was administered to the 8- to 10-year-olds with typical phonological development, to reduce the overall duration of testing. Task presentation order was balanced with the comprehension and production demands of the language- and literacy-related measures.

4.3.2 Method for the Multisyllabic Words (MSWs)

This section begins with the methodology for MSW selection, before turning to the procedures for elicitation, transcription and assignment of MSW mismatch tallies to productions of the words. First, as described in section 3.3.1, 20 MSWs from the *Profile* and *IPE4* of the *CAPES* (Masterson & Bernhardt, 2001) were selected for analysis, based on phonological complexity and word familiarity. To review, word complexities included: (a) representative lengths and stress patterns, i.e. one to two feet comprised of two to five syllables, with unstressed word-initial, final and medial syllables, in addition to sequences of word-final weak syllables; (b) early and later-developing segments; (c) challenging adjacent and non-adjacent consonant sequences of manner, e.g. [±continuant], [nasal], tap and syllabic, and place, e.g. coronal-dorsal, dorsal-coronal and coronal-labial; and (d) diphthongs and lax vowels. Word familiarity was considered in terms of word frequency and Age of Acquisition (AoA), variables associated with phonological accuracy (Gierut & Morrisette, 2012). In order to avoid parental survey data typically used for AoA, word frequency was chosen to indicate familiarity, and was sourced from a large sample (Lee, 2001) of objective child speech data (Morrison, Chappell & Ellis, 1997). ChildFreq (Baath, 2010) was used to calculate frequencies, because probability and neighbourhood density are not central components, and age ranges relevant to the current study are included.

With regard to transcription and assignment of mismatches tallies, conventions for all sub-studies were in accordance with procedures described in sections 3.3.1 and 3.3.3.2, respectively. However, decisions made concerning the prosodic and feature components of particular segments deserve further explanation, keeping in mind that mismatches were conceptualized as differences from optimal adult targets, as opposed to errors. Specifically, to improve research and clinical reliability, uniform opportunities were implemented for independent judgments for schwa, tap, and syllabic consonants across the levels of the phonological hierarchy. In terms of prosody, a unique timing unit was assigned to each segmental type, for the reasons that follow. First, concerning schwa, James (2006) had reported developmental lengthening in MSWs up to age 7 years, implying the existence of a shorter adult timing unit than for other vowels. Second, manner of tap production had been distinguished from that of other consonants in terms of its ballistic nature, or speed and trajectory (Ladefoged 2006; Shockey, 2003). Third, for consonants in syllabic as opposed to non-syllabic contexts, differing acquisition patterns had been suggested (Bernhardt & Stemberger, 1998). Regarding features, schwa was described with respect to traditional height, backness and tenseness vowel features; the laryngeal feature for tap was considered [+voice], and relevant consonant manner, place and laryngeal features were ascribed to syllabic consonants (Bernhardt & Stemberger, 1998). Additional decisions concerning accepted and unaccepted segmental feature variants in general, and with respect to specific words, are listed with the groups.

The mismatch tally data was aggregated to generate three sub-totals: Lexical, Structure and Feature, in addition to a Whole Word total. For the current study, consonant diacritic tallies were included in the Feature sub-total, and not solely in the Whole Word Total, because the pilot study had demonstrated that variable productions of a word could be differentiated whether or not diacritics were included in the calculations. Combining feature and diacritic tallies, therefore, allowed all segmental mismatches to be represented by the Feature subtotal.

4.3.3 Reliability

Inter-rater reliability was considered for both transcription and assignment of MSW mismatch tallies. In all cases, joint training sessions were completed with the thesis author and the other raters, building consensus for 10% of the data. The raters then re-tallied the same productions independently until each reached 100% agreement with the consensus tallies twice in a row.

Transcription reliability was completed for 10% of independent transcriptions, and is reported in turn for the various groups in the sub-studies. For the TD 5-year-olds, point-to-point reliability was high between the author and an SLP graduate of a Masters thesis program (section 3.3.1). For the TD 8- to 10-year-olds, inter-rater reliability and consistency were good between the thesis author and two graduate SLP students, Kappa = .668, 95% CI [.607, .669], $p = .0001$, where Kappa = .7 is considered good agreement; Cohen, 1998). Transcription inter-rater reliability for the 8- to 10-year-olds with PPD was between the thesis author and an assistant professor, also an SLP graduate. Point-to-point reliability was 83.5%, the expected level for samples of children with PPD.

Next, inter-rater reliability was examined in independent mismatch tally assignments, between the thesis author and an SLP graduate of a Master's thesis program, who had also participated in establishing tally reliability for the pilot project. Point-to-point agreement was determined for each group of TD 8- to 10-year-olds, i.e. for the comparisons with TD 5-year-olds, and 8- to 10-year-olds with PPD; the relevant point-to-point agreement values were 99.24% and 98.86%, respectively. To determine reliability and consistency for the TD 5-year-olds and the 8- to 10-year-olds with PPD, the single measure ICC was calculated using a two-way random effects model. For the TD 5-year-olds, reliability and consistency were high for the coding overall and the mismatches alone, ICC(2,1) = 0.925, 0.819, and the 95% CIs [0.921, 0.930], [0.785, 0.849], $p = 0.001$, respectively. Finally, for the children with PPD also, reliability and consistency were high for the coding overall and the mismatches alone, ICC(2,1) = 0.951, 0.872, and the 95% CIs [0.948, 0.955], [0.842, 0.896], $p = 0.001$.

4.3.4 Statistical Analysis

One major purpose of the study was to evaluate children with and without PPD pre- and post-early literacy instruction on a whole word MSW mismatch metric in three comparisons: (a) the same children with typical phonological development, at 5 and 8 years of age; (b) age-matched 8- to 10-year-olds with and without PPD; and (c) 8- to 10-year-olds with PPD and 5-year-old typically developing children. Prior to analysis of the mismatch tallies, the data were examined for average standard score means and group equivalence on two language-related variables: vocabulary comprehension (*PPVT-4*) and working memory (*CELF-4 Number Repetition Forward* or *Backward* sub-tests). In addition to the Total tallies, sub-totals reflecting nonlinear phonological and language processing sub-constructs were also tested: i.e. Lexical, Word Structure and Feature. The data were entered into one-way multivariate (MANOVA, $p < .05$), and planned univariate analyses (ANOVA, $p < .01$), adjusting the probability level of the latter in the situation of unequal variances. Mean differences were examined using Wilk's Lambda and related *F*-tests, and effect sizes (partial eta squared, η^2_p) were calculated. Discriminant analyses were also conducted to determine which MSW variables best accounted for the variance (Wilk's Lambda, with effect size, canonical R^2), and sensitivity and specificity of the metrics were considered in post hoc classification analyses. In the next section, for the three comparisons in sequence, quantitative descriptive and inferential results are presented, followed by phonological analyses of mismatch patterns.

4.4 Results for the Statistical Analyses

The results of three sub-studies will be presented in the following order: (a) a longitudinal study of children with typical phonological development (TD; n = 22) who took part at age 5 years, and then between 8 and 10 years of age; and (b) two cross-sectional studies of children with PPD (n = 12) compared with TD children; first, in comparison with an age-matched group of TD 8- to 10-year-olds (n = 12) and next, with a group of TD 5-year-olds (n = 62). Each section begins with descriptive and statistical comparisons on the group equating *language* variables, i.e. standard scores for vocabulary comprehension (*Peabody Picture Vocabulary Test-4;* Dunn & Dunn, 2007) and working memory (*Number Repetition* subtests of the *Clinical Evaluation of Language Functions-4;* Semel, Wiig, & Secord, 2003), before turning to the main descriptive, statistical and follow-up analyses. All effect sizes were interpreted according to Cohen's (1988) conventions.

Concerning the total mismatches for the study, the proportions for the groups are displayed in Figure 4.1. Given a developmental progression in MSW accuracy, the expectation was that the proportion would be smallest for the 8TD, followed by the 8PPD and 5TD. The predicted standing was upheld for the 8TD relative to the other groups, but not between the 5TD and 8PPD. In comparison with the 5TD, the proportion for the 8TD was about one-third greater which suggested that MSW accuracy increased between the age of 5 years and ages 8 to 10 years, as shown by James (2006). For the 8PPD, the proportion relative to the 5TD was one-tenth larger, however, suggesting more accurate MSW production for the latter group. The statistical analyses examined the comparability and significance of these proportionate differences in the sub-studies, the results of which are reported in the next sections.

4.4.1 Longitudinal Comparison: 5-year-olds and 8- to 10-year-olds with Typical Phonological Development

For the longitudinal comparison, the age range of the TD 5-year-olds (n = 22), was 5;3 to 6;0, M(sd) = 5.64 (3.08 months), and for the TD 8- to 10-year-olds (n = 22), 8;7 to 10;3, M(sd) = 9.44 (6.35 months). The groups were first evaluated for equivalence on the language variables, on which all children scored within the average range. Assumptions for normality, linearity, univariate and multivariate outliers were examined by visually inspecting histograms, and Q-Q, box, and scatter plots, in addition to conducting related statistical tests. For the 5-year-olds, although Shapiro-Wilk's tests of normality were significant for both vocabulary and working memory ($W = .89$, $p = .019$; .92, $p = .062$, respectively), histogram skew and kurtosis were not severe. The number of outliers was unsubstantial: (a) for vocabulary, two 5-year-olds and one 8-year-old whose scores were low average; and (b) for memory, one 5-year-old who scored above average. A significant positive correlation between the language variables was moderate ($r = .36$, $p = .02$), suggesting an acceptable degree of multicollinearity. Concerning variance-covariance homogeneity, Mauchley's test of sphericity was not met, and therefore, Greenhouse-Geisser estimate corrections were applied to the degrees of freedom for the *F*-tests (Field, 2009). The subsequent one-way within-groups multivariate analysis (MANOVA) was not significant, suggesting group equivalence on the measures. The means, standard errors and 95% CIs are reported in Table 4.1.

With respect to the mismatch tallies by participant group, Figure 4.2 through Figure 4.5.

display the proportions of Lexical, Structure and Feature subtotals relative to the Whole Word totals. Figure 4.2 and Figure 4.3 relate respectively, to the entire group of 5TD children (n = 62) and the 5TD sub-group compared longitudinally at 8 to 10 years of age (n = 22). The proportions of mismatch types for the sub-group of 5-year-olds (Figure 4.3), were representative of those for the larger group (Figure 4.2). For 5-year-olds in general, the proportion of lexically related mismatches was particularly small, such that in all practicality, one-quarter of all mismatches concerned word structure, and the remaining three-quarters, features. For typically developing Australian English-speaking children, between ages 5 and 7 years, word structure and feature acquisition for MSWs were also both on-going (James, 2006). Feature mismatches apparently occurred because of challenging sequences in MSWs, and traded off with earlier gains in word structure, particularly with stress patterns. After 7 years of age, matches of MSW components on the prosodic and feature tiers of the nonlinear phonological hierarchy were synchronized.

As shown in Figure 4.4, the sub-total proportions for the 8TD group apparently differed from those of the 5TD. First, with regard to lexically influenced mismatches, there was a slight between-group difference given that there were no tallies for the 8TD. A comparison was therefore made of the Lexical sub-total proportions for both 8TD groups in the larger study, i.e. one for the longitudinal sub-study (Figure 4.4), and the other for the cross-sectional comparison (Figure 4.5). For the 8TD children in the latter group, the proportion of lexically influenced mismatches was small, such that the between-group difference for the 8TD children was in all likelihood unsubstantial. For the 8- to 10-year-olds combined, the Lexical sub-total proportion was not unlike that for the 5-year-olds. Lexical influences, therefore, were apparently minimal and in constant proportion between age 5 years (Figure 4.2 and Figure 4.3) and ages 8 to 10 years (Figures 4.4 and 4.5). Returning to the longitudinal sub-study, in comparison with the 5TD

(Figure 4.2), the proportions of Word Structure and Feature mismatch sub-totals for the 8TD (Figure 4.4) were relatively more balanced, i.e. one-third and two-thirds, respectively.

To assess the MSW mismatch tally differences, a one-way within-groups MANOVA was conducted on the dependent variables: Lexical, Structure and Feature subtotals, in addition to the Total tallies. Meeting of assumptions was evaluated according to the procedures for the language variables, and were considered met, with exceptions as follows. First, the Lexical and Structure subtotals were negatively skewed (W = .64, .91, p = .001, respectively); the skew of the former was accounted for by the restricted range of scores (0 to 2). Second, for both age groups, the variance-covariance matrices revealed particular variables with predictable increasing linear relationships, i.e. Feature in relation to Structure subtotals; and Total tallies in relation to Structure or Feature subtotals. These relationships resulted from the compounding of deletion or insertion tallies across the nonlinear Structure and Feature phonological components, which also contributed to the Total. Third, again by virtue of the theoretical relationship of the subtotals, multicollinearity was possible, confirmed by significant moderate to large positive correlations among the subtotals, and between the Total and subtotals (see Table 4.2). The latter strong correlations suggested the redundancy of the Total tallies, which was therefore removed from further analyses. Despite their strong correlation, the Structure and Feature subtotals were retained as separate variables because of their theoretical importance in the nonlinear phonological model. Finally, Mauchley's test of sphericity of the variance-covariances was not met, and consequently, Greenhouse-Geisser estimate corrections were applied to the degrees of freedom for the *F-tests* in the subsequent analyses (Field, 2009).

The results of the repeated measures MANOVA for the combined dependent variables, i.e. Lexical, Structure and Feature subtotals, indicated a significant within-groups difference, λ =

33.29, p =.001, η^2_p = .84 (a large effect size). For the follow-up univariate analyses, at the adjusted probability level (p < .0125) for the number of tests, all differences were significant, $F(1, 21)$ = 12.58 p = .002; 11.57, p = .003; 79.54 p = .001; η^2_p = .38, .36, .79, respectively. Moderate effect sizes were noted for Lexical and Structure subtotals, whereas the effect size for the Feature subtotal was large. The means, standard errors and 95% CIs are reported in Table 4.3. Examination of the means revealed that the 5-year-olds had consistently more Lexical, Structure and Feature mismatches, and therefore, a higher Whole Word Total, than the 8- to 10-year-olds.

To further explore the relationships among the dependent subtotal variables, a discriminant analysis was conducted. The Eigenvalue revealed one discriminant function that accounted for 100% of the variance, canonical R^2 = .86 (large effect size), and significantly differentiated the groups, λ = .26, χ^2 (3) = 53.91, p = .001. Concerning the loadings of the variables, the Feature subtotal loaded twice as much on the function as either the Structure or Lexical subtotals (r = .78, .27, .35, respectively). A follow-up classification analysis suggested that the function correctly classified 91.0% of the children in their original groupings, 86.4% of the 5-year-olds (19 of n = 22) and 95.5% of the 8- to 10-year-olds (21 of n = 22), implying the validity of the MSW whole word measure.

4.4.2 Cross-sectional Comparison: 8- to 10-year-olds with and without PPD

For the age-matched comparison (N = 24), the age range of the children with PPD was 8;5 to 10;10, M(sd) = 9.38 (9.64 months), and of the 8- to 10-year-old TD children, 7;8 to 10;10, M(sd) = 9.35 (11.72 months). The groups were first evaluated for equivalence on the language-related variables, on which all children scored within the average range. Assumptions for normality, linearity, univariate and multivariate outliers, homogeneity of variance-covariance matrices, and multicollinearity were examined by visually inspecting histograms, and Q-Q, box, and scatter plots, in addition to conducting related statistical tests. In general, there were no serious violations; however, the test of normality for the *Number Repetition Backward* subtest was significant ($W = .914$, $p = .042$,), apparently because most scores were within the average to low average range. A non-significant positive correlation between the language variables was small, suggesting an acceptable degree of multicollinearity. The one-way between-groups MANOVA was without significance, indicating group equivalence on the language-related variables. The means, standard errors and 95% CIs are reported in Table 4.4.

With respect to the mismatch tally comparisons for the 8TD and 8PPD groups, Figure 4.5 and Figure 4.6 respectively, display the proportions of Lexical, Structure and Feature subtotals relative to the Whole Word total. For these groups, mismatch proportions for the Lexical sub-totals were comparable, and in addition, the proportion for the 8PPD was analogous to that for the 5-year-olds (Figure 4.2 and Figure 4.3). Similar to the 8TD group (see section 4.4.1), in comparison with the 5TD, the proportion of lexical mismatches for the 8PPD (Figure 4.2) was comparable (Figure 4.3). Concerning children with PPD, however, longitudinal information would be needed about lexical mismatch frequencies between age 5 years and ages 8 to 10 years, in order to make inferences about developmental stability or variation. Turning to the Word Structure and Feature mismatch sub-totals, in comparison with the relative proportions for the 5TD, those for the 8TD were more balanced. There was therefore a suggestion of the expected developmental progression in segmental feature acquisition for MSWs (James, 2006).

To assess the MSW mismatch tally differences, a one-way between-groups MANOVA was conducted on the dependent variables: Lexical, Structure and Feature subtotals, in addition to Total tallies. Meeting of assumptions was evaluated according to the procedures for the language variables, and considered met with the following exceptions. First, examination of the variance-covariance matrix revealed particular variables with predictable increasing linear relationships, i.e. Structure to Feature subtotals, and Total to Structure and Feature subtotals. These relationships resulted from compounding of deletion or insertion tallies across the nonlinear Structure and Feature components, which also contributed to the Total. Second, again by virtue of the theoretical relationship of the Structure and Feature subtotals, multicollinearity was possible, confirmed by significant, strong positive correlations with the Total to Structure and Feature subtotals (see Table 4.5). A significant positive correlation of the Structure and Feature tallies was moderate, whereas the nonsignificant negative correlations of the Lexical subtotal with all other variables was small. Because the Total tally was the sum of the sub-totals, the correlations predictably indicated its redundancy, and the subsequent removal from further analyses. In spite of the small correlations of the Lexical subtotal, the variable was retained to deliberate theory verification. Finally, Levene's test of homogeneity of error variances was significant for the Feature subtotal, $F(1, 22) = 6.65$, $p = .017$; therefore, the p value for the follow-up univariate tests was reduced to .01 (Field, 2009).

In the main analysis, Wilk's Lambda indicated a significant between-groups difference on the combined dependent variables, $\lambda = .30$, $F(3, 20) =15.71$, $p = .001$, with a large effect size, $\eta^2_p = .70$. At the adjusted probability level ($p < .01$) for the follow-up univariate ANOVAs, the Lexical subtotal difference was without significance, whereas the Structure and Feature subtotals were significant with large effect sizes, $F(1, 22) = 16.82$ ($\eta^2_p = .433$), and 46.52 ($\eta^2_p = .68$), respectively. The means, standard errors and 95% CIs are reported in Table 4.6. Examination of the means showed that the children with PPD had consistently more Structure and Feature mismatch subtotals, and therefore, a higher Whole Word Total, than the age-matched group.

To further explore the relationships among the dependent subtotal variables, a discriminant analysis was conducted. The Eigenvalue revealed one discriminant function that accounted for 100% of the variance, canonical R^2 = .70 (large effect size), and significantly differentiated the groups, λ = .30, χ^2 (3) = 24.82, p = .001. Concerning the loadings of the variables, the Feature subtotal loaded twice as much on the function as the Structure subtotal (r = .95, .57, respectively), whereas the Lexical subtotal loaded very little (r = -.048). A follow-up classification analysis suggested that the function correctly classified 95.8% of participants in their original groupings, 100% of those with PPD and 91.7% of the TD 8-year-olds (i.e. one TD 8-year-old was erroneously classified in the group with PPD). In general, the results of the sub-study indicated that the Lexical subtotal was less relevant to the MSW phonological production of these older children, in comparison with the longitudinal sample that included younger participants (i.e. 5-year-olds).

4.4.3 Cross-sectional Comparison: 8- to 10-year-olds with PPD and 5-year-olds with Typical Phonological Development

The groups for the speech-matched comparison were first evaluated for equivalence on the language-related variables. The age range of the children with PPD was 8;5 to 10;10, M(sd) = 9.38 (9.64 months), and for the TD 5-year-olds, 5;3 to 6;0, M(sd) = 5.64 (3.08 months). All children scored within the average range, with the exception of three children on the *Number Repetition* subtests: two 5-year-olds who scored above and mildly below average, respectively, and one child with PPD who scored moderately below average. Assumptions for normality, linearity, univariate and multivariate outliers, homogeneity of variance-covariance matrices, and multicollinearity were examined by visually inspecting histograms, and Q-Q, box, and scatter plots, in addition to conducting related statistical tests, with no serious violations noted. Although the test of normality for the memory subtest was significant ($W = .914$, $p = .042$,), the histogram displayed no severe skew or kurtosis. A significant positive correlation between the language and memory measures was moderate ($r = .357$, $p = .02$,), suggesting an acceptable degree of multicollinearity. Subsequently, the one-way between-groups MANOVA was without significance, suggesting group equivalence on the language-related measures. The means, standard errors and 95% CIs are reported in Table 4.7.

A one-way between-groups MANOVA was also conducted to test the difference on the MSW dependent variables: Lexical, Structure and Feature subtotals, in addition to the Total tallies. Meeting of assumptions was evaluated according to the procedures for the language variables, and considered met with the following exceptions. First, as observed in the reported sub-study of 8- to 10-year-olds with and without PPD, examination of the variance-covariance matrix revealed particular variables with predictable increasing linear relationships, i.e. Total to Feature subtotals. These relationships resulted from compounding of deletions or insertions across the Structure and Feature tallies, which also contributed to the Total. The plot of the Total to Structure subtotals, however, was random. Second, again by virtue of the theoretical relationship of the subtotals to the total, multicollinearity was possible, confirmed by significant positive correlations between all of the MSW tally measures (see Table 4.8): small for the relationships with the Lexical subtotal; and strong for the Total to Structure and Feature subtotals, and for the Structure to Feature subtotals. Overall, the correlations indicated the redundancy of the Total tally, which was removed from further analyses. Levene’s test of homogeneity of error variances was without significance for any variable.

The main analysis was conducted using Type III Sums of Squares for the unsystematically unbalanced design, i.e. the smaller sample of children with PPD resulted from its low incidence relative to the population of 8-to 10-year olds as opposed to withdrawal for any systematic reasons. Wilk’s Lambda for the between-groups difference on the combined dependent variables was not significant, so the data were not tested further. The means, standard errors and 95% CIs are reported in Table 4.9. The result suggested that the MSW phonological production of the children with PPD was similar to that of the 5-year-olds with typical phonological development. Analogous to the longitudinal sub-study described in an earlier section of this report, for the Lexical subtotal, the significant correlations with the remaining MSW variables indicated its greater relevance to a sample that included younger children (i.e. 5-year-olds).

To summarize, in the current sub-studies, there were four main findings concerning the MSW phonological production of school-age children. First, for TD children, the frequencies of MSW Lexical, Structure and Feature mismatches decreased significantly between ages 5 and 8 years. Second, for children between 8 and 10 years of age, those with PPD demonstrated significantly more Structure and Feature mismatches than an age-matched TD group. Third, the frequencies of MSW Lexical, Structure and Feature mismatches produced by 8- to 10-year-olds with PPD were equivalent to those of TD 5-year-olds. Fourth, a MSW whole word tally measure derived from the sum of tallies for lexical influences (particularly for younger children), and word structure and segmental feature mismatches, appeared promising in terms of validity and reliability. The following section describes the mismatch patterns in detail as observed in the child productions by word types.

