

Winter 12-21-2018

Teacher Judgment Accuracy in the Domain of Spelling

April Spencer

University of Maine, aprils@maine.edu

Follow this and additional works at: <https://digitalcommons.library.umaine.edu/etd>

 Part of the [Elementary Education and Teaching Commons](#), [Language and Literacy Education Commons](#), and the [Special Education and Teaching Commons](#)

Recommended Citation

Spencer, April, "Teacher Judgment Accuracy in the Domain of Spelling" (2018). *Electronic Theses and Dissertations*. 2984.
<https://digitalcommons.library.umaine.edu/etd/2984>

This Open-Access Thesis is brought to you for free and open access by DigitalCommons@UMaine. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of DigitalCommons@UMaine. For more information, please contact um.library.technical.services@maine.edu.

**TEACHER JUDGMENT ACCURACY
IN THE DOMAIN OF SPELLING**

By

April May Spencer

B.A. University of Maine at Farmington, 1989

M.A. Ohio University, 1991

M.Ed. Ohio University, 1997

A DISSERTATION

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

(in Literacy)

The Graduate School

The University of Maine

December 2018

Advisory Committee:

Janet Spector, Associate Professor Emerita of Special Education, Chair

Victoria Bennett-Armistead, Associate Professor of Literacy

William Dee Nichols, Professor of Literacy

Sara Flanagan, Assistant Professor of Special Education

Julie Cheville, External Graduate Faculty

TEACHER JUDGMENT ACCURACY

IN THE DOMAIN OF SPELLING

By April M. Spencer

Dissertation Advisor: Dr. Janet Spector

An Abstract of the Dissertation Presented
in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy
(in Literacy)

December 2018

This study investigated first and second grade teachers' understanding of their students' skill in spelling words of varying difficulty. Specifically, teachers judged their students' item-by-item performance on a twelve-item spelling test adapted from a commercially available developmental spelling inventory. Results showed that teachers' judgments were accurate 77% of the time, a figure similar to accuracy levels reported in previous teacher judgment research in the area of literacy. The study also included an error analysis task that was designed to assess teachers' knowledge of developmental spelling. Results revealed a significant relationship ($r = .42$) between teacher judgment accuracy and performance on the error analysis task. That is, teachers whose error analyses reflected greater knowledge of developmental spelling were more accurate in judging their students spelling skills than teachers whose error analyses attended less to developmental spelling trends.

ACKNOWLEDGEMENTS

This study reflects an enduring passionate interest in education and in language. I would like to express sincere appreciation to the many who have inspired me to pursue that interest and who have encouraged me to pursue doctoral study. I cannot imagine stronger role models than those I have been so lucky as to encounter along my path. To Dan Gunn, Valerie Huebner, Marilyn Shea, David Scribner, the late Allen Flint, the late Alice Bloom, and to all of my professors at Ohio University and the University of Maine at Orono, thank you for so compellingly emboldening my disbelief in such a thing as overthinking things and affirming that not all role models wear athletic shoes. I would also like to express gratitude to my doctoral committee and to extend special thanks to my advisor, Janet Spector. For the wisdom and expertise she has shared, for the patience I have tested, for the integrity she has modeled, for the persistence she has expected, for the humility she has taught me, and for all the time (including the extension into retirement) she has so graciously granted, I am grateful and indebted.

I am also grateful to the teachers who volunteered to participate in my study. It was a true pleasure meeting all of them. I must also acknowledge the many family and friends who have provided continuous support in the completion of this project even when it has involved sacrificing time that might otherwise have been spent with them. In particular, I must extend my perpetual appreciation to Vivianne for her immeasurable support and patience along the way.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iii
LIST OF TABLES	viii
Chapter	
1. STATEMENT OF THE PROBLEM	
Why Accurate Teacher Judgment Matters	1
A Developmental Lens on Learning	2
Teacher Judgment Studies.....	5
Defining Teacher Judgment	5
Methodological Issues and General Findings	6
A Developmental View of Spelling	8
Purpose of the Study.....	9
2. A REVIEW OF THE LITERATURE	13
Research on Teacher Judgment of Students' Literacy Skills.....	13
Teacher Judgment Research Design Considerations.....	14
Student Characteristics as Moderator Variables	17
Teacher Variables Mediating TJ Accuracy	20
Item-Related Variables That May Moderate Judgment Accuracy.....	23
Empirical Gaps and Questions for Further Research	23
Developmental Aspects of Spelling	24
Related Theories and Findings Regarding Spelling Development	31