4.5 Results and Discussion for Phonological Mismatch Patterns

Because of the potential for complex interactions among MSW components, a vast number of mismatch patterns was possible, i.e. word length, feet, stress pattern, syllable structure and segmental sequences. This section therefore first clarifies the presentation organization of the MSW types and outlines the overarching findings, before describing the observed patterns in detail.

4.5.1 Data Organization for Phonological Mismatch Patterns

First, in order to better draw comparisons among observed phonological patterns, the 20 MSWs selected from *CAPES* were organized into six groups (4.5.2-4.5.7), keeping in mind James' (2006) findings concerning the particular vulnerabilities of non-prominent feet and syllables. Grouping decisions, therefore, primarily considered stress pattern in combination with length by feet and syllables, and in addition, whether the word contained a tap/flap. Particular attention was paid to foot and syllable prominence, i.e. right or left, and primary, secondary or weak, respectively. Position of weak syllables was also taken into account, i.e. word-initial (WI), within-word (WW) or word-final (WF), and in addition, the presence of WF weak syllable sequences. With respect to the current study, word length consisted of one foot with two or three syllables, or two feet with four or five syllables. One-footed words were organized into three groups: *weak Strong (wS), weak Strong weak (wSw),* and *Strong weak weak (Sww)*. Words of two feet were few, and fit into two groups: (a) four-syllable words with *{Strong/secondary} weak*

Strong weak (swSw) stress (with no weak syllable sequences); and (b) four- and five-syllable words, containing a WF sequence of two weak syllables, i.e. *weak Strong weak weak (wSww), Strong secondary weak weak (Ssww)*, and *secondary weak Strong weak weak (swSww)*. Words comprising group (a) had trochaic feet, i.e. prominence on the first syllable, the most common pattern for words of English (Shockey, 2003). All words comprising group (b) had WF weak syllable sequences in the second foot, even if stress differed in the first foot, i.e. started with a weak or prominent syllable. The presence of a weak syllable sequence was considered an important distinguishing factor because James (2006) reported the particular vulnerability of weak within word syllables. A sixth grouping consisted of words containing taps from the above lists. For the current study, a strict view was taken of the timing demands of tap production, a result of the *ballistic* nature (Shockey, 2003), or *speed and trajectory* (Ladefoged, 2006) of taps. The frequency of tap mismatches, within the overall timing demands of MSW length and stress pattern, was therefore of interest.

In the following section, results are reported according to the MSW word-type groups delineated above. Within each section, potential developmental constraints and mismatch patterns are first described, followed by a reporting of results common across and unique to participant groups. Patterns are described in terms of progression from relevant lexical effects to successive tiers of the phonological hierarchy. Lexical influences were difficult to predict, firstly, because syllable components of MSWs overlap particularly with those of monosyllabic words, for which broad word neighborhoods exist. Secondly, parallel processing and connectionist theories suggest that for selection, lexical and phonological components interact, such that the relative influences could not be set apart. Component interactions on the same or different tiers could result in shifts of stress prominence in any word, in addition to insertion, deletion or assimilatory patterns concerning word structure and segmental features. With regard to vowel features, in weak syllables, target schwa and the lax vowel variant, /ɪ/, were considered equivalent.

As noted above, mismatch patterns common to participant groups are first described within each section, from fewest to most, and subsequently, patterns unique to any participant group. Data from children with typical phonological development, therefore, are usually described first: 8- to 10-year-olds (8TD), longitudinal and cross-sectional groups jointly, (n = 22; n = 12, respectively) and then 5-year-olds (5TD; n = 62); finally, the data from the 8- to 10-year-olds with PPD (8PPD; n = 12) are reported. Overall, for same-age groups with typical phonological development, phonological patterns were similar; i.e. for 8TD, the groups for the longitudinal, and cross-sectional age-matched comparisons; and for 5TD, the cross-sectional large group and longitudinal sub-group. The 8TD showed few mismatch patterns, the 5TD, more and the 8PPD group, the most. Patterns unique to each of the 5TD and 8PPD were observed, but once again, more often for the 8PPD. With respect to segments, all groups demonstrated at least some developmental place mismatches for sibilants (s, z, ʃ, tʃ, dʒ) and rhotics, (ɹ, ɚ, ɝ). Sibilants were either (inter)dentalized or slightly backed, for both 5TD and 8PPD groups. Sibilant mismatches were evident across all of the words for the 5TD, but for the 8PPD, sibilant place in four of ten words was accurate: for /s/ in *hospital, mosquito,* and *hippopotamus*, and for /ʃ/ in *invitation*. Notably, among the words with accurate sibilants, none contained affricates. Place deletions in rhotics were frequent for both 5TD and 8PPD, irrespective of position in the onset or rime. For the 5TD, rhotic mismatches were observed across the words; for the 8PPD, however, schwar was accurate in two of ten words, *watermelon* and *thermometer.* Such developmental mismatches were included in the tallies reported in section 4.4 and will not be discussed in detail here, unless relevant to phonological interactions within the words.

The section below presents details about patterns other than the developmental place mismatches in order by word group. Reference to patterns for the 8TD group is reserved for exceptions to otherwise negligible mismatch frequencies. The description of patterns considers the interactions of lexical influences with the various word components of the nonlinear phonological hierarchy (Figure 1.1). The phonological tiers were the bases of the mismatch tally sub-totals, Word Structure and Features. The prosodic tiers related to word structure included: feet and stress patterns; syllable structure components of onset and rime, in addition to the rime components of nucleus and coda; and moras on the timing tier. Features were on the manner, place and laryngeal tiers, and those considered for tallies are shown in Table 1.2. In the following sections, vulnerabilities for each word type are presented, followed by results observed. Enumerated lists of key mismatched pronunciations are included with each description. (Alternative adult pronunciations are noted using{ , }.)

4.5.2 Group 1: *wS*

Words: *balloons* /bəˈlũːnz/, *giraffe* /ʤəˈɹæːf/ or /ʤɚˈæːf/, *guitar* /gəˈtʰɑːɹ/, *explodes* /ˌɪkˈsploʊdz/ or /ˌɛkˈsploʊdz/

The vulnerabilities of this disyllabic iambic word group concerned the interaction of the WI weak syllable with specific consonant sequences that might surface as weak syllable deletion in either of two forms. First, concerning *balloons* and *giraffe,* schwa deletion would result in an allowable English onset cluster (i.e. [bl] or [ʤɹ], where the affricate is a possible realization of

/d/ in /dɹ/), but not for *guitar* or *explodes* (*[gt] or *[kspl]). Thus, the latter word pair might show both consonant and vowel deletion. Alternatively, segment insertion could simplify challenging manner and/or place sequences, e.g. [kəspl], with metathesis of the schwa. Words with more later-acquired fricatives or affricates (*explodes, giraffe*) might show more feature mismatch patterns.

Turning to the results, for the 5TD and 8PPD groups, productions of the word *balloons* /bəˈlũːnz/ were the most matched. Starting at the level of the foot, syllable deletions or insertions in any word were solely from the 5TD group, and specifically because schwa either deleted or was inserted. As predicted for *balloons* and *giraffe*, deletion of the word-initial weak syllable related solely to schwa, and not consonant onset deletion, such that the onset of the resultant monosyllable was realized as a consonant cluster. For *balloons*, schwa deletion occurred just twice (2/62), whereas there were 13 schwa deletions for *giraffe*. Apparently, frequency of schwa deletion was related to some extent to early or late acquisition in development, e.g. /b/ and /l/ acquired earlier than /dʒ/ and /ɹ/. Syllable deletion simplified consonant sequences in *guitar* and *explodes*. For *guitar* > [tʰɑːɹ], a single syllable deletion (1/62) was possibly a result of a high-ranked Not[Dorsal_Coronal] sequence constraint. Turning to the examples below, for *explodes* > [sploʊdz], syllable deletion simplified the /kspl/ coda-onset sequence, as did syllable deletion with insertion in 1(a) below. Five of the 5TD children (5/62) deleted the WI unstressed syllable whereas a single 5TD child (1/62) inserted schwa with co-occurring metathesis as in 1(b). Schwa inserted in *explodes* also suggested lexical influences, e.g. from the onomatopoeia of *kablam,* or *kapow*, allowing /k/ to be produced in a preferred prosodic position, the onset.

(1) *explodes*	ˌ{ɛ, ə, ɪ}kˈsploʊdz	a) spəˈloʊdz	(5TD)
		b) kəspəˈloʊdz	(5TD)
		c) ˌɛθˈploʊdð	(5TD)
		d) ˌɛkˈploʊð	(5TD)
		e) ˌɛtsˈploʊd	(5TD)

A comparison of 1(a) (1/62) and 1(b) (1/62) draws attention to the other potential complexities of language processing for MSW production. In 1(b) WI syllable deletion might have co-occurred with schwa insertion in the /spl/ cluster, or alternatively, at the timing unit and/or segmental feature level, deletion of /k/ might have co-occurred with metathesis of [sp]. The two explanations imply that different constraints are foremost in the phonological system. In the first, a weight constraint that disallows heavy (VC) unstressed syllables might be prioritized, and in the second, a coda-onset manner constraint for [-continuant] [+continuant] is possibly more relevant.

Further to presence of consonant timing units, the coda-onset sequence /kspl/ of *explodes* was also simplified once by /k/ deletion (without syllable deletion) (1c, 1/62), and again by /s/ deletion (1d, 1/62). For 1(d), the WF cluster /dz/ was also simplified, by deleting /d/, but in 1(e), /z/ failed to appear in spite of the [-continuant] [+continuant] coda sequence in the first syllable.

For the latter, the morphemic status of WF /z/ potentially was influential, but 1(d) and 1(e) suggest trade-offs as to which sibilant, /s/ or /z/, was produced. In *giraffe*, in contrast, deletion of WF /f/ was unique to one 8PPD child (1/12).

One pattern of consonant deletion, and one substitution pattern were possibly related in terms of lexical influences. For *guitar*, WI [g] was deleted (2a below) by one of the 5TD (1/62), whereas one each from the 5TD and 8PPD (1/62) groups substituted [ð] for the same segment (2b).

(2) *guitar*	ɡ{ə/ɪ}ˈtʰɑːɹ	a) əˈtʰɑːɹ	(5TD)
		b) ðəˈtʰɑːɹ	(5TD, 8PPD)

These examples may show the influence of morphosyntax on phonology, the articles *a* and *the* represented in the schwa and [ðə]. Furthermore, in relation to the pattern of word-initial syllable deletion above, the possible influence of morphosyntax suggests variable resolutions of the same high-ranked Not[Dorsal_Coronal] feature constraint. Because the production in 2(b) also occurred for one PPD child (1/12), for whom the word was presumably more familiar, the likelihood was greater of a phonology-morphosyntactic interaction. Such interactions might be predicted by parallel processing theories.

Consonant insertion was another pattern seen for *guitar*, for each of two 5TD children (1/62). In 3(a) and 3(b) below, the inserted [+nasal] in the weak syllable rime duplicated the coda timing unit of the prominent syllable, and therefore, the syllables were equally weighted. In comparison with duplication of rhotic features as in 3(c) (5TD; 1/62), in 3(a) and 3(b), nasal insertion might have been a coalescence of vowel and/or rhotic sonorance plus [-continuant] from /t/.

(3) *guitar*	g{ə, ɪ}ˈtʰɑːɹ	a) gənˈtʰɑːɹ	(5TD)
		b) kʰɪnˈtʰɑɹ	(5TD)
		c) kʰɚˈtʰɑɹ̥	(5TD)
		d) kəˈtʰɑɹ	(5TD, 8PPD)
		e) {t, d}əˈtʰɑɹ	(5TD, 8PPD)

Moving to the level of features, for *guitar* for the 5TD children, [g] showed [-voice] (3b through 3d) and/or [Coronal] harmony with [t] (3e), with or without the co-occurrence of other patterns. The productions 3(d) and 3(e) were also observed in the data for the 8PPD children, 1/12 and 3/12, respectively. Devoicing was most common for the 5TD (20/62), occurring just once (3d) for the 8PPD (1/12). Coronal place harmony (3e) was more frequent for 8PPD (4/12) than 5TD (2/62) children. The 8PPD only produced [d] for /g/, maintaining the [+voice] feature, whereas Coronal place assimilation was with and without laryngeal harmony for the two 5TD. For the 8PPD, [Coronal] harmony for /g/ was apparently specific to component interactions of *guitar*, because velars (k, g) were highly accurate in the remainder of the data set.

Further to individual consonant features, mismatches were present for 5TD, with more for 8PPD. Additional examples for *giraffe* and *explodes* follow.

(4) *giraffe* /ʤəˈɹæːf/ a) dəˈɹæf (5TD, 8PPD)

b) ɡəˈɹæf (5TD)

For *giraffe*, deaffrication ([-continuant]) in 4(a), /dʒ/ > [d], was proportionately more frequent for the 8PPD (3/12) than the 5TD (7/62). For one 8PPD child, deaffricated /dʒ/ was aspirated, evidence that [+continuant] might have been represented underlyingly; on the surface, however, [-continuant] [+continuant] were not yet homorganic (produced at the same place of articulation). In 4(b), for two 5TD children (2/62), /dʒ/ surfaced as [Dorsal] [g] rather than [Coronal], in spite of place accuracy in *magician* [məˈʤɪʃn̩]. One possibility was of priming interference because *giraffe* immediately followed *guitar* in the elicitation. Another possibility, however, was that features of /dʒ/ interacted with prosodic position in terms of syllable prominence, unstressed for *giraffe* and stressed for *magician*. That is, feature accuracy of /dʒ/ was promoted by the preferred prosodic position in the stressed syllable. Finally, although slight dentalization of /dʒ/ was expected as a developmental place mismatch, frequencies relative to age were unexpected, i.e. higher for the 8PPD (6/12) than the 5TD (4/62).

Sibilant dentalization was also relevant to *explodes*. In contrast to dentalization of /dʒ/ as described in the paragraph above, relative frequencies of dentalized or interdentalized sibilants /s, z/ were as anticipated, i.e. higher for the 5TD (37/62) than the (8PPD, 2/12). Occasionally, for the 5TD, sibilants in *explodes* were lateralized, possibly in response to either anticipatory (5a) or regressive (5b) influence of the /l/. Because the non-lateralized sibilant was dentalized in both cases, there was the likelihood that segmental feature licensing was not completed in the linear order corresponding to the surface string.

(5) *explodes*	ˌ{ɛ, ə, ɪ}kˈsploʊdz	a) ˌɛksˈˈploʊz̥	(5TD)
		b) ˌkʰərs̥pəˈləʊzˡ	(5TD)

With respect to vowels, solely the 8PPD showed mismatches, all of which pertained to *explodes*.

(6) *explodes*	ˌ{ɛ, ə, ɪ}kˈsploʊdz	a) iˌsploʊdz	(8PPD)
		b) ˌɪkˈsplɑʊts	(8PPD)

In 6(a), the WI lax vowel in the unstressed syllable was [+tense], [i] > [ɪ] (1/12); and in 6(b), the diphthong was lowered in the prominent syllable (1/12), /oʊ/ > [ɑʊ], possibly primed by the recency of *doghouse* [ˌdɑgˈhaʊs], in the *CAPES* elicitation order. In 6(a), insertion of [+tense] was possibly a response to accuracy for syllable weight, i.e. the presence of a timing unit for the deleted /k/. In 6(b), there may have been a trade-off in accuracy of the lax vowel [ɪ] and the diphthong, the cue cost of matching the first unstressed vowel inhibiting accuracy of the later stressed /oʊ/.

Certain mismatches were less expected for these Group 1 words, occurring in *guitar*.

(7) *guitar*	g{ə, ɪ}ˈtʰɑːɹ	a) ɣəˈtʰɑɹ	(5TD)
		b) bəˈkʰɑ̟ːɹ	(5TD)
		c) dɪgˈtʰɑːɹ	(8PPD)

Concerning /g/ for one 5TD participant (1/62), the feature [+continuant] surfaced, /g/ > [ɣ] (7a), a type of lenition in a weak syllable. Further to *guitar,* productions in each of the 5TD and 8PPD groups showed interesting interactions, i.e. [bəˈkʰɑ̣ːɹ] and [dɪgˈtʰɑːɹ], respectively. The [Dorsal] feature of WI /g/ apparently migrated to the stressed syllable onset (7b), and to the unstressed syllable coda (7c), with corresponding matched and mismatched syllable structure respectively. In each case, manner and place of /g/ surfaced somewhere in the word; in syllable onsets, consonant voicing also matched, but place mismatches were various. In [bəˈkʰɑ̣ːɹ], [b] [Labial] may have spread from /ɹ/, and in [dɪgˈtʰɑːɹ], [d] duplicated [Coronal] of [tʰ]. For the latter, the possibility also exists that lexical neighbors of [ɪg] were activated.

In summary, in terms of vulnerabilities for this *wS* word group, word specific patterns occurred as expected, more for the younger 5TD group. Weak syllable deletion was observed and consonants generally showed fewer matches in the unstressed syllable, especially the later-acquired consonants and those in complex sequences. There were some unusual word productions, especially for *guitar* (5TD). Only the 8PPD group showed vowel mismatches. Evidence of inter-word priming in list elicitation also may have affected some pronunciations, again primarily in the younger group.

Expected simplification patterns mostly circumvented the initial weak syllable or the complexity of consonant sequences. To begin, when a weak syllable deleted, if an allowable cluster resulted in the onset of the remaining stressed syllable (*balloons, giraffe*) the relevant onset was realized; otherwise, the onset deleted with the vowel (*guitar*). Conversely, an onset

CCC in a stressed syllable was simplified through schwa insertion, whether or not initial weak syllable deletion occurred (*explodes*). Dorsal_Coronal sequences were simplified through feature assimilation (*guitar*). Finally, trade-offs occurred between consonants deleted from a non-preferred weak syllable coda and the onset and/or coda clusters of the stressed syllable (*explodes*).

In some cases, complexity appeared to increase rather than decrease at various tiers of the phonological hierarchy. For example, as a result of schwa tensing, increases to syllable weight and the number of vowel features presumably allowed more time for the weak syllable to be realized. Harmony of syllable weight interacting with segmental features, and possibly lexical influences, increased the complexity of syllable structure as result of coda insertion: [kʰɪnˈtʰɑɹ], [dɪɡˈtʰɑːɹ]. Stress pattern and word length traded off with syllable weight and realization of all consonants (*explodes*), e.g. coda metathesis in the WI syllable and schwa insertion in the CCC onset in the stressed syllable, with the result of three syllables of *wwS* stress. These interactions lend support to the notion of parallel processing of hierarchical linguistic units because variable multiple component interactions are needed to explain them.

4.5.3 Group 2: *wSw*

Words: *umbrella* /ˌʌ̃mˈbɹɛlə/, *electric* /əˈlɛkˌtɹɪk/, *gorilla* /ɡəˈɹɪlə/ or /ɡɚˈɪlə/, *mosquito* /məˈskiːɾoʊ/, *magician* /məˈʤɪʃn̩/, *computer* /kʰəmˈpʰjuːɾɚ/

Concerning the results for this word group, there were confounds likely for *magician* and *electric,* depending on word familiarity. For each word, there was a highly related word, semantically and phonologically, in the *CAPES* elicitation list, i.e. *musician* and *electricity,* respectively. The potential confound was taken into account when considering the phonological patterns for these words in accordance with the tally procedures described in section 4.3.2. As a result of the possible confound, those productions are not focused on in this section.

Apart from the lexical confounds above, this group of trisyllables had vulnerabilities in common with the iambic disyllables in Group 1, where primary stress was not on the word-initial syllable. Furthermore, for words with a vowel onset, *umbrella* and *electric,* interactions with syllable weight were likely. That is, a single mora of schwa in the onset of *electric* could make it vulnerable to deletion, whereas for *umbrella,* the two moras in /ˌʌ̃m/ could promote syllable realization. For *mosquito, magician*, and *gorilla,* each with C + /ə/ in the WI unstressed syllable, if schwa deleted, consonant mismatches could relate to consonant sequence phonotactics. For *mosquito* and *magician*, /m/ was unlikely to surface because *[mz] and *[mdʒ] are illegal clusters in English; the full syllable could therefore delete. However, if schwa deleted in *gorilla,* /g/ could surface because [gɹ] is a legal cluster. For *computer* or *umbrella*, production of syllabic consonant, [m̩] in the WI syllable was an acceptable variant, and therefore, [kʰm̩ˈpʰjuːɾɚ], and [m̩ˈbɹɛlə] might surface; otherwise, the full syllable might delete.

Various processes could simplify challenging adjacent and non-adjacent consonant feature sequences. For instance, in the coda-onset sequences of *computer* and *umbrella,* /mpj/ and /mbɹ/ respectively, deletion could result from a NotTwice[Labial] place constraint (meaning two labials were adjacent at some level). Also vulnerable to deletion was the combination of the adjacent and non-adjacent consonant sequences in *electric*, i.e. [Dorsal-Coronal_Dorsal] for production of /kt_k/. For *gorilla,* the [+sonorant] sequence, /{ɹ, ɚ}_l/ might be interrupted by insertion of a [-sonorant] segment. In each of the two words, sharing of some key features between segments might foster assimilation of all features of one of the segments to the other.

Starting with the foot tier, for the words *computer* and *mosquito*, 5TD children (2/62) deleted the WI weak CVC syllable, realized respectively as [ˈpʰjuːɾɚ] and [ˈskiːɾoʊ], the latter of which may have been a lexical variant in the children's dialect. Syllable deletion was also related to schwa deletion. In *electric,* word-initial schwa deleted for five of the 5TD (5/62) group only. Schwa deletion, and consequently syllable deletion also affected foot structure in *gorilla,* but predictably, WI [g] was realized in the stressed syllable because schwa deletion left an allowable cluster [gɹ]. Both 5TD and 8PPD groups deleted schwa in *gorilla*, but a lower proportion of 5TD (4/62) than 8PPD (3/12). In general, however, few of the 8PPD deleted schwa in the Group 2 words, as was similarly observed for the Group 1 words.

Continuing with vowels, additional mismatch patterns are described below in terms of potential feature assimilations, some of which related to syllable weight and degree of prominence, again affecting foot structure. Consonant-initial and then vowel-initial words are addressed. Beginning with consonant-initial words, for schwa in *gorilla,* three 8PPD children (3/12) inserted caret [ʌ] or a lengthened [ʊː] (assimilating [Labial] from /ɹ/), and therefore, syllable weight and relative prominence increased. In three words, *computer, magician,* and *mosquito,* only 8PPD children replaced schwa with caret (4/12, 1/12, 1/12, respectively). In *mosquito,* for one 8PPD child (1/12), in the production of [ɔ], schwa may have assimilated [Labial] from [m].

Vowel-initial words were *electric* and *umbrella*. As already mentioned, syllabic [m̩] (8a below) was an acceptable variant in *umbrella*, with the sole production of this syllabic variant in the 8PPD group (1/12). For the 8PPD group, feature insertions in these words were often through patterns of feature spreading or assimilation, related to interactions of schwa or caret (for *electric* and *umbrella,* respectively) with [-consonantal] of /ɹ/ (8b and 9a, 9b below).

(8) *umbrella*	ˌʌ̃mˈbɹɛlə	a) m̩ˈbɹɛlə	(5TD)
		b) ˌɛmˈbɰɛlə	(8PPD)
(9) *electric*	ˌ{ə, ɪ, iː}ˈlɛkˌtɹɪk	a) ˌɔˈlɛktɰɪk	(8PPD)
		b) ˌɛˈlɛktɰɪk	(8PPD)

Full vowels replaced schwa in *electric*: [ɔ] (2/12) (9a); and [ɛ] (1/12), a duplication of the prominent vowel (9b), the latter also observed in *umbrella* (8b) for two children (2/12). One child from each of the 8PPD (1/12) and 5TD (1/62) groups inserted a diphthong in *electric* or *umbrella* (10, 11 below).

(10) *electric*	{ə, ɪ, iː}ˈlɛkˌtɹɪk	ˌeɪːˈlɛktɰɪk	(8PPD)
(11) *umbrella*	ˌʌ̃mˈbɹɛlə	ˌaɪmˈbɹɛlə	(5TD)

Turning to consonants, adjacent consonant sequences showed mismatch patterns, both those at syllable margins (i.e. coda-onset) and consonant clusters in onsets of the stressed syllable. For *computer, umbrella,* and *electric,* the former sequence immediately preceded the latter, whereas for one word, *mosquito*, only one sequence was relevant (ambiguous as to coda-onset or onset status). Consonant clusters /pʰj/ in *computer* and /bɹ/ in *umbrella* were accurate by syllable position; however, for the 62 5TD children, /tɹ/ in *electric* showed deletion of /ɹ/ for five children and metathesis for two children (12a, 12b). The cue cost of realizing the preceding coda /k/ may have resulted in trade-offs for accuracy of the following cluster.

(12) *electric*	{ə, ɪ, i:}ˈlɛkˌtɹɪk	a) ˈlɛkˌtʃəɰk	(5TD)
		b) əˈlʌkˌdɵɹk	(5TD)

In addition to simplifying the onset cluster in the stressed syllable, metathesis of /ɹ/ maximized rime weight, i.e. CVCC with two moras, as opposed to CCVC with one, but keeping the same overall number of timing units. The [tʃ] in 12(a) shows the common pattern of affrication in /tɹ/, however with deletion of the /ɹ/ (Bernhardt & Stemberger, 1998). Finally, in 12(b), the default [d] was a result of [+voice] assimilation from /ɹ/ or other voiced segments in the word.

Other coda-onset simplification patterns were observed, in addition to one of potentially increasing complexity. For the /ktɹ/ sequence in *electric*, a larger proportion of 8PPD (4/12) deleted /k/ than 5TD (2/62). Six of the 8PPD group (6/12) showed glottal replacement of the /k/ in *electric*, a possible trade-off for accurate [tɹ] realization. For one of the 5TD (1/62), syllable complexity apparently increased for *mosquito* through insertion of /k/ (13 below); that is, a coda-onset sequence, [ksˈk], surfaced where none existed in the target.

(13) *mosquito* məˈskiːɾoʊ məkˈsˈkiɾoʊ (5TD)

Coda-onset sequences of CCC, /mpɹ/ and /mbj/, were in *umbrella* and *computer,* respectively. Both sequences were vulnerable to a NotTwice[Labial] feature constraint (14, 15).

(14) *umbrella* ˌʌ̃mˈbɹɛlə ˌʌ̃nˈbɹɛlə (5TD, 8PPD, 8TD)

(15) *computer* kʰəmˈpʰjuːɾɚ gəˈbjutʊ (5TD)

For *umbrella* (14)*,* appearance of [n] was a very common pattern for the 5TD (23/62) and 8PPD (4/12), and occurred frequently in the 8TD children (13/34) as well, possibly to resolve a negative same-place sequence constraint (i.e. three tokens of [Labial]). A repeated [Labial] sequence was apparently still difficult for a few of the 5TD (2/62) and relatively more of the 8PPD (3/12). These children also deleted /m/ in *umbrella*, which in addition, was the case for *computer* (5TD, 7/62; 8PPD, 2/12). For one 5TD child (1/62), feature interactions in *computer* (15) led to deletion of /m/ and to stops, /k/ and /p/, assimilating [+voice].

(16) *mosquito* məˈskiːɾoʊ a) bəˈskiːɾoʊ (8PPD)

b) m̩͜bˈskiːtʰoʊ (8PPD)

The productions in (16) are further examples of consonant mismatch, for /m/ in the initial unstressed syllable of *mosquito*. For two 8PPD children, there was possibly a trade-off between overall consonant accuracy and realization of word structure, shown by degrees of denasalized /m/, assimilated from the consonants in the stressed syllable.

Finally, there were notable interactions applicable to mismatches in the non-adjacent [+sonorant] segments in the consonant timing units of *gorilla.*

(17) *gorilla* {gɚˈɪlə, gəˈɹɪlə} a) gwəˈwɪlə (5TD)

b) gɹɪˈlɪlə (5TD)

c) ˈgɝələ (8PPD)

Productions 17(a) and 17(c) from the 5TD (1/62) group demonstrate trade-offs of syllable structure and segments. The 5TD child who produced 17(a) did not demonstrate acquisition of /ɹ/ in the sample. The /w/ was apparently duplicated in the first syllable. With respect to 17(b), the segment /ɹ/ was in this 5TD child's inventory; migration of /ɹ/ occurred, with duplication of the lateral in the onset of the stressed syllable. Last, in 17(c), for one 8PPD child (1/12) stress shifted.

In summary, although the words in Group 2 have been categorized by their stress pattern, the above observations again demonstrated a multiplicity of interactions in productions of the words. The interactions did involve the word-initial weak syllable as expected, and for both the 5TD and 8PPD, may have related to this syllable alone. A pattern for one word, *umbrella*, was common to all three groups (i.e. including the 8TD), with one-third of each group producing [n] for /m/ in *umbrella*. This pattern may have been a result of articulatory difficulty for a same-place sequence, i.e. [Labial-Labial-Labial], with the necessary articulatory precision still developing at age 8 years. Also common to the 5TD and 8PPD groups, though rare, was diphthongization in non-prominent syllables (*umbrella, electric*).

For the 5TD generally, however, the expected patterns were observed, i.e. of weak syllable deletion (*computer, mosquito, electric)*, and more consonant mismatches for clusters in the onsets of stressed syllables, the latter concerning later-acquired consonants, especially liquid /ɹ/ (*electric*). Liquid sequences were also vulnerable to consonant duplication with or without insertion (*gorilla*). These patterns contrasted with those for the 8PPD. Rather than syllable deletion, through vowel feature insertion, the 8PPD equalized syllable weight (moras) and therefore increased syllable prominence (*electric, umbrella, gorilla*). Consonant deletion was related to sequence constraints at the coda-onset juncture, for [Labial-Labial] and especially for [Dorsal-Coronal] sequences (*electric*). For the latter, glottal replacement also occurred. Finally, other consonant feature mismatches were sometimes for the earliest acquired segments (*mosquito*), indicative of trade offs for accuracy of other word components.

4.5.4 Group 3: *Sww*

Words: *animal* /ˈæ̃ːnəmɫ̩/, *hospital* /ˈhɑːspɪɾɫ̩/, *vegetable* /ˈvɛʤtəbɫ̩/, *skeleton* /ˈskɛlətn̩/

All words in this group were potentially vulnerable because of the weak syllable sequence, and in addition, for *hospital*, tap /ɾ/. The word *hospital* is addressed in section 4.5.7, and will not be discussed further here. For the remainder of the Group 3 words, other components vulnerable to simplification were adjacent and non-adjacent consonant sequences. Adjacent sequences included the: (a) s-cluster in *skeleton*; and (b) coda-onset [+continuant] [-continuant] in *vegetable*. Non-adjacent sequences were (a) [Dorsal_Coronal] [k_t] in *skeleton*; and (b) [Coronal_Labial] for [n_m] in *animal*. For *skeleton* and *animal,* the various sequences of sonorants /m, n, l, ɫ̩/ were also vulnerable to interaction with the word-final weak syllable sequence (James, 2006; Kehoe, 2001).

The results for 5TD and 8PPD concerned mismatches of syllable and timing unit numbers, in addition to feature assignment, as seen in these introductory examples for *vegetable* and *skeleton*.

(18) *vegetable* ˈvɛʤtəbɫ̩ ʊ̆ˈvɛʧdəbɫ̩ (8PPD)

(19) *skeleton* ˈskɛlətn̩, ˈskɛlətʰɪn ɪˈskɛlətn̩ (8PPD)

One 8PPD (1/12) participant demonstrated the only syllable insertions (18, 19 above), a result of initial lax vowel insertion (excrescence). In (18), [Labial] spread from /v/ to the inserted vowel slot. In (19), the inserted [ɪ] may have resulted from spread of [+high] from [Coronal] of /s/. Given that /s/ and /k/ are [-voice], the feature [+voice] may have spread at least in part from adjacent vowels on the vowel tier to the inserted vowel, also a possibility in (18). At higher levels of the nonlinear hierarchy, therefore, excrescence apparently increased word complexity in terms of length by syllables (four as opposed to three) and thus by feet (two as opposed to one), and in terms of non-preferred word-initial *weak Strong* stress. Simplification, however, was potentially in terms of additional time on the vowel tier, i.e. off-setting anticipated consonant sequencing complexity within unstressed syllables because of affricate-stop, and syllabic consonants {ɫ̩, n} with or without tap.

Deletion and insertion were observed on several phonological tiers for *skeleton* and also *animal*. Syllable deletion was a result of within-word schwa deletion. For the 5TD, schwa deleted in both words but for the 8PPD, only in *skeleton*. A particular 5TD production of *animal*, i.e. [ˈæ̃ːmdɫ̩], possibly resulted from metathesis of /n/ to the coda (avoiding a sequence constraint for [Coronal_Labial_Coronal], /n_m_ɫ̩/); if a nasal-lateral [nɫ̩] sequence was disallowed, a Stop-Stop, [Labial-Coronal], [md] sequence appeared. Timing unit insertions were relevant to syllabic consonants and surfaced as schwa plus consonant for both groups of children, but more frequently for the 8PPD. At the level of consonant features, both the 5TD and 8PPD commonly produced /ɫ̩/ as [+back] vowels, i.e. schwa, /ɔ/, or /ɑ/, more predictable at 5 years than at 8 years of age developmentally (McLeod, 2009).

Returning to *vegetable* at the level of consonants, there were several mismatches which differed for the groups in terms of type and frequency, both for individual consonants and consonant sequences. First, for the 8PPD, the WI /v/ of *vegetable* showed three mismatch patterns: [-consonantal] [w] (1/12), [-continuant] [b] (1/12), and [-labiodental] /β/ (1/12), whereas mismatches for the 5TD (6/62) were solely [-continuant] [b].

Concerning the medial coda-onset sequence of *vegetable*, /ʤt/, an alternative voicing variant /tʃt/ was noted in the study. Mismatches for the 8PPD primarily concerned the voicing feature sequence. The same sequence was relevant to mismatches for the 5TD, as were segmental insertions. Beginning with the acceptable variants, ten 5TD children (10/62) produced /ʤt/, and therefore the [+voice] [-voice] sequence, whereas /ʤt/ was not observed for the 8PPD (0/12). Instead, the variant /tʃt/, a sequence without voicing alternation, occurred twice for the

8PPD (2/12), but only five times for the 5TD (5/62). Mismatched voicing patterns were various and more frequent for the 8PPD than for the 5TD, with or without deaffrication. These included: (a) [-voice] [+voice], [tʃd] or [tʃɾ], 8PPD (6/12), 5TD (6/62); (b) [+voice] [+voice] [ʤd], 8PPD (3/12), 5TD (10/62); and (c) [-voice] [-voice] with deaffrication in [ˈvɛʃ̺təbɫ̩], 8PPD (2/12), 5TD (3/62). These patterns represented the other half of productions for the 8PPD, and about one-third of those for the 5TD. For the 5TD, patterns of insertion resolved [ʤt] sequencing constraints (20 below). At the foot level, one child (1/62) increased word length by syllables, and thus altered stress through schwa insertion, either between (20a) or preceding (20b) deaffricated [ʤ] (i.e. [ʃ, ʒ]) and [t]. In 20(b), [+syllabic] from /ɫ̩/ may have migrated to the [ʤ] timing unit in addition.

(20) *vegetable*	ˈvɛʤtəbɫ̩	a) ˈvɛʒətəbəl	(5TD)
		b) ˈvɛə̆͜ʒdəbə	(5TD)

Other children resolved constraints on [ʤt] at the timing unit level (21 below), and equalized the weight of the perceptually salient first and last syllables. For nine children (9/62), an inserted nasal /n/ (21a) preceded the affricate, suggesting the importance of sonorant influences,

(21) *vegetable*	ˈvɛʤtəbɫ̩	a) ˈvɛndʒ̺təbɫ̩	(5TD)
		b) ˈvɛʔʃtəbəɫ̩	(5TD)
		c) ˈbɛʒ̺dʒəbə̪ɫ̩	(5TD)
		d) ˈvɛkʃ̩əbɫ̩	(8PPD)

i.e.: [+sonorant] harmony with vowels and [ɫ̩]; [+anterior] assimilated from [ɫ̩] in combination with /t/. Maintaining activation of the features of sonorants, may have made more time available to organize the anticipated [ʤt] sequence. Glottal insertion (21b) possibly resolved a constraint on licensing [-continuant] [+continuant] in one timing unit, i.e. for /dʒ/, whereas in 21(c), metathesis of the timing unit for /t/ allowed feature harmony through alternating [-continuant] [+continuant] across segments on the consonant tier. The presence of the [ʒ̤dʒ] sequence suggested the simultaneous processing of [-continuant] [+continuant]. For one 8PPD child, metathesis of /t/ surfaced in an apparently unusual production (21d). In addition to deaffrication of /ʤ/, [Dorsal] from [ɫ̩] was apparently duplicated in the metathesized timing unit for /t/.

To summarize, the challenging aspects of the productions were again as predicted, i.e. within-word weak syllables were vulnerable because of interactions with consonant sequences. Syllable deletion as a result of schwa deletion occurred more often for the 5TD (*animal, skeleton*) than the 8PPD (*skeleton*). One unexpected pattern was the initial vowel excrescence for one child with PPD, possibly a unique strategy in which word length by syllables traded off with faithfulness to segments. The word, *vegetable,* was particularly challenging in terms of feature sequences, including [Labial-Labial +labiodental] for the 8PPD, and more so, [-continuant] [+continuant] for both groups. Resolutions to constraints on the latter sequence were highly variable for the 5TD, and were mostly relevant to /dʒt/, i.e. insertions at the foot and timing unit levels that interacted with manner and place sequencing. For the 8PPD, mismatches also primarily concerned /dʒt/ but the majority of interactions were of voicing with [-continuant]

[+continuant], such that [+voice] [-voice] did not surface. The myriad of strategies for production of /dʒt/ that are difficult to explain using a linear processing framework, are clarified in terms of parallel interactive processing of multiple hierarchical phonological units.

4.5.5 Group 4: ***Swsw, swSw***

Words: *alligator* /ˈæːləˌgeɪɾɚ/, *watermelon* /ˈwɑːɾɚˌmɛln̩/, *invitation* /ˌĩnvəˈtʰeɪʃn̩/

Word structure and feature accuracy (with the exception of developmental substitutions for late-acquired consonants) was considered more likely for these two-footed words. The foot was balanced for syllable number, i.e. two, and stress prominence was on the preferred first syllable (whether primary or secondary). In comparison with *alligator,* and *watermelon*, however, *invitation* might have been less familiar to the children, such that lexically influenced mismatches might occur (see below). The word *invitation* was also more complex in terms of having secondary rather than primary stress word-initially and because of the adjacent coda-onset consonants, /nv/.

With the exception of schwar-tap sequences, which are discussed in 4.5.7, both 5TD and 8PPD produced the target elements of *alligator* and *watermelon*, i.e. there were no deletions of syllables or segments. For *alligator,* one exception relevant to an 8PPD child was WI insertion of /h/.

Mismatches in the Group 5 words primarily concerned the word *invitation.* Infrequently, the 5TD (3/62) simplified stress in the non-prominent foot through WI syllable deletion or stress equalization (2/62). For two different children in the 8PPD group (2/12), stress shifted to the WI syllable, together with production of [ɛ] > [ˈɛ̃nvəˌtʰeɪʃn̩], or [æ] > [ˈæ̃ːnvəˌtʰeɪʃn̩], as a result of vowel harmony in height features, i.e. [-high], with the steady state of the stressed diphthong /eɪ/. The production of [æː] may also have reflected morphosyntactic influence from the article, *an*.

For one of the 8PPD group (1/12), consonant insertion was related to a constraint on syllable weight. In item (22) below, insertion of [n] in the second syllable increased syllable complexity but harmonized the weight (two moras) in the non-prominent syllables across the word.

(22) *invitation*	ˌĩnvəˈtʰeɪʃn̩	ˌĩnˌvĩnˈtʰeɪʃn̩	(8PPD)

With respect to consonants of *invitation*, the 5TD (56/62) predictably had slight developmental place shifts for /ʃ/, whereas the 8PPD matched this consonant. The remainder of patterns produced by the 5TD and 8PPD affected the first two syllables, and were particularly relevant to [Labial] in the /nv/ sequence.

(23) *invitation*	ˌĩnvəˈtʰeɪʃn̩	a) ˌɛ̃mbəˈtʰeɪʃ̟n̩	(5TD)
		b) ˌĩɱəˈtʰeɪʃ̟ə̃n	(5TD)

In 23(a) above, where /v/ surfaced as /b/ (5TD, 7/62), /n/ assimilated [Labial]. For about one-third of the 8PPD group (5/12), and one-sixth of the 5TD children (11/62), however, one or the other segment was deleted altogether (and therefore its timing unit). When /n/ deleted, /v/ was produced correctly (5TD 5/62; 8PPD, 3/12), possibly because of its preferred prosodic position in the syllable onset. When /v/ deleted, /n/ assimilated [Labial] with or without

[+labiodental] (5TD 6/62; 8PPD, 2/12), suggesting that [Labial] activated in the preferred prosodic position of /v/, perhaps in combination with its activation for /ʃ/, resulted in [m] or [ɱ] (23b; 5TD, 2/62) on the surface.

Overall, the Group 4 words showed minimal patterns and again, a challenging consonant sequence /nv/ was particularly vulnerable in a weak prosodic position at the coda-onset interface of the non-prominent foot. For the 8PPD, stress shifts in interaction with vowel feature insertions in the non-prominent syllable word-initially may have increased the time available for correct sequencing of consonant features. Increased complexity in the structure of one non-prominent syllable (22 above) apparently traded off with the simplicity of moraic harmony across the sequence of non-prominent syllables. Potentially higher activation of a feature relevant to both /v/ and /ʃ/, [Labial], may have increased the likelihood of its realization in either of the first two syllables, with or without consonant deletion. These patterns again support the notion of parallel processing and patterns of activation in multiple levels of the hierarchy.

4.5.6 Group 5: ***wSww, Ssww, swSww***

Words: *thermometer* /θɚˈmɑ̃ːməɾɚ/, *cash register* /ˈkʰæːʃˌɹɛʤɪstɚ/, *hippopotamus* /ˌhɪpəˈpʰɑːɾəməs/

In addition to *invitation,* these three words were possibly among the least familiar to the children, such that lexically influenced mismatches might be more probable. While the phonological components of the second foot were consistent in syllable number and stress, more diverse components in the first foot might augment the interaction variability. Notably, in *cash* *register,* prominent stress occurs on adjacent syllables (primary and secondary), and the children might avoid this stress clash. Other interactions with sequences of later developing fricatives and affricates, /s, ʃ, θ/, and /ʤ/, respectively, with taps and/or rhotics, might increase the frequency and variability of mismatches across hierarchical components. For instance, in *thermometer,* accuracy of early developing /m/, with recurrence in adjacent syllable onsets (i.e. on the syllable onset plane), might be compromised by a NotTwice place constraint. The word *cash register* contains potentially challenging manner and place sequences, i.e. alternations of [+continuant] and [-continuant]; a [Dorsal_Coronal] sequence, /k_ʃ_ʤ_s/; and [-anterior] /ʃ_ɹ_ʤ/ to [+anterior] /st/ to [-anterior] /ɚ/. The vulnerabilities of *hippopotamus* relate more to word structure than features, because of a higher number of syllables but developmentally earlier segments. Specifically, non-prominent components might be deleted, including the first foot, in addition to one or more weak within-word syllables.

In terms of results, for *thermometer,* lexical unfamiliarity, stress and segmental feature mismatches were apparent. A pattern characteristic of solely the 5TD concerned tap production and is discussed in 4.5.7.

Beginning with stress, two of the 8PPD (2/12) altered stress prominence, either through a shift to the WI foot/syllable, or by weak syllable deletion. At the timing unit level, for the 5TD (4/62), productions of [hɚ] or [hə] in the WI weak syllable might have had morphosyntactic influence from the third person feminine possessive pronoun, *her*. But in any case, while [-voice], [+spread glottis] and [+continuant] matched the target /θ/ with [h], oral articulation and [-sonorant] were sacrificed in the initial weak syllable. Continuing with features of WI /θ/, a larger proportion of 5TD (14/62) than 8PPD (1/12), replaced [Coronal] with [Labial, +labiodental], i.e. [f] (an acquired segment with perceptual similarity to the low intensity interdental). The 8PPD, however, substituted other manner or place features that the 5TD did not, e.g. with resulting [s] (loss of [-grooved]; 1/12) or [t] (loss of [+continuant]; 1/12), respectively.

For the majority of words, the 8PPD, and not the 5TD, had unique patterns, but with respect to *cash register*, both groups demonstrated variable patterns across children. Of the MSWs evaluated, *many* of the 8TD and *most* of the 5TD and 8PPD showed mismatches for *cash register*. The numbers of children, with the *Mean* (*standard deviation)* number of tallies was as follows: (a) 8TD: 20/34; Mean = 1.09 (1.80); (b) 5TD: 59/62; Mean = 6.70 (4.8); (c) 8PPD: 11/12, Mean = 5.58 (3.75). A sample of the various productions is provided below.

(24) *cash register* ˈkʰæːʃˌɹɛdʒɪstɚ

a) ˈkʰæʃ̟ˌɰiʃɔʊ (5TD)

b) ˈkʰædəˌɹɛʒdɪʃtɚ (5TD)

c) ˈkʰæθdɚˌɹɛðdʊ (5TD)

d) ˈkʰæʃ̟ˌɹɛʃ̟stʊɾʊ (5TD)

e) ˈkʰæːʃˌɹɛθɹəθstɚ (5TD)

f) ˌkʰæːʃ̟ˈwɛʒ̞əʃ̞̟ʊː (8PPD)

g) ˈtʰæːʃˌɹɛʒɪstɚ (8PPD)

Stress mismatches were present for both groups, but differed as to type. Several of the 8PPD (3/12) shifted prominent stress to the initial syllable in the second foot, e.g.