Research on Teacher Knowledge.....	34
Direct Measures of Teachers’ Knowledge.....	34
Indirect Assessments of Teachers’ Knowledge	36
Assessment of Teachers’ Knowledge of Spelling	37
3. METHOD.....	43
Procedures.....	43
Sample.....	43
Data Collection	44
Spelling test administration.....	44
Student selection.....	46
Teacher interviews.....	47
Measures.....	49
An Overview of Ganske’s (2013) Developmental Spelling Assessment	49
Details of the Modified Spelling Inventory Used in the Present Study	51
Error Analysis Task	52
Development of the error analysis task.....	52
Administration of the error analysis task.....	55
Scoring the error analysis task – Step One.....	55
Scoring examples.....	59
Scoring the error analysis task – Step two.....	60
Reliability of scoring.....	61

Data Analysis.....	61
4. RESULTS.....	62
Sample.....	62
Teacher Characteristics.....	62
Education and experience.....	63
Familiarity with spelling programs.....	63
Student Characteristics.....	64
Teacher Judgment Accuracy.....	65
The Relationship Between Teacher Characteristics and Judgment Accuracy.....	68
Grade level.....	68
Experience and education.....	68
Familiarity and experience with spelling programs.....	68
The Relationship Between Student Characteristics and Teacher Judgment Accuracy.....	70
The Relationship Between Item Characteristics and Judgment Accuracy.....	70
Error Analysis Task.....	73
The Relationship Between Teacher Characteristics and Developmental Spelling Knowledge.....	73
The Relationship Between Error Analysis Scores and Hit Rate.....	74
5. DISCUSSION.....	75
Teacher Judgment Accuracy.....	76

Factors Impacting Judgment Accuracy	78
Teacher Related Factors	78
Student Related Factors.....	80
Item Related Factors.....	81
Teacher Knowledge and Teacher Judgment	82
Limitations and Implications for Future Research.....	87
Sample of Teachers (add additional issues discussed in oral).....	87
Assessment of Teacher Knowledge	88
Conclusions	89
REFERENCES	91
APPENDICES	
APPENDIX A. Recruitment email	104
APPENDIX B Spelling inventory and instructions for administration.....	107
APPENDIX C. Interview protocol as approved by IRB.....	109
APPENDIX D. List of misspellings used in error analysis task	112
APPENDIX E. Developmental Spelling Assessment.....	113
APPENDIX F. Sample of a transcribed response to the error analysis task	114
BIOGRAPHY OF THE AUTHOR.....	116

LIST OF TABLES

Table 4.1	Self-Reported Familiarity with Published Spelling Programs	64
Table 4.2.	Percentage of Grade 1 Teachers' Over and Underestimations	66
Table 4.3.	Percentage of Grade 2 Teachers' Over and Underestimations	67
Table 4.4.	Percentage of Students Spelling Each Item Correctly (Item Stage and Feature).....	71

CHAPTER 1

STATEMENT OF THE PROBLEM

Why Accurate Teacher Judgment Matters

A ubiquitous topic of introductory educational psychology courses is Piaget's theory of cognitive development (1964). Students of educational psychology become acquainted, for example, with Piaget's seminal conservation task experiments, in which young children of various ages are asked to consider whether the amount of liquid changes or stays the same when transferred from one container to another. Replications of Piaget's experiments (e.g. Ashton, 1975) have revealed that, as a group, children's ability to reason about this and other problems changes substantially over time in quite predictable ways. Such findings represent foundational information – metaphorical yardsticks – for teachers, for whom being able to judge an individual child's achievement may, in part, be dependent on such pre-established summaries of age-appropriate expectations, whether those expectations be related to general cognitive development, mathematical skill development or the acquisition of literacy skills.

In the field of early literacy, deftness of teacher judgment matters because although giving considerable thought to lesson planning in advance is important, teachers need to be able to make numerous decisions in real time (Artelt & Rausch, 2014; Scales, et al., 2018). For example, they are called upon to make decisions about how to respond to students' misconceptions, and which examples will best illustrate concepts for particular students (Moats, 2009) and also about whether advancement to some next level of challenge might be premature, timely, or overdue. Accurate, reliable teacher judgment of students' abilities also affords more time for teaching because it may reduce the time required for more formal, standardized

assessments that are the source of increasing anxiety for parents, teachers, policy makers, and students alike (e.g., Kohn, 2000).