[ˌkʰæːʃˈɹɛdʒɪstɚ]; for one child (1/12), two syllables received equal stress, the WI syllable, but in addition, the WF weak syllable. Stress mismatches occurred in a small proportion of the 5TD, but unlike the 8PPD, were a result mainly of weak syllable deletion in the second foot (24a; 7/62). Other singular stress mismatch patterns included: a combination of syllable reduplication and insertion (24b; 1/62); a combination of syllable deletion and insertion (24c; 1/62). The groups also differed somewhat with respect to timing unit patterns, i.e. for 8PPD, timing units consistently matched, but the 5TD group showed diverse occasional mismatches. For one child, there was apparent metathesis of the WF syllables (24d; 1/62). In some instances tense vowels surfaced in all syllables (24a; 1/62), or coda insertion in the weak CV syllables reduplicated the CVC structure of the prominent foot (24e; 1/62). In other examples, consonants inserted in the onset of the second foot (*register*) created a cluster, either [dɹ] (1/62) or [ʤɹ] (1/62), reduplicating some or all features of /ʤ/. In contrast, the /st/ cluster in the final syllable was simplified by deletion, affecting /s/ (9/62) more often than /t/ (6/62).

Regarding feature mismatches of *cash register,* considerable interaction was expected, given the number of rhotics, /ɹ, ɚ/ and coronals /ʃ, ʤ, s/ in the word. For the coronals, mismatches related primarily to deaffrication, various combinations of coronal place shifts, and laryngeal assimilations, as was typical across all words. The 5TD typically matched at least one of these segments, usually /ʃ/ (52/62), although the 8PPD tended to show mismatches on more than one coronal fricative/affricate (e.g. 24e, 24f; 6/12). For the 8PPD, mismatches were more frequent for /ʤ/ (8/12) than /s/ (6/12), but nearly equivalent for the 5TD (/ʤ/, 22/62; /s/, 19/62);

for both groups, however, when either was mismatched, there were also mismatches on at least one other sibilant.

A handful of the 8PPD demonstrated feature mismatches that the 5TD did not. For example, for two children (2/12), WI /k/ assimilated coronal from other consonants in the word (24g). Concerning /ɹ/ in the onset of the second foot, in one production (1/12), [h] surfaced, possibly assimilating [-voice] from the preceding /ʃ/ but maintaining [-consonantal] and [+continuant] of /ɹ/; and in another instance, a bilabial fricative [β] (1/12) occurred for /ɹ/.

Turning to results for *hippopotamus*, although nearly all of the 8TD produced the adult target, two did not, both of whom deleted a different weak syllable (2/12): one from the first foot, and the other, from the second foot. Possibly, the entire word was unfamiliar to some children, because an abbreviation, *hippo,* was customary in casual Canadian-English speech. For the 5TD, within-word weak syllable deletions were the most common pattern (8/62), but were restricted to the second foot, and represented the only stress alterations. Two of the 8PPD (2/12), however, shifted stress, either to the WI or WF syllable, without syllable deletion. For one of the 8PPD (1/12), mismatches were at the timing unit level, i.e. the timing unit and features of the WI onset /h/ transposed to the second syllable onset, deleting the features of /p/, i.e. [ˌɪhəˈpʰɑːɾəmɪs].

Major examples of interesting interactions for this word group primarily concerned *cash register* and *hippopotamus.* To conclude this section, two productions of *cash register* from the 5TD children are presented, i.e. [ˈkʰæʃ̟ˌɰɛɾɪtʃʊ], and [ˈkʰæʃ̟ˌɰɛgs̺tʊ]. Apart from the mismatches for rhotics, in the first production, [ˈkʰæʃ̟ˌɰɛɾɪtʃʊ], a short tap, as opposed to a longer stop, {/t/, /d/}, surfaces in the non-prominent foot, arising possibly from metathesis of /dʒ/ and /st/ with deletion of the /s/. The opposing adjacent place and manner sequences were avoided, and open syllables (CV) were thus maintained across the foot. In the second production, the weak syllable within the second foot was deleted, while keeping the syllable onset timing unit, which was transposed into the coda of the preceding syllable. The productions of *cash register* exemplify all of the other patterns shown in the other word groups, but amplified due to its higher complexity. Further discussion concerning the interactions is presented in section 4.6.2.2.

4.5.7 Group 6: Words with Tap /ɾ/

Words: *mosquito* /məˈskiːɾoʊ/, *computer* /kʰəmˈpʰjuːɾɚ/, *hospital* /ˈhɑːspɪɾl̩/, *alligator* /ˈæːləˌgeɪɾɚ/, *watermelon* /ˈwɑːɾɚˌmɛln̩/, *thermometer* /θɚˈmɑ̃ːməɾɚ/, *hippopotamus* /ˌhɪpəˈpʰɑːɾəmɪs/

The following section provides additional examples of segment-structure interactions and elaborates those more fully as the culminating examples. In the current study, production of taps was of interest because of the interaction of their underlying phonology with phonetics. Tap production has been described as *ballistic* in nature (Shockey, 2003), and as having *speed and trajectory* (Ladefoged, 2006), setting them apart from most other consonants. The relative difficulty of taps has also been described in terms of articulatory motion (Derrick, 2011) and phonetic context (Klein & Altman, 2002). Taps were therefore considered likely to interact with the multiple elements of MSWs, and in particular, when preceding syllabic consonants. Thus, a special section is devoted to the production patterns observed.

Adults may use stops for taps on occasion, e.g. in single word productions as in this data set. Production of North American English taps has importance because of their obligatory nature in customary connected speech (Shockey, 2003) and even in some careful speech. Concerning developmental acquisition, Klein and Altman (2002) reported a longitudinal case study of tap production for four children, ages two to five years. Tap (also referred to as flap) production was considered in trochaic words in several contexts: (a) intervocalic, i.e. /ɪɾi/, /_ɾɚ/, and /_ɾɪŋ/; and (b) syllabic consonantal, i.e. lateral release, /ɫ̩/, and nasal release, /n̩/. Klein and Altman suggested that by age five years, no children produced the lateral or nasal release tap allophones, but that the [+voiced] stop [d] was primarily produced for taps.

For the current study, there are various contexts of interest for taps: intervocalic /ɪɾi/ (*mosquito*), and /ɑːɾə/ *hippopotamus*); /_ɾɚ/ (before syllabic schwar), i.e. *computer, alligator, watermelon, thermometer*; and before syllabic /ɫ̩/ and /n̩/, might be pertinent to *hospital* and *skeleton,* respectively. Interactions of taps and syllabic consonants might also involve differing complexities as a result of their hierarchical organization (see Figure 1.1). For example, the timing unit(s) of syllabic consonants might be doubly linked to the nucleus, or alternatively, singly linked to the rime, affecting the relative moraic weight, i.e. two moras as opposed to one, respectively (Bernhardt & Stemberger, 1998). The double linking to the nucleus might result in more patterns for resolving constraints on taps sequenced with syllabic consonants.

Considering the above discussion, a strict view was taken of tap production for this section. In the descriptions that follow, full coronal consonant insertion, i.e. of [d] or [t], is reported as a mismatch pattern, and for brevity, is referred to as tap stopping. The next section describes the mismatch patterns for *hospital* and *thermometer* to show examples of the various patterns that occurred across this word set. (Figures 4.7 and 4.8 show tap stopping patterns for all Group 6 words.)

For each word in Group 6, the most common mismatch pattern was a consonant feature change, i.e. tap stopping, generally to [d], but also to [t]. The mean mismatch proportions (with standard deviation in brackets) were lowest for the 8TD (Mean = 11.76% [7%]), and were highest, and also equivalent for the 8PPD and 5TD groups (23.8% [16.27%]; 24.65 [8%]). Proportions of stopped taps across words were consistently lowest for the 8TD. The largest proportion of stopped taps was produced by the 8PPD in *mosquito, hospital* and *thermometer*, indicative of interaction with consonantal sequence complexity of the words overall. For the 5TD, more stopped taps were produced in *computer, alligator* and *watermelon*, suggesting that tap interacted with schwar, also a later segment in acquisition with articulatory complexity. The 8TD group produced one word without tap stopping, i.e. *alligator*.

With respect to voicing, the younger children demonstrated a preference for [+voice] [d] over [-voice] [t], whereas the tendency was slightly reversed for older children with and without PPD. For the 5TD, [+voice] (75%) was considerably more common than [-voice] (25%), with proportions as reported by Klein and Altman (2002). For the 8TD and 8PPD, [-voice] (55%; 60%) was more frequent than [+voice] (45%; 40%). Within the words, [-voice] almost never occurred in *alligator* (5TD, 1/62), nor in three additional group-specific words, i.e. *watermelon*, *hippopotamus*, and *computer* for the 8TD, 8PPD, and 5TD, respectively. For *hospital* and *thermometer*, frequencies of [-voice] were highest (13/190; 18/90, respectively), and for the children were larger proportionately for the 5TD (6/62; 12/62), and largest for the 8PPD (3/12; 3/12). For the remainder of the words, there were few tokens of [-voice] (from one to three) across groups, but the proportions were nevertheless at least twice as large for the PPD, keeping in mind the small sample. Klein and Altman (2002) suggested that the tendency of young children to produce [+voice] [d] for taps is relevant to place and manner similarities in spite of a timing distinction. For children with PPD, however, the interacting phonological and phonetic components of tap may prohibit its production, and consequently, default insertion of [-voice] [t].

For patterns other than tap stopping, i.e. deletion, nasal substitution and glottal stop insertion, there were also few tokens across groups, and these were mainly word specific. For the 8TD (n = 34) and the 5TD (n = 62), tokens of deletion and nasals were negligible, and also tokens of glottal stop for the latter group. There was no glottal stop insertion for the 8TD. Keeping in mind the small sample of children with PPD (n = 12), each of deletion, nasal substitution and glottal stop insertion was twice as frequent in comparison with the 5TD. Concerning the words, tap deletion was the most common pattern for all groups of children in *hospital* and *watermelon*, but deletion was just as frequent as nasal substitutions were in *hippopotamus*. Furthermore, for the 5TD, tap was nasalized in the remainder of words that contained a [+nasal] weak syllable onset (/m/ for the current study), i.e. *mosquito* and *thermometer*, and in addition, in *hospital*. Glottal stops appeared in *watermelon* for the 5TD and 8PPD, but also in *alligator* for the latter group. The patterns suggested that unless deleted, taps were vulnerable to feature assimilation. For example, the presence of a nasal elsewhere in the word increased the likelihood of tap assimilating nasal, and similarly, the existence of a stop predicted glottal replacement. Patterns for two words are described next as examples for MSWs and taps: *hospital* and *thermometer*.

For *hospital* /ˈhɑːspɪɾl̩/, patterns primarily reflected the weak syllable sequence stress pattern sequence (Sww) in interaction with the consonant sequences: the /sp/ cluster and tap-/l̩/ sequence. Regarding tap stopping for *hospital*, Figure 4.7 illustrates the highest frequency for the 8PPD group (4/12), followed in turn by the 5TD (12/62) and 8TD (5/34). Concerning voicing, the 5TD were divided evenly as to production of [d] or [t], whereas more often, the 8TD (3/4) and 8PPD (3/5) produced [t].

Mismatch patterns also showed interactions related to deletion or stopping of taps (25a and 25b, respectively) and schwa insertion before syllabic /l̩/.

(25) *hospital*	ˈhɑːspɪɾl̩	a) ˈhɑːspɪ{d, t}əl̩	(8TD, 5TD)
		b) ˈhɑːs̺pɪəl̩	(5TD)

For two-thirds of the 5TD (41/62) and a single 8TD (1/34), stopping of taps co-occurred with schwa insertion (25a), thereby adding a vowel timing unit, with features of /l̩/ filling a consonant timing unit. For one 5TD child (1/62), the same timing unit adjustments for syllabic /l̩/ co-occurred with tap deletion (25b), possibly a trade-off for matching the target /sp/ cluster.

The mainly single productions in (26) below exemplify other possible patterns of word component trade-offs, generally involving tap and the /sp/ cluster: i.e. syllable deletion (26a, b) and metathesis (26c, d), and in 26(e), feature interactions.

(26) *hospital*	ˈhɑːspɪɾl̩	a) hɑːsˈp l̩ː	(8PPD)
		b) ˈhɑːs̪{p, b}ɪɫ	(8TD, 5TD)
		c) ˈhɑːθəbl̩	(5TD)
		d) ˈhɑːstəbɤ	(5TD)
		e) ˈhɑːθbɪðəɫ	(5TD)

In 26(a) and 26(b), one of the weak syllables deleted, and no tap was produced. The production in 26(a), from one 8PPD child (1/12), had the only stress shift, which apparently increased stress pattern complexity in terms of a non-preferred iambic foot. Lengthening of the /l̩/ might have been compensatory for syllable deletion. Given that the underlying mora would be added to the moras of syllabic /l̩/, in comparison with the WI syllable, WF syllable weight would have been greater, and thus, primary stress might have been attracted rightward. For 26(b), the 8TD (2/34) produced solely target [p] in the WF syllable onset, whereas one 5TD participant (1/62) produced [b]. Realization of [ɪ] in the WF syllable in combination with tap deletion resolved a tap-syllabic consonant sequence constraint.

In 26(c, d), further interactions with the /sp/ cluster are observed with metathesis occurring, i.e. of /p/ and tap, potentially resolving a [Coronal-Labial] place sequence constraint within the /sp/ cluster or a [Coronal-Labial_Coronal] place sequence constraint between the cluster and the tap. Because of metathesis, features relevant to the cluster were no longer adjacent, allowing [Labial] to be produced in the WF syllable onset.

In 26(e), vowel insertion preceded the syllabic consonant timing unit, transposing syllabic /ɫ̩/ to the coda position. Segmental slots were otherwise preserved with sacrifice of features, in particular, [+voice] assimilated in the word final weak syllable sequence.

For the word *thermometer* /θɚˈmɑ̃ːməɾɚ/, mismatches were more likely to occur in the unstressed syllable word initially and in the weak syllable sequence word finally. Nevertheless, tap mismatches almost entirely involved stopping, which occurred with the following frequencies: 8TD (6/34), 8PPD (5/12), and 5TD (21/62). The 8- to 10-year-olds were divided evenly as to [t] or [d], whereas 5-year-olds produced [t] more often than [d] one-third of the time. The proportion of stopped taps was highest for the 8PPD, followed by the 5TD and 8TD.

Exceptions to stopping were negligible and occurred in the 5TD or 8PPD groups. A sole 5PPD (1/32) participant nasalized tap. Productions in 27(a) through 27(c) were also unique but all suggested tap interaction with /ɚ/.

(27) *thermometer*	θɚˈmɑ̃ːməɾɚ	a) θəˈmɑ̃ːnəʧʊ̆	(8PPD)
		b) θɚˈmɑ̃ːnəʃɚ	(5TD)
		c) fɚˈmɑ̃ːməɽɚ	(5TD)

In 27(a), because [ə] was an acceptable dialect variant for /ɚ/ in the WI syllable of *thermometer*, the status of this production as a mismatch was unclear. In 27(b) and 27(c), the schwar was present. Regarding 27(a) and 27(b), respectively, in the prominent foot /ˈmɑ̃ːməɾɚ/, the tap timing unit was filled by [ʧ] and [ʃ]. Lexical influences in 27(a) might have related to activation of *temperature*, which also contains /ʧɚ/ in the word final syllable. Phonologically, production of [ʧ] inferred that [-continuant] [+continuant] features of tap were retained. [Labial] did not surface for /m/ in the onset of the unstressed syllable within the prominent foot, but likely transposed to the final syllable. The feature [+continuant] also might have assimilated from word initial /θ/, independent of, or in combination with [+continuant] from tap.

In summary, for MSWs containing taps, for the TD children, there was apparently a development progression between 5 and 8 years of age, an extension of the available information between 3 and 5 years of age (Klein & Altman, 2002). In comparison with the TD groups, tap accuracy was lowest for the 8PPD. The most common phonological pattern for all groups was tap stopping; voicing of tap, however, was the most problematic for the 8PPD, who more often produced [-voice] [t] than [+voice] [d]. For the 8PPD children in addition, patterns of deletion, nasal substitution and glottal replacement were more frequent. As noted for *thermometer* and *hospital*, as for all other word groups, challenging stress patterns and segment sequences interacted with tap production, such that mismatches suggested interactions across levels of the nonlinear hierarchy.

4.6 Overarching Discussion

The discussion first reviews the results reported for the quantitative data and then the description of key phonological patterns. The implications of the quantitative results for each sub-study are briefly deliberated in turn, before drawing conclusions. The findings concerning phonological patterns will be discussed in relation to previous research concerning MSWs, primarily the study by James (2006), the first to examine phonological development of a large number and variety of MSWs, across age groupings of children. Although Australian children were studied, James' findings are pertinent to the current study because monolingual English-speaking school-aged children were included.

4.6.1 Review of Quantitative Results

The MSW phonology of three samples of school-aged children, 5-year-olds with typical phonological development and 8- to 10-year-olds with and without PDD, was quantitatively examined on a MSW metric that tallied Lexical, Structure and Feature mismatches. There were three main findings from the multivariate analyses of the MSW mismatch totals. First, Lexical, Structure and Feature mismatches decreased significantly between the 5TD and 8TD groups. Second, in comparison with the age-matched 8TD group, the 8PPD demonstrated significantly more Structure and Feature mismatches. Third, the frequency of Lexical, Structure and Feature mismatches for the 8PPD and 5TD were equivalent. Implications of each of these results will be discussed in turn with respect to the Lexical, Structure and Feature totals.

Beginning with the Lexical totals, the effect size of the significant difference between the 5TD and 8TD was moderate, whereas the difference for the 8TD and 8PPD was without significance. It appeared that with developmental vocabulary growth and familiarity, lexical influences were less important. The large effect sizes for all significant comparisons of Structure and Feature totals, however, suggested that each of these aspects of MSW phonology continued to develop between ages 5 and 8 years. James (2006) used a phonological process frequency analysis to examine word structure and features in MSW production between ages 5 and 7 years, and reported similar results. That is, deletions and insertions of word structure components, (syllables, consonants and/or vowels) decreased during this period. Patterns related to feature production and sequencing also declined after age 5 years, particularly for words longer than two syllables.

Together with James' findings, the current results generally support a hierarchical theory of phonology, and the utility of a whole word metric for measuring MSWs. With respect to the metric, the Lexical total differentiated performance of the 5-year-olds from the 8- to10-year-olds with and without PPD, implying a connection between this level of language organization and phonology, at least until words are familiar. The interaction of both Structure and Feature totals distinguished the performance of typically speaking 5-year-olds, or 8- to 10-year-olds with PPD, from 8- to 10-year-olds with typical phonological development, verifying the contribution of both nonlinear theoretical constructs. Data entry into spreadsheets provided a visual display of mismatch tallies, color-coded for the multiple phonological patterns that often occurred within a single word. Such treatment of the data aided interpretation of interactions across a word, interactions that remained unexplained when patterns were described in isolation. Exposing such interactions, in turn provided verification of the parallel processing of interconnected components.

The discriminant analyses suggested a fourth finding, i.e. the MSW whole word tally metric that included tallies for Lexical, in addition to Structure and Feature mismatches, accurately classified children according to MSW developmental age group, 5TD or 8TD, and with regard to the typicality of phonological development, i.e. 8TD or 8PPD. To date, no other studies of MSWs have quantified phonological structure using a nonlinear framework. James (2006), however, reported traditional Percent Vowels Correct (PVC) and Percent Consonants

Correct (PCC), an estimate of nonlinear segmental accuracy. After age 5, PVC was 94% or better, and PVC, 90% or better, with 1% growth in each of successive years up to age 7 years. These data suggested little developmental distinction from ages 5 through 7 years (the oldest of the groups studied), and indeed, phonological mastery of the words. The MSW whole word metric, therefore, appeared to have promise as a method for quantifying differences in the phonological development of a group of complex words, which were minimized using traditional quantification methodology.

4.6.2 Review of Phonological Patterns by Participant Group and Phenomena

The results section described the observed mismatch patterns by MSW word types. Here the data are summarized first in terms of patterns for the participant groups, and then by key developmental phenomena.

4.6.2.1 Participant Group Phenomena

For the 8TD, phonological mismatch patterns were negligible, even if words were possibly less familiar, e.g. *invitation, thermometer,* or *cash register*, or vulnerable to lexical confusion, e.g. *electric* or *magician.* Thus, consistent with many studies of phonological development, by age 8 years (see McLeod, 2009), the Canadian English-speaking children showed mastery of many long words. Productions of *umbrella* and *magician* were the most frequently mismatched. For *cash register,* a long word with numerous components, more frequent mismatches relative to shorter words were anticipated for all groups; even so, for the 8TD, the range of mismatches on *cash register* was small, from zero to four, with six or seven tallied for only 2/34 children. Mismatches typically involved some combination of feature deletion or assimilation to resolve NotTwice constraints on the consonant tier, i.e. NotTwice[+continuant], a manner constraint, interacting with a NotTwice place constraint, i.e. [Coronal] [Coronal] sequence: [dʒ_st]. Finally, the 8TD apparently had yet to master tap production, in that about half used [-voice] [t] substitutions. For adults, taps are normally [+voice] (Shockey, 2003), but whether the adults in this dialect region use taps consistently is not known at this time. Nonetheless, the 8TD were apparently still refining the articulatory precision necessary for accuracy of MSW production.

Turning next to the 5TD, where some phonological mismatches would still be expected, both for later-developing segments (especially sibilants, rhotics) and long words, stress shifts seldom occurred in the 2-syllable to 3-syllable words evaluated in the study. Instead, consistent with the research (James, 2006; Kehoe, 2001), stress pattern change resulted from deletion of weak initial or within-word syllables, and occasionally, of a word-initial syllable with secondary stress. Weak syllables interacted with a following sonorant onset, such that an allowable English consonant sequence resulted, either a liquid cluster or adjacent sonorants, respectively, e.g.: (a) clusters, [gɹ], in *gorilla* produced as [ˈgɹɪlə]; or [lt], in *skeleton* produced as [skɛltn̩]; and (b) the sonorant sequence, [mn], as in *animal* produced as [æmnl̩] (1/62). Weak syllable interaction with sonorant onsets was also consistent with previous research (James, 2006; Kehoe, 2001).

Concerning the 8PPD, a variety of patterns was observed related to production of weak syllables. Weak syllables within the foot were subject to deletion. As was the case for the 5TD, in disyllabic words, weak syllables were also deleted. This was expected for the younger children in that, developmentally, trochaic stress (*Strong weak*) is mastered before iambic stress (*weak Strong*). The result for the 8PPD children, being older, emphasizes the importance of observing particular mismatch patterns across a variety of word lengths, to examine the impact of length and stress on successful production. Additional patterns unique to the 8PPD interacted with weak syllables. Concerning lax vowels, feature insertions such as raising or lowering (e.g. *electric*, *invitation*), or production of caret (e.g. *computer*), created slight increases to syllable prominence.

In words of all lengths once again, certain 8PPD children inserted various components in relation to sequence constraints. For instance, in *vegetable* and *skeleton*, insertion of a word-initial lax vowel (excrescence) created feet of equal iambic stress, possibly interacting with the within-word weak syllable, to prevent deletion. Alternatively, a word-initial [-sonorant] insertion, [h], in *alligator,* prevented syllables without consonant onsets. NotTwice place constraints also resulted in coda deletion in contiguous coda-onset sequences, e.g. *electric, computer*. More often, consonant feature assimilations affected words of every length, primarily for sequences of sonorants, e.g. or of non-contiguous [Dorsal_Coronal]. A final observation for this section was that later developing sibilants, particularly [-anterior] palatoalveolars, were produced correctly when a single segment was present in the word, e.g. /ʃ/ in *invitation*, but were less likely to all be accurate for multiple segments, e.g. *cash register*. Further interactions were relevant to taps, schwa and syllabic /ɫ̩/, discussed next under phenomena.

4.6.2.2 Phonological Phenomena in MSWs

The key phonological phenomena observed in these data occurred across the phonological hierarchy, both at autonomous levels and in interaction between levels. Starting at the level of the foot, the key pattern observed was with the 8PPD group, i.e. stress shift, only rarely found in 5TD. This went along with their higher tendency for vowel mismatches, which included both quality changes and lengthening. Together with mismatches for tap, the suggestion was of a specific difficulty for that group of 12 children with the timing of units, particularly short ones. In terms of syllables, deletion often affected the weak syllables at the beginning and in the middle of words as expected and as was found by James (2006) and others. In some cases, relevant onsets were retained in the coda of the prominent syllable, as also found by James (2006). Consonant deletion, if it occurred, was more likely to resolve complex sequences (contiguous or non-contiguous), again as was expected. Individually, segments described as late acquired appeared to be late acquired here also (sibilants, liquids). Taps were noted to be one of the later acquired elements.

The previous data add to the small database on children's acquisition of long words with general patterns similar to those reported to James (2006), but for a Canadian English-speaking group of children in a small urban-rural community. However, what was more striking about the current data, was the interaction between the various components of the phonological hierarchy. The nonlinear phonological theories provided a way to describe those interactions.

Key interactive phenomena occurred both within prosodic levels and between prosodic and segmental levels. For example, prosodic interactions included: (a) adjustments of foot structure in two-footed words, thereby either maintaining the target number of syllables in the word in spite of mismatches for specific feet (24c, *cash register* [ˈkʰæθdɚˌɹɛðdʊ]), or increasing the total number of syllables in order to provide similar structure for the two feet (i.e. trochaic; 24b, [ˈkʰædəˌɹɛʒdɪʃtɚ]); (b) adjustments to foot or syllable structure such that a syllable or timing unit might surface in a less preferred prosodic position (24d, [ˈkʰæʃ̟ˌɹɛʃ̟stʊɾʊ]); or (c) adjustments of syllable weight to increase the prominence of the weak syllable, e.g. consonant insertion (*guitar*, 3a, 3b), vowel lengthening (*electric*, 12b), and diphthongization (*electric*, 10). Prosodic-segmental interaction examples were: (a) mismatches for earlier-acquired segments, e.g. /m/, /k/ /g/, /p/, when appearing in weak prosodic positions (*wSw* or *Sww*) or challenging sequences, e.g. *animal* [ˈæ̃ːmdl̩], *explodes* (1a, 1d), *guitar* (3), whether because of repeated features (three labials in *umbrella)* or different features [Coronal_Labial] (as in *animal* 1a); (b) mismatches for later-acquired segments (sibilants, liquids) when there was more than one such type within a longer word (*gorilla, cash register*); and (c) effects of phonotactics, such that deletion might be permitted if an allowable sequence would occur, e.g. deletion of schwa in *balloon* or *giraffe*, but not in *guitar*; alternatively, an illegal sequence might be permitted to allow more timing units and/or more features to be realized.

With some of these examples, there were paradoxical interactions in that simplification at one level may have resulted in complexity at another or vice versa. For example, in example 24c, *cash register* [ˈkʰæθdɚˌɹɛðdʊ] above, there was a paradox such that apparent increased complexity at the foot level (i.e. insertion of a syllable to increase length of the first foot) actually was a type of simplification whereby insertion resolved a markedness constraint at the level of the prosodic word (i.e. a sequence of two stressed syllables). The resulting stress patterns in the feet were harmonized, i.e. two trochees. In contrast, deletion of a schwa in an unstressed syllable, a simplification of foot structure, resulted in increased complexity at the level of the syllable onset, e.g. deletion of schwa in *giraffe* resulting in a [dʒɹ] cluster. These patterns suggest that to construe such mismatches as inconsistent linear deletions, substitutions or patterns of assimilation, provides a surface level understanding of what are multiple component interactions within complex phonological structure. The current study has provided many more examples of such interactivity within the phonological (processing) system. In the final chapter we will explore further the relationship of these phenomena to nonlinear phonology and phonological processing theories.

Table 4.1 Means for language-related skills for typically developing 5- and 8-year-olds

Language measure	Age group (years)	Mean	Standard error	95% Confidence Interval	
				Lower limit	Upper limit
Vocabulary (*PPVT-4*)	5	118.27	2.81	112.61	123.94
	8 to 10	116.55	2.81	110.88	122.21
Working memory (*CELF-4*)	5	10.18[a]	.50	9.17	11.20
	8 to 10	9.27[b]	.50	8.26	10.29

Note. TD = Typical phonological development. *PPVT-4 = Peabody Picture Vocabulary Test-4* (Dunn & Dunn, 2007). *CELF-4 = Clinical Evaluation of Language Functions-4* (Semel, Wiig, & Secord, 2003).
[a]*Number Repetition Forward* subtest. [b]*Number Repetition Backward* subtest.

Table 4.2 MSW tally correlations for typically developing 5- and 8-year-olds

		Subtotals		Total
		Structure	Feature	Whole word
Subtotals	Lexical	.46*	.59*	.60*
	Structure		.77*	.87*
	Feature			.99*

Note. MSW = multisyllabic words. Correlation = Pearson *r*.
*Correlation is significant at the .01 level (2-tailed)

Table 4.3 MSW tally means for 5- and 8-year-olds with typical development

Tally	Age group (years)	Mean	Standard error	95% Confidence Interval	
				Lower limit	Upper limit
Lexical	5	.56	.13	.29	.82
	8 to 10	.05	.05	-.05	.14
Structure	5	11.69	1.80	7.95	15.43
	8 to 10	6.10	.70	4.64	7.56
Feature	5	39.96	3.33	33.04	46.88
	8 to 10	9.81	1.28	7.14	12.47
Whole word	5	52.21	4.96	41.90	62.53
	8 to 10	15.95	1.88	12.05	19.85

Note. MSW = multisyllabic words.

Table 4.4 Means for language-related skills of 8- to 10-year-olds with typical versus protracted phonological development.

Language measure	Speech group	Mean	Standard error	95% Confidence Interval	
				Lower limit	Upper limit
Vocabulary (*PPVT-4*)	PPD	110.75	2.89	104.77	116.74
	TD	114.25	2.89	108.27	120.24
Working memory (*CELF-4*)[a]	PPD	8.75	.64	7.42	10.09
	TD	10.42	.64	9.08	11.75

Note. PPD = Protracted phonological development. TD = Typical phonological development. *PPVT-4 = Peabody Picture Vocabulary Test-4* (Dunn & Dunn, 2007). *CELF-4 = Clinical Evaluation of Language Functions-4* (Semel, Wiig, & Secord, 2003).
[a]*Number Repetition Backward* subtest.

Table 4.5 MSW tally correlations for 8- to 10-year-olds with protracted versus typical development

		Subtotals		Total
		Structure	Feature	Whole word
Subtotals	Lexical	-.19	-.16	-.16
	Structure		.68*	.81*
	Feature			.98*

Note. MSW = multisyllabic words. Correlation = Pearson *r*.
*Correlation is significant at the .01 level (2-tailed)

Table 4.6 MSW tally means for 8- to 10-year-olds with typical versus protracted phonological development

Tally	Speech group	Mean	Standard error	95% Confidence Interval	
				Lower limit	Upper limit
Lexical	PPD	.33	.17	-0.018	0.69
	TD	.42	.17	0.065	0.77
Structure	PPD	15.50	1.56	12.26	18.74
	TD	6.44	1.56	3.20	9.68
Feature	PPD	46.25	3.45	39.10	53.40
	TD	12.99	3.45	5.83	20.14
Whole word	PPD	62.08	4.19	53.39	70.77
	TD	19.86	4.19	11.17	28.55

Note. MSW = multisyllabic words. PPD = Protracted phonological development. TD = Typical phonological development.

Table 4.7 Means for language-related skills of 8- to 10-year-olds with protracted phonological development and typically developing 5-year-olds

Language measure	Speech group	Mean	Standard error	95% Confidence Interval	
				Lower limit	Upper limit
Vocabulary (*PPVT-4*)	PPD	110.75	3.86	103.06	118.44
	TD	117.11	1.70	113.73	120.50
Working Memory (*CELF-4*)	PPD	8.75[a]	.59	7.59	9.92
	TD	10.18[b]	.26	9.67	10.69

Note. PPD = Protracted phonological development. TD = Typically developing. *PPVT-4 = Peabody Picture Vocabulary Test-4* (Dunn & Dunn, 2007). *CELF-4 = Clinical Evaluation of Language Functions-4* (Semel, Wiig, & Secord, 2003).
[a]*Number Repetition Backward* subtest. [b]*Number Repetition Forward* subtest.

Table 4.8 MSW tally correlations for 8- to 10-year-olds with with protracted phonological development and typically developing 5-year-olds

		Subtotals		Total
		Structure	Feature	Whole word
Subtotals	Lexical	.35*	.25	.32*
	Structure		.73*	.87*
	Feature			.97*

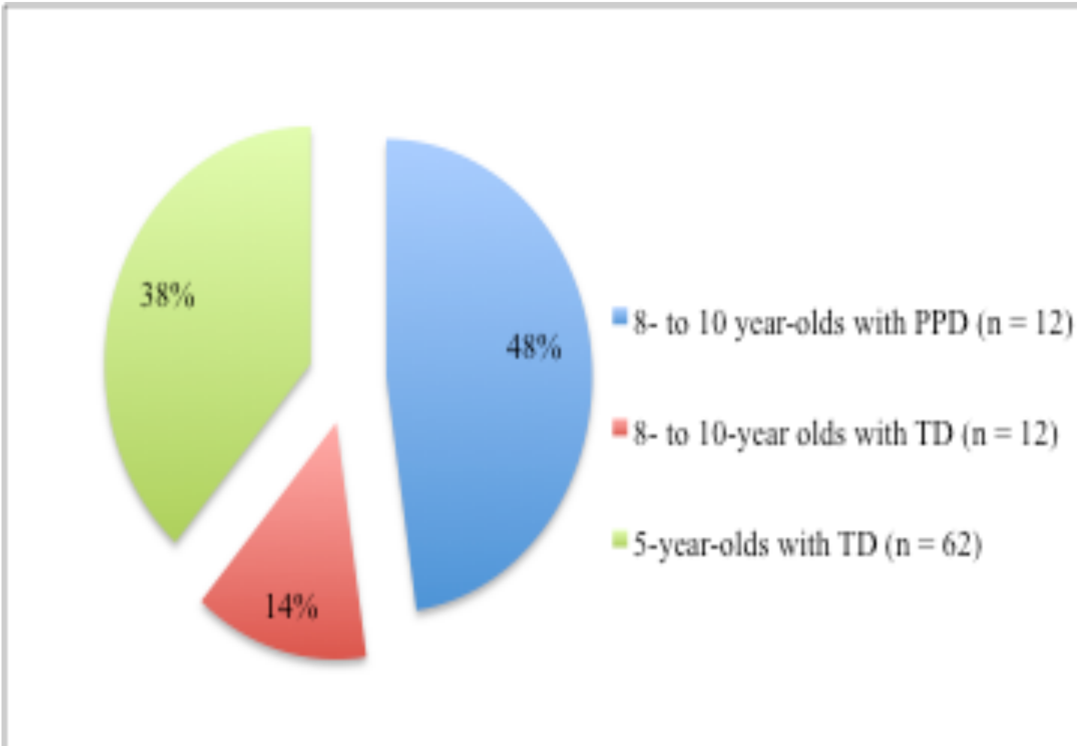
Note. MSW = multisyllabic words. Correlation = Pearson *r*. PPD = Protracted phonological development. TD = Typically developing.
*Correlation is significant at the .01 level (2-tailed)

Table 4.9 MSW tally means for 8- to 10-year-olds with protracted phonological development and typically developing 5-year-olds

Tally	Speech group	Mean	Standard error	95% Confidence Interval	
				Lower limit	Upper limit
Lexical	PPD	.33	.18	-.02	.69
	TD	.40	.08	.24	.55
Structure	PPD	15.50	2.38	10.76	20.24
	TD	11.22	1.05	9.14	13.31
Feature	PPD	46.25	5.05	36.19	56.31
	TD	38.63	2.22	34.21	43.06
Whole Word	PPD	62.08	7.01	48.11	76.06
	TD	50.27	3.09	44.12	56.42

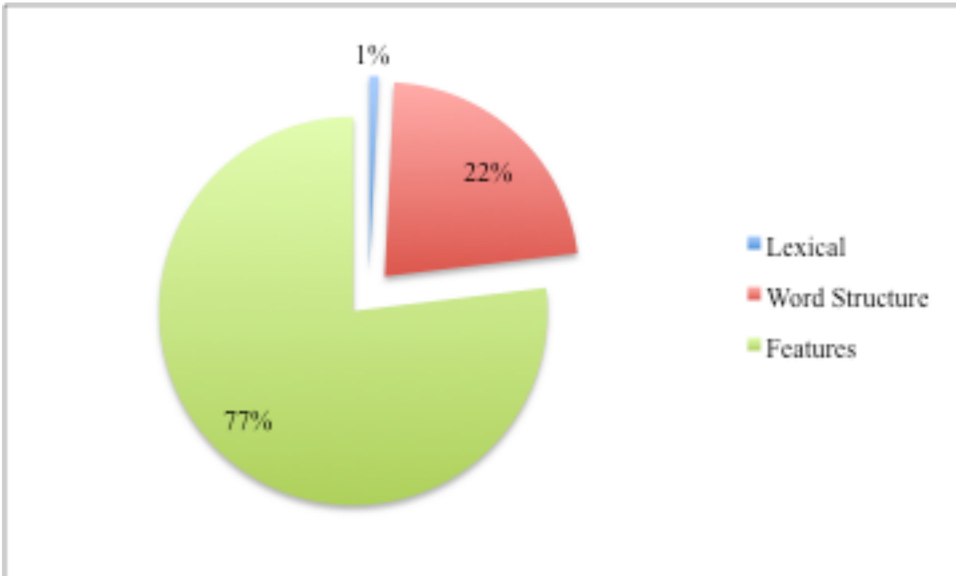
Note. MSW = multisyllabic words. PPD = Protracted phonological development. TD = Typically developing.

Figure 4.1 Group proportions of total mismatches for study



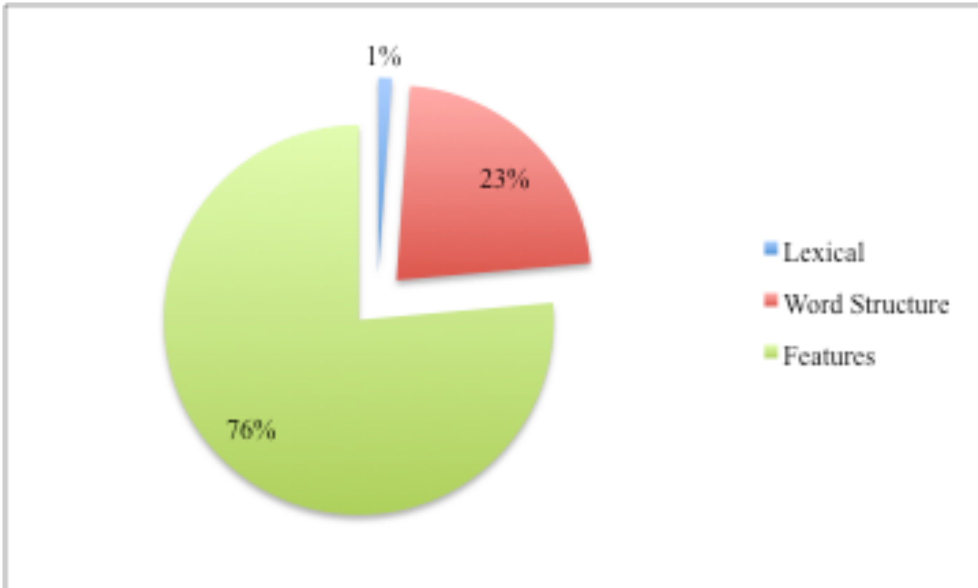
Note. PPD = Protracted phonological development; TD = typically developing.

Figure 4.2 Proportion of mismatch types for all TD 5-year-olds (n = 62)



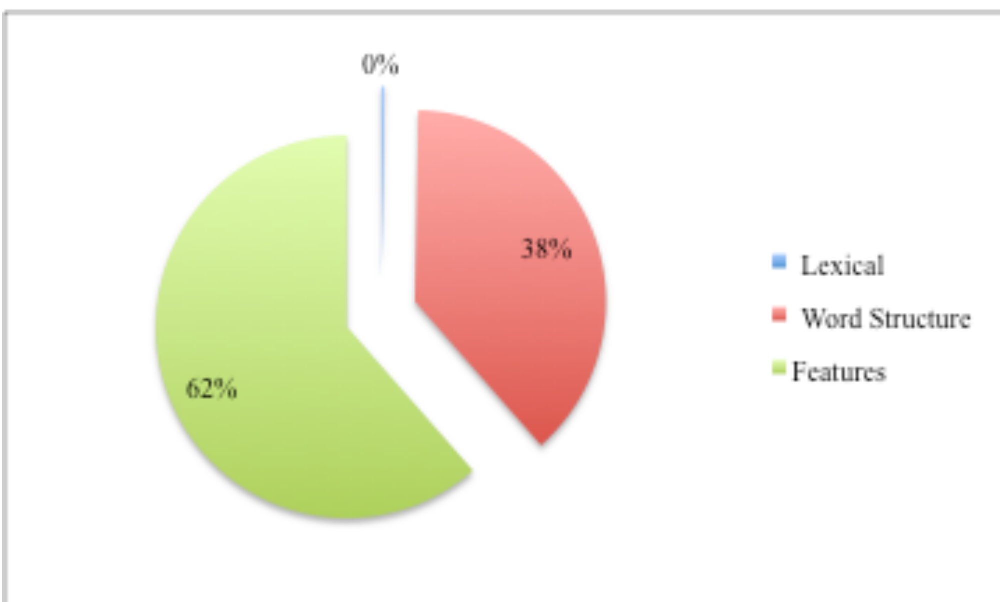
Note. Whole word total = 3142; TD = typically developing.

Figure 4.3 Proportion of mismatch types for longitudinal TD 5-year-olds (n = 22)



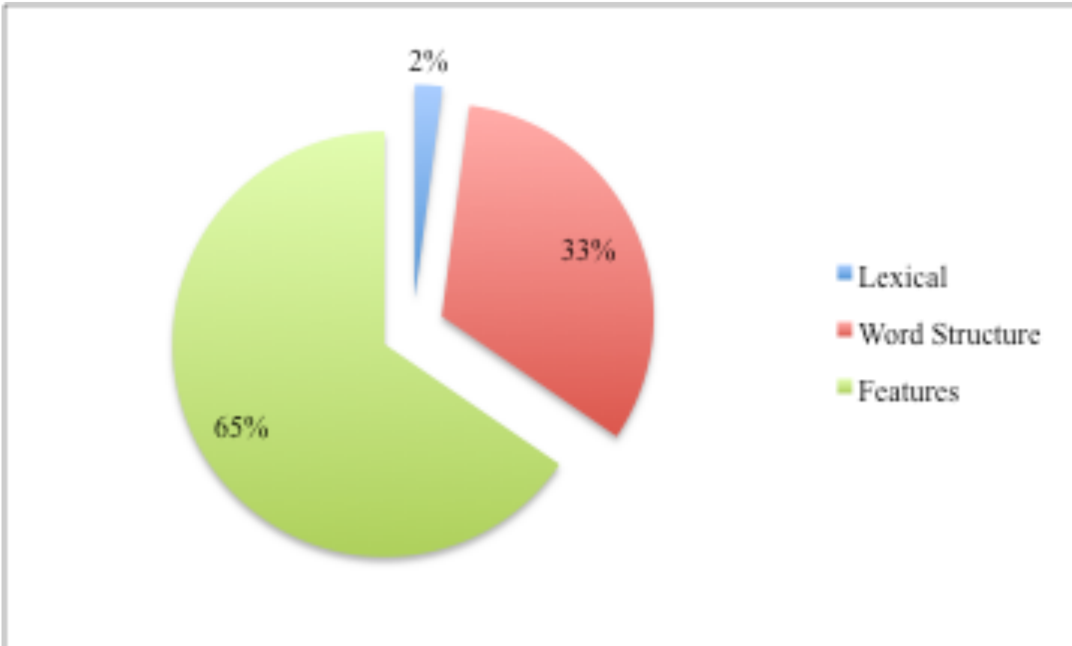
Note. Whole word total = 1172; TD = typically developing.

Figure 4.4 Proportion of mismatch types for longitudinal TD 8- to 10-year-olds (n = 22)



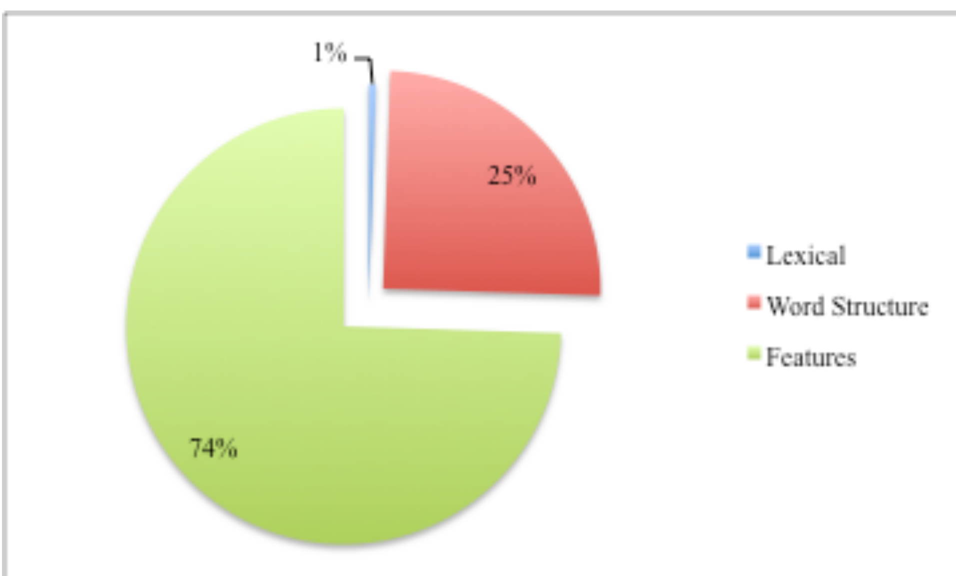
Note. Whole word total = 351.; TD = typically developing.