Accurate indicators of students' literacy skills are important sources of data. For example, they may cue the initiation of a timely intervention or a subsequent referral to special education services for students who struggle despite intervention (e.g. Gersten & Dimino, 2006; Shaywitz, Shaywitz, Fletcher & Escobar, 1990). As Leinhardt pointed out, while it may not be possible to know whether or to what extent an individual teacher applies knowledge about a particular student's abilities, "it is clear that the teacher cannot use it if s/he does not have it" (Leinhardt, 1983, p. 7). Both ostensibly subjective teacher-judgment-based identification procedures and more formal, more objective assessment processes run the risk of misclassifying students, i.e., as in need or not in need of specialized instruction (Begeny et al., 2011). Students who are misidentified as needing special services may unnecessarily strain limited resources, but students whose abilities are unwittingly overestimated raise a more serious ethical dilemma, resulting possibly in a student falling through the proverbial cracks – floundering and in need of services but not identified as such.

A Developmental Lens on Learning

Both Piaget's (1964) and Vygotsky's (1978) views of cognitive development provide a broad conceptual lens through which to view findings on children's spelling development. Although their theories of cognitive development differ in a number of ways, both Piaget and Vygotsky proposed active models of learning, models that underpin the rationale for the current study. Children, they both asserted, are not passive receptors of knowledge; instead, they actively construct knowledge. Both also viewed infants as being born with similar raw materials and physiological capacities for cognitive development.

Piaget's stage theory included four distinct stages. In the sensorimotor stage, which lasts from birth to approximately age two, infants and toddlers construct knowledge based largely on sensory experiences and the manipulation of objects. In the next preoperational stage (ages 2-7), children begin to think symbolically and to learn how to use words and pictures to represent objects; they tend to be egocentric and have difficulty seeing things from the perspective of others. In this stage, they struggle with the concept of constancy. In the next concrete operational stage (ages 7-11) they begin to think logically and become capable of reasoning from specific information to a general principle, but still struggle with abstractions. In the formal operational stage children become more successful with deductive logic, reasoning from a general principle to specific information (Piaget, 1964). A hypothesized relationship between such stages and observed spelling strategies shall be described later. For the moment, it is worth clarifying that Piaget stressed that development must precede learning. In other words, his theory implies the near impossibility of teaching/learning certain concepts until some level of cognitive maturation has occurred within the child.

Although Vygotsky (1978) did not refer to stages in the way that Piaget did, his notion of a *zone of proximal development* did recognize the importance of taking into account a child's current understandings to guide more advanced learning. Vygotsky also stressed the central role of social interaction in the process of learning (1978). According to Vygotsky, there is an important difference between what a learner can do without help and what he or she can do with help. It follows that to move a child toward independence in reading, writing or spelling, a teacher must first have a clear understanding of what the child can do independently and what they might reasonably be expected to be successful in tackling next.

Vygotsky and others (e.g. Wood, Bruner & Ross, 1976) also emphasized the role of more knowledgeable tutors in facilitating discussion and collaborative work with learners – interchanges that positively influence their further learning. As children actively attempt to understand instructions provided by teachers, they internalize such information to the extent that they are capable to guide current and future performance during challenging tasks.

Jeanne Chall's influential work has provided a literacy specific context for interpreting the gradual development of both reading and writing skills (1983). Chall described six levels of development, three of which are most pertinent to this study. At stage 0, a pre-reading stage, it is primarily children's oral language proficiency that is developing and which will provide the foundation for later literacy acquisition. In this stage, as young children are read to by parents and caregivers, they may begin to acquire concepts of print specific to their language. They may observe, for example, that in English, print is read from left to right and from top to bottom. In this stage, too, children may learn to sing the ABC song, become able to recognize a few letters, and perhaps learn to write their own names. In stage 1, children next learn the alphabetic principle and master the correspondences between letters and the sounds they represent. She described readers at this stage as "glued to print" because they have to devote most of their cognitive resources to decoding. In many cases, such intense focus may impact the ability to simultaneously attend closely to meaning. In Stage 2, children learn to deal with increasingly complex phonic elements and read stories composed of increasingly complex words. Through practice, they become much more fluent. Together, Stages 1 and 2, wrote Chall, represent a "learning-to-read stage" (1983, pp. 10-24). Related research by Ehri (e.g.,1987) and by Gentry (2000) is summarized in chapter two and includes discussion of parallel stages of spelling development.

Although Piaget's and Vygotsky's views on development differ, the current study is based on what I regard as commonalities between them and also on Chall's domain specific view of literacy development, namely the understanding of learning as progressing gradually in predictable ways and the notion that instructional scaffolding can positively affect such development. This study is also based on the assumption that understanding typical and atypical development is important for teachers in that it may positively impact adaptive word-level literacy instruction for students of all achievement levels.