Figure 4.5 Proportion of mismatch types for cross-sectional TD 8- to 10-year-olds (n =12)



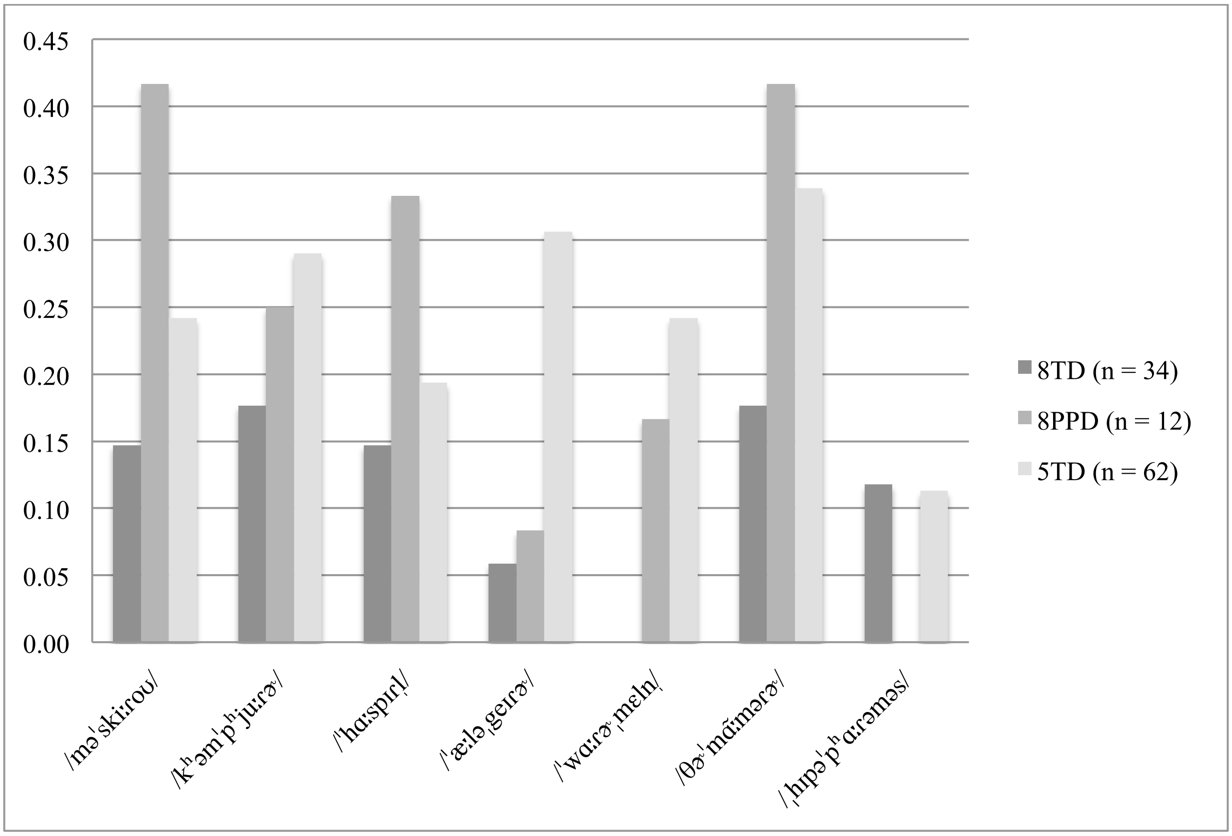
Note. Whole word total = 238; TD = typically developing.

Figure 4.6 Proportion of mismatch types for cross-sectional 8- to 10-year-olds with PPD (n = 12)



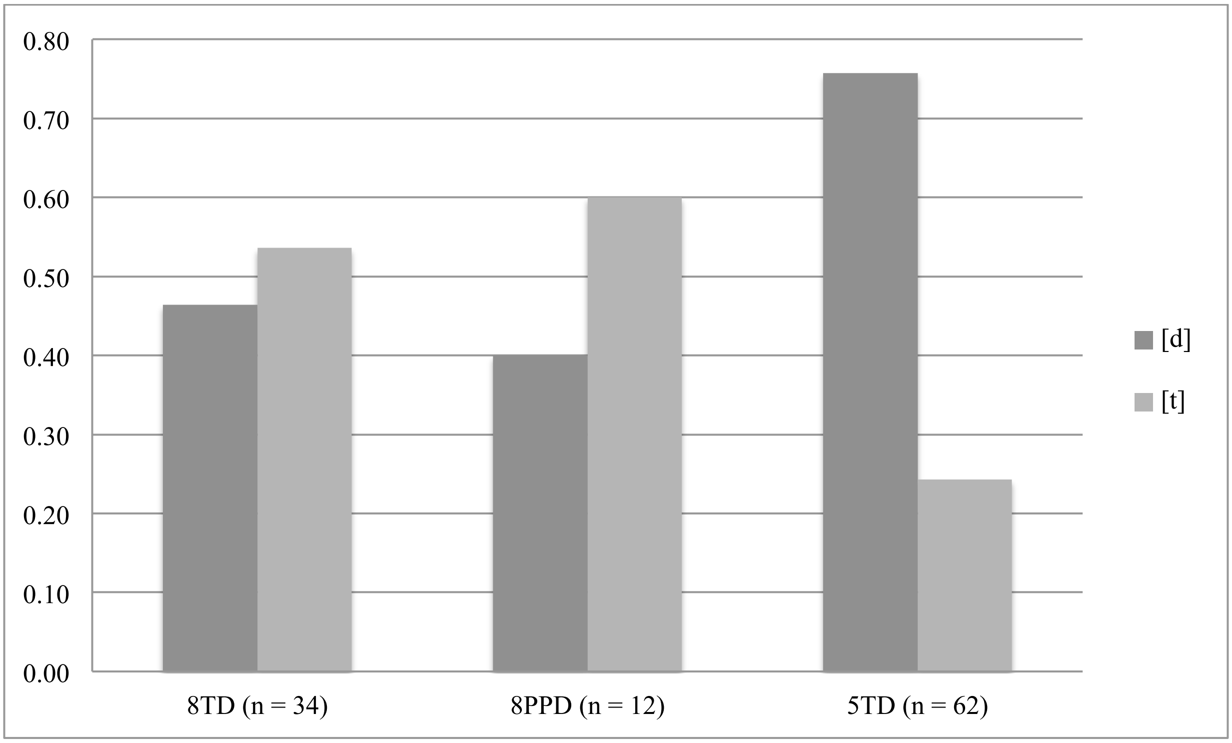
Note. Whole word total = 745; PPD = protracted phonological development.

Figure 4.7 Proportion of word productions with taps produced as coronal stops



Note. 8TD = 8- to 10-year olds with typical phonological development, those compared longitudinally at 5-years of age (n = 22), and those compared cross-sectionally with 8- to 10-year-olds with PPD (n = 12). 8PPD = 8- to 10-year-olds with PPD (n = 12). 5TD = 5-year-olds with typical phonological development (n = 62).

Figure 4.8 Proportion of taps produced as [+voice] or [-voice] coronal stops



Note. 8TD = 8- to 10-year olds with typical phonological development, those compared longitudinally at 5-years of age (n = 22), and those compared cross-sectionally with 8- to 10-year-olds with PPD (n = 12). 8PPD = 8- to 10-year-olds with PPD (n = 12). 5TD = 5-year-olds with typical phonological development (n = 62).

Chapter 5: Conclusion

This thesis concerned multisyllabic word (MSW) phonological development of school-aged children in an urban-rural region of British Columbia, Canada. A brief overview of the studies is first presented, as a background to discussion of the validity of the MSW whole word metric and the theoretical bases thereof: nonlinear phonology and parallel interactive language processing. The discussion will continue with the clinical implications of the thesis, before turning to limitations, in terms of the current studies and of the MSW metric, providing suggestions for future research. The chapter will conclude by outlining the contributions of the current research.

5.1 Overview of the Studies

The primary aims of the thesis were to compare MSW phonological production in school-aged children with and without PPD, and to establish a criterion reference for phonological accuracy in these long and complex words. The research was motivated by a meta-analysis (Chapter 2) that identified a relationship between MSW phonological production and literacy, and the utility of MSW evaluation for identifying ongoing PPD. Given this significant result, for the main research, a whole word metric was designed in order to quantify the complex interactions of phonological components across a MSW production, because a theoretically grounded metric for quantifying the data was unavailable. The metric was structured from complementary principles of nonlinear linguistic theory and parallel interactive theories of language processing. A pilot study (Chapter 3) evaluated the reliability and validity of the MSW metric, which was found to better differentiate variable word productions than traditional phonological metrics, e.g. PCC (Shriberg & Kwiatkowski, 1982), or sums for calculating PMLU

(Ingram, 2002). Having established the reliability and validity of the whole word MSW metric, in the main study (Chapter 4), the metric compared MSW accuracy between school-aged children with and without typically developing phonology (TD and PPD, respectively), asking three major questions. The first question addressed developmental aspects of MSWs in TD children pre- versus post-early literacy instruction (age 5 years, and ages 8- to 10 years) in a longitudinal comparison. As predicted, the older TD children showed significantly fewer mismatches than the younger children (Lexical, Structure, Features), a finding similar to that of James (2006) for Australian English-speaking school-aged children. The second question asked whether there would be a difference between children with and without a history of PPD at 8 to 10 years of age. Again, as expected, the TD group had significantly fewer mismatches in structure and features than the group with PPD; in fact, the TD children showed near-mastery of all words on the whole word metric. However, there were no significant differences in lexical mismatches. The third question asked whether children post-early literacy instruction with history of PPD would show lower accuracy than TD 5-year-olds. This was not confirmed for word structure and feature mismatches, the 8PPD group showing similar mismatch proportions to that of the 5TD group. However, the 8PPD group had significantly fewer lexical mismatches than the 5TD group. In addition to mastery levels, differences with respect to phonological patterns were expected in each comparison, because such differences have also been demonstrated for short words (Leonard, 1992). Overall, a whole word metric that included tallies for lexical influences (particularly for younger children) in addition to word structure and segmental features, accurately classified children with and without PPD in their original groups. This is discussed further below in terms of validity of the measure and the theoretical foundations for the metric.

In the current study, phonological analyses of the mismatch patterns showed some similarities between the 5TD and 8PPD groups, but there were also unique patterns observed in both groups, with more of those appearing in the 8PPD group (e.g. vowel insertions, lengthening and quality differences). Leonard (1992) reported a similar finding for short words in preschool children, noting that the phonological patterns of children with PPD were less systematic, and presumably within a larger vocabulary than those of younger TD children. Presuming that MSW words are less familiar to younger children than to older children with PPD (lexical effects), the current study suggests if evaluated on the same MSW elicitations, some patterns of children with PPD will be exceptional, nevertheless.

In the next two sections, the results are discussed in terms of validity of the MSW metric and in relation to the theoretical bases thereof, i.e. nonlinear phonology and parallel interactive language processing. Particular attention is paid to the ability of the metric to explain interactions and variability. Timing will be emphasized, in particular, for syllable units, schwa, syllabic consonants, and taps, because of the theoretical importance of timing units in nonlinear phonology, and for the timing of activation levels during language processing.

5.2 Validity of the MSW Metric

In the development of a metric, it is crucial to establish its validity in a variety of ways. The validity of the MSW metric was considered in several respects, including for construct, convergent, and divergent validity, and in addition, sensitivity and specificity. First, construct validity of the metric was established with respect to the theoretical bases of nonlinear phonology and parallel interactive language processing, discussed further below. Next, the pilot study demonstrated convergent validity in relation to strong correlations with two other phonological metrics, PCC and the sums for calculating PMLU. Third, sensitivity of the metric was shown by more rankings for variable word productions, and for children. Furthermore, differences in the sequential order of children reflected divergent validity, and as expected, were relative to the number of structure mismatches tallied. That is, in comparing the three metrics, dissimilarity of rank ordering was reflected in the new metric having more structure mismatches and fewer feature mismatches, because the comparison metrics accounted less for components of the structure. The MSW metric, therefore provided more information from which to judge the relative amount of mismatch among children. Finally, in Chapter 4, sensitivity and specificity of the MSW metric were indicated by the post-hoc discriminant analyses, because developmentally, over 90% of children with TD were correctly classified by age. The metric was also sensitive to predictable increases in lexical familiarity, in that the most lexical mismatches were made by the 5TD group, whereas there was no difference for 8- to 10-year-olds with or without PPD. Children were also correctly classified as to history of PPD. The latter two results suggested that phonological difficulty, and not word familiarity, was key for children with PPD, post-early literacy instruction. Given that children with PPD had more mismatches than younger TD children, even if not significantly different, it appeared the metric had the ability to identify ongoing PPD. The metric is discussed further with regard to limitations and strengths, in the last two sections of the chapter. The following section further discusses construct validity in relation to the theoretical foundations of the metric.

5.3 Construct Validity and Nonlinear Phonology

This section will pull out some key constructs of nonlinear and parallel interactive language processing theories (see sections 1.9.2 and 1.9.3) that speak to the construct validity of the metric. First, there are two fundamental constructs in nonlinear phonology: (a) the hierarchical organization of linked phonological units; and (b) the autonomy of those units. The consequence of these two constructs is that there is tension between autonomy of processing on one level and interactions with processing on other levels of the hierarchy. Constraints within autonomous domains may therefore be affected by constraints in another domain. The metric was designed to instantiate the two fundamental constructs above. *Autonomous units* across the phonological hierarchy were designated for tallying of mismatches. Mismatch tallies were then totalled as a way to encapsulate *interactions across the hierarchy* through a whole word measure. The data supported both of these constructs: (1) patterns affected individual units, e.g. feet, prominence, syllables, segments, features; and (2) MSWs showed many examples of multiple interactions between tiers. Further to the constructs, in Chapter 4, words were categorized by length (number of feet and syllables) and stress patterns, with the assumption that similarities and differences would be more apparent given these groupings. In fact, in spite of the groupings, numerous productions showed many, unique, often child-specific interactions (section 4.6), e.g. 1(a), *explodes*; 7(c) *guitar*; 8(b) *umbrella*; 12(b) *electric*; 17(b) and 17(c) *gorilla*; 19(a) *vegetable* and 19(b) *skeleton*; 21(a) through 21(e) *cash register*; and 22(g) *hospital*. The consequence of the many interactions was variability across children in pronunciation of individual words and word groups. In spite of the variability, however, there were general trends across all word types showing interactions of multiple tiers. Interactions were more likely: (a) in weak prosodic (less acoustically salient) positions such as non-prominent word-initial or -medial syllables (as noted in James, 2006; Kehoe, 2001); (b) when there were sets of complex sequences, either contiguous or non-contiguous (as noted in Bernhardt & Stemberger, 1998); (c) when there were later-developing segments, especially when more than one was in the word; and (d) there were timing constraints on production (short vowels, taps).

A third construct concerns adjacency, reviewed here in terms of three principal conditions: (a) there is no intervening element on the same tier (tier adjacency, e.g. the consonant tier); (b) there is nothing between the Root nodes dominating two elements (surface-adjacent or contiguous); or (c) two elements are situated in two adjacent syllables (Bernhardt & Stemberger, 1998). Instantiation of these principles was also apparent in the children's MSWs productions. Keeping in mind the interactions, the examples from Chapter 4 below highlight mismatches reflecting tier adjacency on various tiers of the phonological hierarchy:

Prosodic word tier: stress shift	(17c) *gorilla*, (26a) *hospital*
Foot tier: syllable transposition between feet	(24b) *cash register*
Foot tier: syllable reduplication between feet	(24e) *cash register*
Syllable tier: simplified coda-onset sequences	(1d) *explodes*
Consonant tier: simplified onset clusters	(12b) *electric*
Vowel tier: vowel place harmony	(8b) *umbrella*
Feature tier: consonant manner and place harmony	(24g) *cash register*
Feature tier: consonant voicing harmony	(15) *computer*

The fourth major construct concerns syllable weight and timing. In the productions of children with PPD, syllable weight was possibly increased in unstressed syllables, facilitating its appearance in the surface production of the word, shown in *electric* as [ˌeɪːˈlɛktɰɹɪk] (10) and another variant of *cash register* below.

(24) *cash register* ˈkʰæːʃˌɹɛʤɪstɚ a) ˈkʰæʃˌɹɛˌʤɪʃstɚ (8PPD)

The examples above demonstrate the variable strategies for increasing syllable weight, i.e. diphthongization (10) and coda insertion (24a). Once again, the pattern is not without the interactions because of other word complexities, including the complex consonant sequences. For instance, in 24(a), [ʃ] insertion of weak syllable [ʤɪʃ] within the word has arguably increased the difficulty of the Coronal consonant sequence. Faithfulness to syllables has apparently taken precedence over a NotTwice[+continuant] manner constraint. Constraints within autonomous domains may therefore be affected by constraints in another domain. Explanations for such constraints are also brought to light by considering models of language processing, the next topic of discussion.

To explain the interactions in MSW productions, models of nonlinear phonology can be mapped onto parallel interactive models of language processing. Whereas the phonological hierarchy provides a description of the relationships among the various units of MSWs, models of parallel interactive language processing operationalize the activities within the hierarchy. Mapping is possible because of alignment among a number of constructs of the two types of theories. First, the models account for the link between the lexicon and the prosodic word such that both semantics and phonology could influence MSW output. The relationship of semantic similarity and familiarity to phonological output is thus taken into account. Second, relationships among component domains, or nodes and hierarchical networks, are explained by levels of unit activation and the feed forward-feed backward of information. Third, trade-offs in word structure and segmental accuracy are clarified by the relative activation weights of weaker and stronger prosodic domains. Incorrect weights compromise the amount of cognitive capacity needed to accurately access and select (license) phonological (reliable and valid) cues with the appropriate timing. This competition within networks may result in output that is less complex at one level (deletion of an unstressed schwa) but more complex at another, as in creation of a [dʒɹ] cluster when the schwa of *giraffe* disappears. The phenomena suggest that more that one factor contributes to phonological mismatches in the output of a word, at least until the relative weights of connections are stabilized in the system.

5.4 Clinical Implications

The perspectives discussed above concerning nonlinear phonology and parallel interactive models of language processing suggest implications for clinical practice in the area of child phonology, especially for school-aged children. The practicalities of clinical service provision necessitate tools that contribute to efficient evidence based assessment and intervention methodologies. The current research has provided an initial theoretically based whole word metric for quantifying the complex interactions in MSWs. The metric therefore has utility for efficiently establishing baselines and measuring progress in the accuracy of MSW production. The metric also provides information about relative proficiencies at the lexical, word structure and segmental feature levels that might inform the design of word-specific strategies for teaching contextually relevant words in speech and print. In the pilot study, the utility of the metric was also examined in another Canadian language dialect, i.e. French. The metric thus has potential utility in languages relevant to children in multilingual contexts.

After early-literacy instruction, children need to read and comprehend many new multisyllabic words (Cunningham, 1998; Duncan, 2009). The current research has also provided meta-analytic evidence of the connection of MSWs to literacy. As a result, for children with a history of PPD, phonological evaluation without MSWs may mean that ongoing protracted development goes undetected, and in turn excludes children from intervention to support both speech and literacy.

In general, the current study has developed an initial metric for the measurement of MSW accuracy in school-aged children. Continued use and evaluation is needed to establish the clinical validity and reliability of the metric, and in turn, the ability to generalize its use.

5.5 Limitations of the Studies and Directions for Future Research

Limitations relate to the design of the studies and of the MSW metric, which restrict the generalization of the results, and are discussed in turn below. The section begins by suggesting the limitations of the studies, i.e. the meta-analysis (Chapter 2), the pilot study of the MSW metric (Chapter 3), the main study (Chapter 4); and ends with the limitations of the metric. Directions for future research are indicated throughout.

Limitations of the meta-analysis (Chapter 2) concerned sample size, which were relatively small for the number of included studies, and in addition, for samples within studies. First, the small sample of included studies, in turn, resulted in small samples of the literacy constructs (*Phonological awareness, Word decoding, Word fluency, Nonword decoding* and *Spelling*), and therefore, the number of moderators that could be tested in relation to the probability of a Type I error. Although considered one of the most important predictors of reading comprehension (Puranik et al., 2008), *Fluency* was excluded from testing on the moderators, because that sample was particularly small. Second, other important moderators of early literacy might have been omitted, e.g. language development. To some extent, however, this variable was controlled by an inclusion criterion of average language abilities for study participants. Third, sampling procedures for included studies were various, and recruitment was often difficult; consequently, broad age ranges were examined, potentially widening the children's developmental differences. An additional issue was that children with PPD had been classified by a variety of clinical means, such that heterogeneity was likely. For each literacy construct, future research must include children with PPD within narrower age ranges, and for whom MSW output has been measured on a continuous metric. Studies of MSW production must also examine the relationship to reading comprehension because of the increasing frequency of MSWs to be read in text, which carry most of the meaning.

Concerning the pilot study of the MSW metric, limitations related to small convenience samples, in this case, of words in addition to children. Mitigating this factor, however, might have been the increased likelihood of mismatches in the samples, because: (a) the sub-sample of selected words included the longest and/or most complex from the main study; (b) the sub-sample of children included those with the most mismatched productions on the selected words; and (c) the metric was highly correlated with the comparison measures, i.e. PCC and sums for calculating PMLU. Nevertheless, further study of the metric was warranted, and was reported in the main study for the thesis (Chapter 4).

Concerning the main study, the sample was predominantly monolingual, and representative of the demographics of a small urban-rural region of central British Columbia, thus restricting generalizability of the results. Because of recruitment difficulties, the sample of children with PPD was very small (n = 12) and heterogeneous as to the profile of PPD. The age range was also wider than originally proposed, i.e. between ages 8 and 10 years, as opposed to 8 years old only. Regarding one cross-sectional comparison, however, the wide range was possibly offset by the age-matched sample, and for each pair, testing that took place during the same season of the school year (i.e. fall, winter or spring). To make further generalizations about

MSW production in school-aged children, future studies should extend to other geographic regions and multilingual contexts.

Next, concerning limitations of the MSW sample, there were inconsistencies as to morphemic composition, i.e. mono- or multimorphemic. Words were sometimes compound nouns (i.e. *watermelon*, *cash register*), verb derivations (e.g. *invitation*), or were inflected verbs or pluralized nouns (e.g. *explodes*, *balloons*). Inclusion of these words, however, provided opportunities to observe the interactions of more phonological components within the nonlinear hierarchy. It was presumed that children were not yet aware of derivational relationship between some of the words. A final consideration limiting the word elicitation validity was of priming effects. Priming was possible as a result of the immediate or preceding word elicitation prompts, or from a preceding word in the elicitation order. In future, the influence of priming effects could be considered in relation to variability in MSW production, but in general, attention to reducing such effects should contribute to study designs.

The final limitations for this discussion are with respect to the MSW metric. The first of these concerns decision-making about lexical influences. Specific guidelines about distinguishing the effects of word neighborhood activations from those of the phonological system were not developed, potentially decreasing tally reliability. The intra-class correlations (ICC) indicated high coding reliability overall, however, reliability of lexical coding was not specifically calculated because the proportion of tallies was small. Larger studies are needed to consider the confound of lexical and phonological tallies, because the respective subtotals require independence in order to be additive. For example, for *guitar* produced as [dɪgt^{h}ɑɹ], the overlap of phonological mismatches is in question, i.e. insertion of a consonant timing unit and segmental features, and in addition, metathesis of /g/. Nevertheless, although the most infrequent, lexical tallies appeared to have merit for two reasons: (a) in the pilot study (Chapter 3), the contribution to rank order distinctions; and (b) for the main study, the significant differences between 5-year-olds and 8- to 10-year-olds with and without PPD. There was therefore potential for loss of valuable information as a result of eliminating lexical considerations from the metric. For instance, productions of older children with many lexical tallies might have significance for clinical decision-making. Further examination of lexical effects is required, for monomorphemic, and in addition, for multimorphemic MSWs.

Another apparent limitation of the metric was the strict view taken of tap production, decided because of the *ballistic* nature (Shockey, 2003), or *speed and trajectory* (Ladefoged, 2006) of taps, in comparison with other consonants. The relative difficulty of taps had also been described in terms of articulatory motion (Derrick, 2011) and phonetic context (Klein & Altman, 2002). Taps production was therefore of interest because of the interaction of their underlying phonology with phonetics (as described in Chapter 4). Although a strict view of taps possibly inflated the proportion of mismatches, especially for the 8TD group, the view of taps for the current thesis had merit for observing phonological component interactions. For example, in the sample of MSWs, 5TD children stopped three-quarters of taps, irrespective of word length. However, in spite of syllable adjacency, nasal harmony of tap with /m/ was more likely in a weakly stressed syllable within the word, e.g. *hippopotamus* versus *thermometer*, which differ with respect to the sequences of tap with /m/ in the onsets of the last two syllables. Derrick (2011) suggested that differences in sub-phonemic organization could account for production variability of tap, at least for adult speakers. For children, however, the converse might be true, i.e. that supra-phonemic hierarchical associations are responsible. In order to assess the impact of MSW elements that challenge the timing of phonological output, future studies could examine larger word sets that provide more opportunities for tap production in the various word contexts.

5.6 Strengths of the Thesis

The major strength of the current research is the contribution to the evidence base about MSW phonological production for children of school age. The study further supports the need to include MSWs in phonological evaluations of school-aged children, and for Canadian English-speaking children, provides criterion reference data for productions of a representative set of culturally relevant MSWs. Up to the present, the evidence base concerning MSW phonology for children of school age, has been limited primarily to James' (2006) phonological pattern analyses of Australian-English speaking children with typically developing phonology. The current research was therefore the first to examine MSW phonology in children with a history of PPD. Further to this was the design of an inaugural tool for holistically quantifying MSW production with apparent utility for research and clinical practice.

The evidence base provided by the current thesis came from a combination of sources. The first is a meta-analysis that demonstrated the relationship of MSW phonology to literacy of children with PPD. The second is the establishment of construct validity for a new whole word MSW metric that was based on two types of theories of phonological production: theories of phonology and language processing. For the thesis, current models of nonlinear phonology and parallel interactive processing were relevant. The inclusion of vowels within the metric, with their importance for stress and syllable structure, also allows processing interactions of all linguistic components to be considered (e.g. weak syllables within a foot, tap and syllabic segments), and are necessary for expanding explanations of linguistic timing unit hierarchies. Third, the sensitivity and specificity of the metric was considered in relation to phonological development, i.e. for TD children, pre- and post-early literacy instruction; and in addition, for children with a history of PPD, pre-early literacy, and with ongoing protracted phonology, post-early literacy.

Another strength of the metric is its utility for research and clinical application. That is, a reliable and valid tool, with uniform administration and scoring, is available for holistically quantifying the multiple interactions that occur within MSW productions. Another advantage is that such quantification provides a continuous scale of measurement, and therefore, more power to detect change than arbitrary categorizations (Peterson et al., 2009). The whole word total has the ability to detect weaknesses and track progress in global MSW phonological production, while the subtotals can be used to examine sub-components of MSW processing, i.e. lexical influences, word structure, and segmental features. In spite of the many word-specific interactions in MSWs, the subtotals might inform phonological patterns for a sub-group of words with certain components in common. The information gained from the various sums within the metric is therefore applicable to intervention planning for children with PPD, and in addition, to future studies of MSW phonological development.

As a result of the numerous complexities involved in examining MSW production, considerable data remain for analysis with respect to the coded feature mismatches, e.g. for manner, place and laryngeal categories, including variability data. In addition, analyses of literacy data collected for the assessment protocol will be a future endeavor.

In conclusion, for school-aged children with a history of PPD, MSWs are a necessary component of evaluation for research and clinical practice. Otherwise, children may be excluded from receiving appropriate integrated phonological and literacy intervention that will enhance their success at school and their options in life more broadly.

Bibliography

References marked with an asterisk indicate studies included in the meta-analysis.

Adams, J., & Weakliem, D. L. (2011). August B. Hollingshead's Four Factor Index of Social Status : From unpublished paper to citation classic. *Yale Journal of Sociology, 8,* 11-19. Retrieved from http://www.yale.edu/sociology/yjs/yjs_fall_2011.pdf, 2011, September 30.

American Speech-language-hearing Association (ASHA) (2013). *Speech-language disorders and the speech-language pathologist.* [Web site posting]. Retrieved from http://www.asha.org/careers/professions/sld.htm, 2013, May 31.

*Apel, K., & Lawrence, J. (2011). Contributions of morphological awareness skills to word-level reading and spelling in first-grade children with and without speech sound disorder. *Journal of Speech, Language, and Hearing Research, 54,* 1312-1327.

Apel, K., Wolter, J. A., & Masterson, J. J. (2006). Effects of phonotactic and orthotactic probabilities during fast mapping on 5-year-olds' learning to spell. *Developmental Neuropsychology, 29,* 21-42.

Arias, J., & Lleo, C. (2013). Rethinking assessment measures of phonological development and their application in bilingual acquisition. *Clinical Linguistics and Phonetics, 28,* 153-175.

Augustine, J. M., Cavanagh, S. E., & Crosnoe, R. (2009). *Maternal education, early child care and the reproduction of advantage*. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2910916/, 2011, September 30.

Baath, R. (2010). ChildFreq: An online tool to explore word frequencies in child language. *LUCS, Minor, 16.* Retrieved from http://childfreq.sumsar.net/, 2014, March 15.

Babayağit, S. (2009). Reading and spelling development in transparent alphabetic orthographies: Points of convergence, divergence and arising issues. In C. Wood & V. Connelly (Eds.), *Contemporary perspectives on reading and spelling* (pp. 133-148). New York, NY: Routledge.

Bahr, R., Silliman, E. R., & Berninger, V. (2009). What spelling errors have to tell about vocabulary learning. In C. Wood & V. Connelly (Eds.), *Contemporary perspectives on reading and spelling* (pp. 109-130). New York, NY: Routledge.

Baker, E., Croot, K., McLeod, S., & Paul, R. (2001). Psycholinguistic models of speech development and their application to clinical practice. *Journal of Research, 4,* 685-702.

Battacharya, A., & Ehri, L. C. (2004). Graphosyllabic analysis helps adolescent struggling readers read and spell words. *Journal of Learning Disabilities, 37,* 331-348.

Bernhardt, B. H., & Stemberger, J. P. (2012). Translation to practice: Transcription of the speech of multilingual children. In N. Muller, & M. Ball (Eds.), *Multilingual aspects of speech sound disorders in children* (pp. 182-190). North York, ON, Canada: Multilingual Matters.

Bernhardt, B. H., & Stemberger, J. P. (1998). *Handbook of phonological development: From the perspective of constraint-based nonlinear phonology*. San Diego, CA: Academic Press.

Bernhardt, B. May, Stemberger, J. P., & Charest, M. (2010). Intervention for speech production in children and adolescents: Models of speech production and therapy approaches. Introduction to the issue. *Canadian Journal of Speech-Language Pathology and Audiology, 34,* 157-167.

Berko, J. (1958). The child's learning of English morphology. *Word, 14,* 150-177.

Bird, J., & Bishop, D. (1992). Perception and awareness of phonemes in phonologically impaired children. *European Journal of Disorders of Communication, 27,* 289-311.

*Bird, J., Bishop, D. V., & Freeman, N. H. (1995). Phonological awareness and literacy development in children with expressive phonological impairments. *Journal of Speech and Hearing Research, 38*, 446-462.

*Bishop, D. V. M., & Clarkson, B. (2003). Written language as a window into residual language deficits: A study of children with persistent and residual speech and language impairments. *Cortex, 39,* 215-237.

Blachman, B. (2000). Phonological awareness. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research: Volume III* (pp. 483–502). Mahwah, NJ: Erlbaum.

Blishen, B. R., Carroll, W. K., & Moore, C. (1987). The 1981 socioeconomic index for occupations in Canada. *Canadian Review of Sociology and Anthropology, 24*, 465-487.

Bowey, J. A. (2005). Predicting individual differences in learning to read. In M. J. Snowling & C. Hulme (Eds.), *The Science of Reading: A handbook. Blackwell handbooks of developmental psychology* (pp. 135-154). Malden: Blackwell.

British Columbia Ministry of Education (2006). *English language arts curriculum: Kindergarten to Grade 7.* Retrieved from https://www.bced.gov.bc.ca/irp/course.php?lang=en&subject=English_Language_Arts&course=English_Language_Arts_Kindergarten_to_Grade_7&year=2006 2015, September 1.

*Burrus, M. (2007). *Spelling skills in children with phonological disorders* (Unpublished master's thesis). Missouri State University.

Cain, K. (2009). Children's reading comprehension difficulties: A consideration of the precursors and consequences. In C. Wood, & V. Connelly (Eds.), *Contemporary perspectives on reading and spelling* (pp. 59-75). New York, NY: Routledge.

Canadian Association of Speech-language Pathologists & Audiologists (2005). *Fact Shee*t. [Web site posting]. Retrieved from http://www.caslpa.ca/PDF/fact%20sheets/speechhearingfactsheet.pdf, 2011, September 30.

Carneiro, P., Costas, M., & Parey, M. (2011). *Maternal education, home environments and the development of children and adolescents.* Retrieved from http://www.ucl.ac.uk/~uctppca/MS2307_revised_manuscript.pdf, 2011, September 30.

*Carroll, J. M., & Snowling, M. J. (2004). Language and phonological skills in children at high risk of reading difficulties. *Journal of Psychology and Psychiatry, 45,* 631-640.

*Catts, H. W. (1991). Early identification of dyslexia: Evidence from a follow-up study of speech-language impaired children. *Annals of Dyslexia, 41,* 163-177.

*Catts, H. W. (1993). The relationship between speech-language impairments and reading disabilities. *Journal of Speech and Hearing Research, 36*, 948-959.

Catts, H. W. (1996). The relationship between speech-language impairments and reading disabilities. *Journal of Learning Disabilities, 19*, 504-508.

Catts, H. W., Fey, M. E., Tomblin, J. B., & Zhang, X., (2002). A longitudinal investigation of reading outcomes in children with language impairments. *Journal of Speech, Language, and Hearing Research, 45*, 1142-1157

Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (2001). Estimating the risk of future reading difficulties in kindergarten children: A research-based model and its clinical implementation. *Language, Speech, and Hearing Services in Schools*, *32,* 38-50.

*Chabon, R. S. (1980). *A comparison of normal and articulation impaired children's performance on four measures of phonemic awareness* (Unpublished doctoral dissertation). University of Pittsburgh. Pennsylvania.

Chall, J. (1983). *Stages of reading development.* New York: McGraw-Hill.

Cohen, J., Cohen, P., West, G. S., & Aiken, L. S. (2003). *Applied multiple regression/correlational analysis for the behavioral sciences.* Mahwah, NJ: Lawrence Erlbaum.

Coltheart, M. (2005). Modeling reading: The dual-route approach. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook. Blackwell handbooks of developmental psychology,* (pp. 6-23). Malden: Blackwell.

Compton, D.L. (2002). The relationships among phonological processing, orthographic processing, and lexical development in children with reading disabilities. *Journal of Special Education, 35,* 201-210.

*Core, C. W. (2004). *Phonological awareness skills in kindergarten children with and without phonological impairment* (Unpublished doctoral dissertation). University of Florida.

Cowan, N. (1996). Short-term memory, working memory, and their importance in language processing. In K. G. Butler, & R. B. Gillam (Eds.). *Working memory and language impairment: New perspectives, Topics in Language Disorders, 17,* 1-18.

*Cowan, W. E., & Moran, M. J. (1997). Phonological awareness skills of children with articulation disorders in Kindergarten to third grade. *Journal of Children's Communication Development, 18,* 31-38.

Critten, S., Pine, K. J., & Steffler, D. (2007). Spelling development in young children: A case of representational redescription? *Journal of Educational Psychology, 99,* 207-220.

Critten, S., & Pine, K. J. (2009). Viewing spelling in a cognitive context: Underlying representations and processes. In C. Wood, & V. Connelly (Eds.), *Contemporary perspectives on reading and spelling* (pp. 92-108). New York, NY: Routledge.

Cunningham, P. (1998). The multisyllabic word dilemma: Helping students build meaning, spell, and read "big" words. *Reading and Writing Quarterly: Overcoming Learning Disabilities, 14,* 189-219.

Decker, G., Simpson, G. B., Yates, M., & Locker, L. Jr. (2003). Flexible use of lexical and sublexical information in word recognition. *Journal of Research in Reading, 26,* 280-286.

Dell, G. S. & Kittredge, A. (2010). Connectionist models of aphasia and other language impairments. In J. Guendouzi, F. Loncke, & M. J. Williams (Eds.), *The handbook of psycholinguistic and cognitive processes* (pp. 169-188). New York, NY: Psychology Press.

Demuth, K., & Johnson, M. (2003). Truncation to subminimal words in early French. *Canadian Journal of Linguistics, 48*, 211-241.

Deonandan, R., Campbell, K., Ostbye, T., Tummon, I., & Robertson, J. (2000). A comparison of methods for measuring socio-economic status by occupation or postal area. *Chronic Diseases in Canada, 21,* Retrieved from http://www.ruor.uottawa.ca/en/bitstream/handle/10393/20369 /lcdce.pdf?sequence=1, 2011, September 30.

Derrick, D. (2011). *Kinematic patterning of flaps, taps and rhotics in English* (Unpublished doctoral dissertation). The University of British Columbia, Canada.

Dodd, B., Hua, Z., Crosbie, S., Holm, A., & Ozanne, A. (2002). *Diagnostic Evaluation of Articulation and Phonology (DEAP).* San Antonio, TX: Pearson.

Dodd, B., Crosbie, S., MacIntosh, Teitzel, T. & Ozanne, A. (2003). *Pre-Reading Inventory of Phonological Awareness (PIPA)*. San Antonio, TX: Pearson.

Doidge, N, (2007). *The brain that changes itself.* NY, NY: Viking Press.

Dole, J., Brown, K., & Trathen, W. (1996). The effects of strategy instruction on the comprehension performance of at-risk students. *Reading Research Quarterly, 31*, 62-88.

Dollaghan, C. A., & Horner, E. A. (2011). Bilingual language assessment: A meta-analysis of diagnostic accuracy. *Journal of Speech, Language, and Hearing Research, 54,* 1077-1088.

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.

Dubow, E. F., Boxer, P., & Hesmann, L. R. (2009). Long-term effects of parents' education on children's educational and occupational success: Mediation by family interactions, child aggression, and teenage aspirations. *Merrill Palmer Quarterly, 55,* 224-249. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2853053/ 2011, September 30.

Duncan, L. G. (2009). Acquiring complex reading skills: An exploration of disyllabic reading. In C. Wood, & V. Connelly (Eds.), *Contemporary perspectives on reading and spelling* (pp. 42-58). New York, NY: Routledge.

Duncan, L. G., & Seymour, H. K. (2003). How do children read multisyllabic words? Some preliminary observations. *Journal of Research in Reading, 26,* 101-120.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test (4th ed) (PPVT-4).* Toronto, ON: Psycan.

Dunn, L. M., Thériault-Whalen, C. M., & Dunn, L. M. (1993). *Échelle de vocabulaire en image Peabody. [Peabody picture vocabulary test].* Toronto, ON: Psycan.

Durost, W.N., Bixler, H.H., Hildreth, G.H., Lund, K.W., & Wrightstone, J.W. (1959). *Metropolitan Achievement Tests.* New York, NY: Harcourt Brace and World.

Ehri, L. C. (1992). Reconceptualizing the development of sight word reading and its relationship to recoding. In P. B. Gough, L. C., Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 107-143). Hillsdale, NJ: Lawrence Erlbaum.

Ehri, L. C. (2005). Development of sight word reading: Phases and findings. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook. Blackwell handbooks of developmental psychology* (pp. 135-154). Malden: Blackwell.

Eisenberg, S. L., & Hitchcock, E. R. (2010). Using standardized tests to inventory consonant and vowel production: A comparison of 11 tests of phonology. *Language, Speech, and Hearing Services in Schools, 41,* 488-503.

*Ekelman, B. L. (1993). *Examining the relationships among speech-language and reading skills in children with a history of speech-language or reading disorders* (Unpublished doctoral dissertation). Case Western Reserve University. Ohio.

Elman, J. L. (1999). The emergence of language: A conspiracy theory. In B. MacWhinney (Ed.), *The emergence of language* (pp. 1-28). Mahwah, NJ: Lawrence Erlbaum.

Fey, M. E., Catts, H. W., & Larrivee, L. S. (1995). Preparing preschoolers for the academic and social challenges of school. In M. E. Fey, J. Windsor, & S. F. Warren (Eds.), *Language intervention: Preschool through the elementary years* (pp. 3-38). Baltimore, MD: Paul H. Brookes.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). Thousand Oaks, CA: SAGE.

Fikkert, P., Freitas, M. J., Grijzenhout, J., Levelt, C. & Wauquier S. (2004, April). *Syllabic markedness, segmental markedness, rhythm and acquisition*. Paper presented at conference of Generative Linguistics in the Old World, Thessaloniki, Greece.

Foy, J. G., & Mann, V. A. (2012). Speech production deficits in early readers: Predictors of risk. *Reading and Writing, 25,* 799-830.

Friend, A., & Olson, R. K. (2008). Phonological spelling and reading deficits in children with spelling disabilities. *Scientific Studies of Reading, 12,* 90-105.

Frith, U. (2002). Resolving the paradoxes of dyslexia. In G. Reid & J. Wearmouth (Eds.), *Dyslexia and literacy: Theory and practice* (pp. 45-68). West Sussex, UK: John Wiley and Sons.

Gierut, J. A., & Dale, R. A. (2007). Comparability of lexical corpora: Word frequency in phonological generalization. *Clinical Linguistics and Phonetics, 21,* 423-433.

Gierut, J. A., & Morrisette, M. L. (2012). Age of word acquisition effects in treatment of children with phonological delays. *Applied Psycholinguistics, 33,* 121-144.

Goldman, R., & Fristoe, M. (1969). *The Goldman Fristoe Test of Articulation (GFTA).* Circle Pines, MN: American Guidance Service.

Goldman, R., & Fristoe, M. (1986). *The Goldman Fristoe Test of Articulation (2nd ed.) (GFTA-2).* Circle Pines, MN: American Guidance Service.

Goldman, R., Fristoe, M., & Woodcock, R. W. (1974). *The Goldman Fristoe Woodcock Sound Symbol Tests (GFWSST).* Circle Pines, MN: American Guidance Service.

Gravetter, F., & Wallnau, L. (2007). *Statistics for the behavioral sciences*. Belmont, CA: Wadsworth.

Hall, R., Adams, C., Hesketh, A., & Nightingale, K. (1998). The measurement of intervention effects in developmental phonological disorders. *Journal of Language and Communication Disorders, 33 Supplement,* 445-450.

*Harris, J., Botting, N., Myers, L., & Dodd, B. (2011). The relationship between speech impairment, phonological awareness and early literacy development. *Australian Journal of Learning Disabilities, 16,* 111-125.

Harris, John (2007). Representation. In P. de Lacy (Ed.), *The Cambridge handbook of phonology* (pp. 119-138). Cambridge, UK: Cambridge University Press.

Hatcher, P. J., Hulme, C., & Snowling, M. J. (2002). The phonological representations hypothesis of dyslexia: from theory to practice. In G. Reid & Wearmouth (Eds.), *Dyslexia and literacy: Theory and practice* (pp. 69-84). West Sussex, UK: John Wiley & Sons.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Thousand Oaks, CA: Sage.

Hermann, J. A., Matyas, T., & Pratt, C. (2006). Meta-analysis of the nonword reading deficit in Specific Reading Disorder. *Dyslexia, 12,* 195-221.

Hester, C., & Hodson, B.W. (2004). The role of phonological representation in decoding skills of young readers. *Child Language Teaching and Therapy, 20*, 115-133.

Hollingshead, A. B. (1975). *Four factor index of social status* (Unpublished manuscript). Yale University.

*Holm, A., Farrier, F., & Dodd, B. (2008[1]). Phonological awareness, reading accuracy and spelling ability of children with inconsistent phonological disorder. *International Journal of Language and Communication Disorders, 43*, 300-302.

*Holm, A., Farrier, F., & Dodd, B. (2008[2]). Phonological awareness, reading accuracy and spelling ability of children with inconsistent phonological disorder. *International Journal of Language and Communication Disorders, 43*, 300-302.

Howell, D. C. (2002). *Fundamental statistics for the behavioral sciences (5th ed.).* Belmont, CA: Wadsworth Thomson Learning.

Hunter, J. S., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.

Ingram, D. (2002). The measurement of whole-word productions. *Journal of Child Language, 29,* 713-733.

James, D. G. H. (2006). *Hippopotamus is so hard to say: Children's acquisition of polysyllabic words* (Unpublished doctoral dissertation). University of Sydney, Australia.

James, D., van Doorn, J., & McLeod, S. (2001). Vowel production in mono-, di- and polysyllabic words in children aged 3;0 to 7;11 years. In L. Hewat, & S. Wilson (Eds.), *Proceedings of the 2001 Speech Pathology Australia National Conference: Evidence and Innovation* (pp. 127-135). Melbourne: Speech Pathology Australia.

James, D., van Doorn, J., & McLeod, S. (2008). The contribution of polysyllabic words in clinical decision making about children's speech. *Clinical Linguistics and Phonetics, 22,* 345-353.

Kager, R. (2007). Feet and Metrical Stress. In P. de Lacy (Ed.), *The Cambridge handbook of phonology* (pp. 195-228). Cambridge, UK: Cambridge University Press.

Kehoe, M. M. (2001). Prosodic patterns in children's multisyllabic word productions. *Language, Speech, and Hearing Services in Schools, 32,* 284-294.

Kemp, N. (2009). The acquisition of spelling patterns: Early, late or never? In C. Wood, & V. Connelly (Eds.) *Contemporary perspectives on reading and spelling* (pp. 76-92). New York, NY: Routledge.

Kemp, N. (2006). Children's spelling of base, inflected, and derived words: Links with morphological awareness. *Reading and Writing, 19,* 737-765.

Khan, L., & Lewis, N.P. (2002). *Khan-Lewis Phonological Analysis (2nd ed.) (KLPA-2).* San Antonio, TX: Pearson.

Kintsch, W., & Rawson, K. A. (2005). Comprehension. In M. Snowling, & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 735-742). Oxford: Blackwell.

*Kirk, C., & Gillon, G. T. (2007). Longitudinal effects of phonological awareness intervention on morphological awareness in children with speech impairment. *Language, Speech, and Hearing Services in Schools, 38*, 342-352.

Klein, H. B., & Altman, E. K. (2002). The acquisition of medial /t, d/ allophones in bisyllabic contexts. *Clinical Linguistics and Phonetics, 16*, 215-232.

Klein, H. B., & Spector, C. C. (1985). Effect of syllable stress and serial position on error variability in polysyllabic productions of speech-delayed children. *Journal of Speech and Hearing Disorders, 50,* 391-402.

Kurtosis. *How to draw a funnel plot.* Retrieved, from http://www.kurtosis.co.uk/index.htm, 2012, January 15.

Ladefoged, P. (2006). *A course in phonetics (5th ed.).* Boston, MA: Thomson Wadsworth.

Larrivee, L. S., & Catts, H. W. (1999). Early reading achievement in children with expressive phonological disorders. *American Journal of Speech-Language Pathology, 8,* 118-128.

Larsen, S.C., Hammill, D.D., Moats, L.C., (1999). *Test of Word Spelling (4th ed.) (TWS-4).* Austin, TX: Pro-Ed.

Lee, C. H. (2007). Phonological activation in multi-syllabic word recognition. *Journal of Psychological Research, 36*, 1-14.

Lee, C. J. L. (2003). Evidence-based selection of word frequency lists. *Journal of Speech-Language Pathology and Audiology, 27,* 170-173.

Leitão, S., & Fletcher, J. (2004). Literacy outcomes for students with speech impairment: Long-term follow-up. *International Journal of Language and Communication Disorders, 39,* 245-256.

Lehtonen, A. (2009). How do children and adults conceptualize phoneme-letter relationships? In C. Wood & V. Connelly (Eds.), *Contemporary perspectives on reading and spelling* (pp. 149-167). New York, NY: Routledge.

Leonard, L. B. (1992). Models of phonological development and children with phonological disorders. In C. A. Ferguson, L. Menn & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 495-508).Timonium, MD: York.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). Lexical access in speech production. *Behavioral and Brain Sciences, 22,* 1-75.

Lewis, B. A. (2007). *Short- and long-term outcomes for children with speech sound disorders.* Retrieved from www.cllrnet.ca, 2007, March 20.

*Lewis, B. A., & Freebairn, L. A. (1992). Residual effects of preschool phonology disorders in grade school, adolescence and adulthood. *Journal of Speech and Hearing Research, 35,* 819-832.

Lewis, B. A., Freebairn, L. A., & Taylor, H. G. (2000[1]). Academic outcomes in children with histories of speech sound disorders. *Journal of Communication Disorders, 33*, 89-92.

Lewis, B. A., Freebairn, L. A., & Taylor, H. G. (2000[2]). Follow-up of children with early expressive phonology disorders. *Journal of Learning Disabilities, 33,* 433-445.

Lewis, B. A., Freebairn, L. A., Hansen, A. J., Iyengar, S. K., & Taylor, H. G. (2004). School-age follow-up of children with childhood apraxia of speech. *Language, Speech, and Hearing Services in Schools, 35,* 122-140.

Lewis, B. A., Freebairn, L. A., Hansen, A. J., Stein, C. M., Shriberg, L. D., Iyengar, S. K., & Taylor, H. G. (2006). Dimensions of early speech sound disorders: A factor analytic study. *Journal of Communication Disorders, 39,* 139-157.

Lipka, O., Lesaux, N. K., & Siegel, L. S. (2006). Retrospective analyses of the reading development of grade four students with reading disabilities: Risk status and profiles over 5 years. *Journal of Learning Disabilities, 39*, 364-378.

Lippke, B. A., Dickey, S.E., Selmar, J. W., & Soder, A. L. (1997). *Photo Articulation Test (3rd ed.) (PAT-3).* Austin, TX: Pro-Ed.

Lipsey, M. W. & Wilson, D. B. (2000). *Practical Meta-analysis.* Thousand Oaks, CA: Sage.

Lyon, G.R. (1998). The NICHD research program in reading development, reading disorders and reading instruction: A summary of research findings. *Keys to Successful Learning: A National Summit on Research in Learning Disabilities, 1-9,* National Center for Learning Disabilities, Inc., New York, NY.

MacLeod, A. N., Laukys, K., & Rvachew, S. (2011). The impact of bilingual language learning on whole-word complexity and segmental accuracy among children aged 18 and 36 months. *International Journal of Speech-Language Pathology, 13,* 490-499.

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

Marantz PMD660 [Portable Solid State Recorder]. (1997). Kleinburg, Ontario, Canada: Marantz, Canada, LLC.

Mason, G. K., & Bernhardt, B. M. (2014). The impact of protracted phonological disorders on literacy outcomes in children: A meta-analysis. In P. A. Ysunza (Ed.), *Speech, language and voice pathology: Methods, challenges and outcomes* (pp. 49-116). New York: Nova.

Mason, G. K., Bernhardt, B. M., & Masterson, J. J. (2010, May). *Variability in the production of multisyllabic words in a group of five-year-old children: Is knowing how to say phonemes all there is to it?* Poster session presented at the annual conference of the Canadian Association of Speech-language Pathologists and Audiologists, Whitehorse, Yukon.

Mason, G. K., Bérubé, D., Bernhardt, B. M., & Stemberger, J. P. (2015). Evaluation of multisyllabic word production in Canadian English- and French-speaking children within a nonlinear phonological framework. *Clinical Linguistics and Phonetics, 29*(8-10), 666-685. doi 10.3109/02699206.2015.1040894.

Masterson, J., & Bernhardt, B. (2001). *Computerized Articulation and Phonology Evaluation System (CAPES).* Toronto: The Psychological Corporation.

Masterson, J. J., Bernhardt, B. H., & Hofheinz, M. K. (2005). A comparison of single words and conversational speech in phonological evaluation. *American Journal of Speech-Language Pathology, 14,* 229-241.

McCauley, R. J., & Hargrove P. (2004). A clinician's introduction to systematic reviews in communication disorders: The course review paper with muscle. *Contemporary Issues in Communication Science and Disorders, 31,* 173–181.

McLeod, S. (2009). Speech sound acquisition. In J. E. Bernthal, N. W. Bankson & P. Flipsen Jr. (Eds.), *Articulation and phonological disorders: Speech sound disorders in children* (pp. 63-119). Boston, MA: Pearson Education.

McCormack, J., McLeod, S., Harrison, L., & McAllister, L. (2007). *Prevalence of Communication Impairment in Australian Children*. Poster presented at the meeting of the American Speech-language-hearing Association, Boston, MA. Retrieved from http://www.csu.edu.au/__data/assets/pdf_file/0015/154014/poster-1.pdf, 2011, May 31.

McDougall, S., & Hulme, C. (1994). Short-term memory speech rate, and phonological awareness as predictiors of learning to read. In C. Hulme, & M. Snowling (Eds.), *Reading development and dyslexia* (pp. 31-44). London: Whurr.

*McDowell, K. D. (2004). *Examining relations among expressive phonology, phonological processing, and early decoding skills in children* (Unpublished doctoral dissertation). Florida State University.

McLeod, S. (2009). Speech sound acquisition. In J. E. Bernthal, N. W. Bankson, & P. Flipsen Jr. (Eds.), *Articulation and phonological disorders: Speech sound disorders in children (*pp. 63-120). Boston, MA: Pearson Education Inc.

Menn, L., & Matthei, E. (1992). The two-lexicon account of child phonology: Looking back, looking ahead. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: models, research, implications* (pp. 211-247). Timonium, MD: York Press.

*Meredith, L. N. (2002). *Spelling, reading, and phonological processing abilities of children with histories of phonological disorders* (Unpublished doctoral dissertation). University of Memphis. Tennessee.

Metsala, J. L., Stanovich, K. E., & Brown, G. D. A. (1998). Regularity effects and the phonological deficit model of reading disabilities: A meta-analytic review. *Journal of Educational Psychology, 90,* 279-293.

Morgenstern, A., Caet, S., Collombel, M., Parisse, C., Sekali, M., & Yamagushi, N. (2010, June). *From buds to flowers: The blossoming of child language and multimodal analyses.* Paper presented at PhonBank workshop, St. John's, NL, Canada.

Morissette, M. L. (1999). Lexical characteristics of sound change. *Clinical Linguistics and Phonetics, 13,* 219-238.

Morrison, C. M., Chappell, T. D., & Ellis, A. W. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *The Quarterly Journal of Experimental Psychology, 50,* 528-559.

Muter, V. (1994). The influence of phonological awareness and letter knowledge on beginning reading and spelling development. In C. Hulme, & M. Snowling (Eds.), *Reading development and dyslexia* (pp. 45-62). London: Whurr.

Murnane, R., Sawhill, I., & Snow, C. (2012). Literacy challenges for the twenty-first century: Introducing the issue. *The Future of Children, 22,* 3-15.

*Nathan, L., Stackhouse, J., Goulandris, N., & Snowling, M.J. (2004). Educational consequences of developmental speech disorder: Key Stage 1 National Curriculum Assessment results in English and Mathematics. *British Journal of Educational Psychology, 74,* 173-186.

Nation, K. (2005). Children's reading comprehension difficulties. In M. Snowling, & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 248-266). Oxford: Blackwell.

Newbold, E. J., Stackhouse, J., & Wells, B. (2013). Tracking change in children with severe and persisting speech difficulties. *Clinical Linguistics and Phonetics, 27,* 521-539.

Norton, E. S., Kovelman, I., & Petitto, L. (2007). Are there separate neural systems for spelling? New insights into the role of rules and memory in spelling from functional magnetic resonance imaging. *Mind, Brain and Education, 1,* 48-59.

O'Connor, R.E., Bell, K., Harty, D., Kappel, L., & Sackor, S. (2000). The influence of test difficulty on children's reading comprehension growth. *Paper presented at the Annual Meeting of the American Educational Research Association*, *1-12,* New Orleans, LA.

*Overby, M. S., Trainin, G., Smit, A. B., Bernthal, J. E., & Nelson, R. (2012). Relationships among speech sound perception, speech sound production, and phonological spelling in second grade children. *Language, Speech, and Hearing Services in Schools, 43*, 97-115.

Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In M. Snowling, & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 227-247). Oxford: Blackwell.

Peterson, R. L., McGrath, L. M., Smith, S. D., & Pennington, B. F. (2007). Neuropsychology and genetics of speech, language and literacy disorders. *Pediatric Clinics of North America, 54,* 543-561.

*Peterson, R. L., Pennington, B. F., Shriberg, L., & Boada, R. (2009). What influences literacy outcome in children with speech sound disorder? *Journal of Speech, Language, and Hearing Research, 52*, 1175-1188.

Pierrehumbert, J.B. (2006). The next toolkit. *Journal of Phonetics, 34,* 516-530.

Presson, N., & MacWhinney, B. (2010). The competition model and language disorders. In J. Guendouzi, F. Loncke, & M. J. Williams (Eds.), *The handbook of psycholinguistic and cognitive processes* (pp. 31-48). New York, NY: Psychology Press.

*Preston, J., & Edwards, M. L. (2007). Phonological processing skills of adolescents with residual speech sound errors. *Language, Speech, and Hearing Services in Schools, 38,* 297-308.

Preston, J. L., Ramsdell, H. L., Oller, D. K., Edwards, M. L., & Tobin, S. J. (2011). Developing a weighted measure of speech sound accuracy. *Journal of Speech, Language, and Hearing Research, 54,* 1-18.

*Puranik C. S., Petscher, Y., Al Otaiba, S., Catts, H. W., & Lonigan, C. J. (2008). Development of oral reading fluency in children with speech or language impairments: A growth curve analysis. *Journal of Learning Disabilities, 41,* 545-560.

*Raitano, N. A., Pennington, B. F., Tunick, R. A., Boada, R., & Shriberg, L. D. (2004). Pre-literacy skills of subgroups of children with speech sound disorders. *Journal of Child Psychiatry and Psychology, 45,* 821-835.

Robertson, C., & Salter, W. (2007). *Phonological Awareness Test (2nd ed.) (PAT-2).* Austin, TX: Linguisystems, Pro-Ed.

Robey, R., & Dalebout, S. (1998). A tutorial on conducting meta-analyses of clinical outcome research. *Journal of Speech, Language, and Hearing Research, 41,* 1227-1241.

Romani, C., Olson, A., & Di Betta, A. M. (2005). Spelling disorders. In C. Hulme & M. J. Snowling (Eds.), *The science of reading: A handbook* (pp. 431-448). London: Wiley-Blackwell.

Rose, Y. (2000). *Headedness and prosodic licensing in the L1 acquisition of phonology* (Unpublished doctoral dissertation). McGill University, Canada.

Rose, Y., & dos Santos, C. (2008). Stress domain effects in French phonology and phonological development. In K. Arregi, Z. Fagyal, S. Montrul, & A. Tremblay (Eds.), *Selected papers from the 38th linguistic symposium on Romance languages* (pp. 89-106). Amsterdam: Johns Benjamins.

Rose, Y., & MacWhinney, B. (2014). The PhonBank initiative. In J. Durand, U. Gut, & G. Kirstoffersen (Eds.), *Handbook of corpus phonology* (pp. 380-401). Oxford: Oxford University Press.

Rose, Y., MacWhinney, B., Byrne, R., Hedlund, G., Maddocks, K., O'Brien, P., & Wareham, T. (2005). Phon: A computerized database for research in phonological acquisition. 10th International Congress for the Study of Child Language. Berlin, July.

Rosenthal, R. (1991). *Meta-analytic Procedures for Social Research.* Thousand Oaks, CA: Sage.

*Rvachew, S. (2007). Phonological processing and reading in children with speech sound disorders. *American Journal of Speech-Language Pathology, 16,* 260-270.

Rvachew, S., & Brosseau-Lapre, F. (2012). *Developmental phonological disorders: Foundations of clinical practice.* San Diego, CA: Plural.

Rvachew, S., Marquis, A., Brosseau-Lapré, F., Paul, M., Royle, P., & Gonnerman, L. (2013). Speech articulation performance of francophone children in the early school years: Norming of the *Test de Dépistage Francophone de Phonologie* [*Screening Test of French Phonology*]. *Clinical Linguistics & Phonetics, 27*, 950-968.

*Rvachew, S., Ohberg, A., Grawburg, M., & Heyding, J. (2003). Phonological awareness and phonemic perception in 4-year-old children with delayed expressive phonology skills. *American Journal of Speech-Language Pathology, 12,* 463-471.

Sander, E. (1972). When are speech sounds learned? *Journal of Speech and Hearing Disorders, 37,* 55-63.

Schonell, F., & Schonell, F. (1956). *Schonell Graded Word Spelling Test*. London, UK: Oliver & Boyd.

Schretlen, C. (2013). *Prosodic structure patterns in multisyllabic word productions of Granada Spanish-speaking children with typical versus protracted phonological development* (Unpublished master's thesis). The University of British Columbia, Canada.

Schuele, C. M. (2004). The impact of developmental speech and language impairments on the acquisition of literacy skills. *Mental Retardation and Developmental Disabilities Research Reviews, 10,* 176-183.

Semel, E., Wiig, E. H., & Secord, W. (2003). *Clinical Evaluation of Language Fundamentals (4th ed.) (CELF-4).* San Antonio, TX: Pearson.

Shankweiler, D., Lundquist, E., Katz, L., Stuebing, K.K., Fletcher, J.M., Brady, S., Fowler, A., Dreyer, L., Marchione, K.E., Shaywitz, S.E., & Shaywitz, B.A. (1999). Comprehension and decoding: Patterns of association in children with reading difficulties. *Scientific Studies of Reading, 3*, 69-94.

Shattuck-Hufnagel, S. (1992). The role of word structure in segmental serial ordering. *Cognition, 42,* 213-259.

Shattuck-Hufnagel, S., & Klatt, D. (1979). The limited use of distinctive features and markedness in speech production: Evidence from speech error data. *Journal of Verbal Learning and Verbal Behavior, 18,* 41-55.

Shockey, L. (2003). *Sound patterns of spoken English*. Malden, MA: Blackwell.

Shriberg, L. D. (1994). Five subtypes of developmental phonological disorders. *Clinics in Communication Disorders, 4,* 38-53.

Shriberg, L. D., Austin, D., Lewis, B. A., McSweeny, J., & Wilson, D. L. (1997). The Speech Disorders Classification System (SDCS): Extensions and life span reference data. *Journal of Speech, Language, and Hearing Research, 40,* 708-722.

Shriberg, L. D., & Kwiatkowski, J. (1982). Phonological disorders I: A diagnostic classification system. *Journal of Speech and Hearing Disorders, 47,* 226-241.

Shriberg, L. D., & Morrison, J. (1992). Articulation testing versus conversational speech sampling. *Journal of Speech and Hearing Research, 35,* 259-273.

Siegel, L. S., & Smythe, I. S. (2005). Reflections on research on reading disability with special attention to gender issues. *Journal of Learning Disabilities, 38,* 473-477.

*Silva, P. A., Justin, C., McGee, R., & Williams, S. M. (1984). Some developmental and behavioural characteristics of seven-year-old children with delayed speech development. *British Journal of Disorders of Communication, 19*, 147-154.

Skahan, S. M., Watson, M., & Lof, G. L. (2007). Speech-language pathologists' assessment practices for children with suspected speech sound disorders: Results of a national survey. *American Journal of Speech-Language Pathology, 16,* 246-259.

Smit, A. B. (2007). General American English speech acquisition. In S. McLeod (Ed.), *International guide to speech acquisition* (pp. 128-145). Clifton Park, NY: Thomson Delmar Learning.

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.

Snowling, M. J., & Carroll, J. M. (2004). Language and phonological skills in children at high risk of reading difficulties. *Journal of Child Psychology and Psychiatry, 45,* 631-640.

*Snowling, M. J., & Stackhouse, J. (1983). Spelling performance of children with developmental verbal dyspraxia. *Developmental Medicine and Child Neuropsychology, 25,* 430-437.

*Stackhouse, J. (1982). An investigation of reading and spelling performance in speech disordered children. *British Journal of Disorders of Communication, 17,* 53-60.

Stackhouse, J., & Wells, B. (1997). *Children's speech and literacy difficulties*. London: Whurr.

Stackhouse, J., Vance, M., Pascoe, M., & Wells, B. (2007). *Compendium of speech tasks: Children's speech and literacy difficulties* (pp. 1-26). Sussex, UK: John Wiley & Sons.

Stanovich, K. E. (1992). Speculations on the causes and consequences of individual differences in early reading acquisition. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 307-342). Hillside, NJ: Erlbaum.

Stein, N., & Glenn, C., (1979). An analysis of story comprehension in elementary school children. In R. Freedle (Ed.), *New directions in discourse processing, vol. 2* (pp. 53-120). Norwood, NJ: Ablex.

Stemberger, J. P. (1992). A connectionist view of child phonology: Phonological processing without phonological processes. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development, models, research, implications* (pp. 165-189). Timonium, MD: York Press.

Stoel-Gammon, C. (2010). The Word Complexity Measure: Description and application to developmental phonology and disorders. *Clinical Linguistics and Phonetics, 24,* 271-282.

Stoel-Gammon, C., & Bernhardt, B. M. (2013). Child phonology, phonological theories, and clinical phonology: Reviewing the history of our field. In B. Peter, & A. A. N. McLeod (Eds.), *Comprehensive perspectives on speech sound development and disorders: Pathways from linguistic theory to clinical practice* (pp. 3-30). New York, NY: Nova.

*Sutherland, D., & Gillon, G. T. (2006). Development of phonological representations and phonological awareness in children with speech impairment. *International Journal of Communication Disorders, 42,* 229-250.

Swanson, R. A., & Law, B. D. (1993). Whole-Part-Whole learning model. *Performance Improvement Quarterly, 6,* 43-53.

Templin, M. (2004). *Templin Archive*. Retrieved from http://www.numec.k-state.edu/fshs/templinarchive/, 2012, May 31.

Treiman, R. (1992). The role of intrasyllabic units in learning to read and spell. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 65-106). Hillside, NJ: Erlbaum.

Tunmer, W. E., & Hoover, W. A. (1992). Cognitive and linguistic factors in learning to read. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 175-214). Hillside, NJ: Erlbaum.

Turner, H. M. III, & Bernard, R. M. (2006). Calculating and synthesizing effect sizes. *Contemporary Issues in Communication Science & Disorders, 33,* 42-55.

U.S. National Institute on Deafness & Other Communication Disorders (USNIDCD) (2010). *Statistics on Voice, Speech and Language*. [Web site posting]. Retrieved from http://www.nidcd.nih.gov/health/statistics/pages/vsl.aspx, 2010, September 30.

U. S. National Institute of Child Health and Human Development. National Reading Panel Reports of the Subgroups (USNICHHD) (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction.* Retrieved from http://www.nichd.nih.gov/publications/pubs/nrp/documents/report.pdf, 2010, September 30

Vernon, P. (1977). *Graded Word Spelling Test.* London, UK: Hodder and Stoughton.

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive Test of Phonological Processing (CTOPP).* Austin, TX: PRO-ED.

Wauquier, S. & Yamaguchi, N. (2013). French description in a templatic approach. in M. Vihman, & T. Keren-Portnoy (Eds.), *Readings in phonological development: Templatic approaches, cross-linguistic data* (pp. 317-342). Cambridge: University Press.

*Webster, P. E., & Plante, A. S. (1992). Effects of phonological impairment on word, syllable, and phoneme segmentation and reading. *Language, Speech, and Hearing Services in Schools, 23,* 176-182.

*Webster, P. E., Plante, A. S., & Couvillion, L. M. (1997). Phonologic impairment and prereading: update on a longitudinal study. *Journal of Learning Disabilities, 30*, 365-375.

Wechsler, D. (1992a). *Wechsler Individual Achievement Test (WIAT).* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1992b). *Wechsler Intelligence Scale for Children – revised (WISC-R)* (Anglicised version). New York: Psychological Corporation.

*Wellman, R. L., Lewis, B. A., Freebairn, L.A., Avrich, A. A., Hansen, A. J., & Stein, C.M. (2011). Narrative ability of children with speech sound disorders and the prediction of later literacy skills. *Language, Speech, and Hearing Services in Schools, 42*, 561-579.

Wheeler, D., & Touretzky, D. S. (1997). A parallel licensing model of normal slips and phonemic paraphasias. *Brain and Language, 59,* 147-201.

Willson, V.L., & Rupley, W.D. (1997). A structural equation model for reading comprehension based on background, phonemic, and strategy knowledge. *Scientific Studies of Reading, 1*, 45-63.

Windsor, J. (2000). The role of phonological opacity in reading achievement. *Journal of Speech, Language, and Hearing Research, 43*, 50-61.

Wood, C., Wade-Woolley, L., & Holliman, A. J. (2009). Phonological awareness: Beyond phonemes. In C. Wood & V. Connelly (Eds.), *Contemporary perspectives on reading and spelling* (pp. 7-23). New York, NY: Routledge.

Woodcock, R. W. (1973, 1987, 1998). *Woodcock Reading Mastery Tests (WRMT).* Circle Pines, MN: American Guidance Service.

Woodcock, R. W., & Johnson, M. E. B. (1977). *Woodcock-Johnson Tests of Achievement (W-JTA).* Itasca, IL: Riverside.

Zec, D. (2007). The syllable. In P. de Lacy (Ed.), *The Cambridge handbook of phonology* (pp.161-194). Cambridge, UK: Cambridge University Press.