Teacher Judgment Studies

Defining Teacher Judgment

In lay terms, the term *judgment* may connote opinions of a person's moral character. In other arenas, such as competitive individual sports like ice skating or diving, the term also implies subjective opinions about, for example, creative expression and execution of required moves. In law, as well, judgment also signifies an opinion which, while based on evidence, is still subjective. In other cases, judgment denotes something slightly different – and relates instead to an ability to accurately forecast an outcome which is purportedly uninfluenced by opinion. In the medical field for example, Doust (2012) described clinical acumen as the ability to recognize “among all the women who complain of feeling tired, the one who has life-threatening Addison's disease” or among all the children “practitioners see with diarrhea, the one with Crohn's disease requiring urgent surgery” (para. 2).

In teacher judgment (TJ) studies, the word judgment is closer in meaning to its use in the medical example. It is certainly arguable whether poor instruction belongs in the same category as misdiagnosis of Addison's or Crohn's disease, but it is certainly true that students deserve classroom teachers with keen insight into their unique struggles. In some educational research,

the phrase “diagnostic competence” has also been used (e.g., Hoth, Döhrmann, & Kaiser, 2016; Klug, Bruder, Kalava, Spiel & Schmitz, 2013; Wuttke & Seifried, 2012). In the TJ literature, the term diagnostic competence refers to teachers’ ability to accurately judge what students know and can do. Less frequently, the term has also been used to refer to teachers’ ability to correctly judge task demands (e.g., Artelt & Raush, 2014).

Finally, it is also important to disambiguate the term *teacher judgement* as used in the present study from similar terms, such as *professional judgment*, which have been used in other contexts to describe the responsibility of teachers to make autonomous decisions that may or may not align with expected practice (e.g. Dottin, 2009; Scales, et al., 2018). In this study, teacher judgment is operationalized narrowly, referring to teachers’ ability to accurately estimate student achievement.

Methodological Issues and General Findings

Teachers’ understanding of their students and of what their students understand about the subject matter they teach has been investigated from multiple viewpoints. The body of research on teacher judgment (as defined above) provided the conceptual and methodological framework for the present study. Studies within this domain have sought to determine teachers’ accuracy in appraising their students’ mathematical and/or reading abilities, and have often involved interviews during which teachers shared their expectations about how individual students in their class had done or would do on specific reading or math tasks (e.g., Hoge & Coladarci, 1989; Südkamp, Kaiser, & Möller, 2012). I was unable, however, to locate any research reports that specifically investigated teachers’ judgment accuracy in estimating the spelling skills of their students. This gap is surprising given that children’s later literacy skills, including spelling, have

been found to correlate strongly with their early literacy skills (e.g., Ehri, 2000; Lonigan & Shanahan, 2008; Perfetti, 1985, 1989).

A comprehensive summary of relevant teacher judgment studies is included in chapter 2, but the general conclusions of meta-analyses by Hoge and Coladarci (1989) and by Südkamp, Kaiser, and Möller (2012) have been that overall, teachers' judgments of their students' mathematical and literacy skills are trustworthy. Hoge and Coladarci, when examining sixteen such studies, found a median correlation of .67 between teacher judgment and student performance on standardized tests of reading and math; in the Südkamp et al. review of 75 teacher judgment studies, the overall mean correlation was similar, $r = .63$, with a median correlation of .53.

Investigators have provided different interpretations of the correlation coefficients found in teacher judgment studies. First, as Madelaine and Wheldall (2005) have asserted, if the relationship between teacher judgement of student achievement and students' actual performance on a reading test is thought of as similar to an expected relationship between two similar but different tests of the same reading construct, a validity coefficient of .67 or .63 might be considered relatively low. Second, while studies of teacher judgement have shown that in the aggregate teachers are able to provide moderately accurate appraisals of students' reading-related knowledge, in reality, teachers' accuracy scores differed considerably among the various studies, and individual performance was significantly dependent on the individual students being judged and the teachers' familiarity with the specific task being judged (e.g., Coladarci, 1986; 1992).

In the majority of studies included in the Hoge and Coladarci meta-analysis, correlation coefficients were calculated by comparing a teacher's predictions of a student's relative ranking among peers with the student's actual score or ranking on some assessment. While such values

reflected the relative relationship between student scores and teacher predictions, the researchers pointed out that such figures did not necessarily reveal much about the accuracy of judgment. An alternative to the correlational approach involves calculating what Leinhardt (1983) termed “hit rates” (p. 9). In studies where a hit rate was calculated, researchers have done so by asking teachers to predict student performance on an item-by-item basis and then dividing the number of times a teacher’s predictions were correct by the total number of predictions the teacher made. The hit-rate method, Coladarci (1986) contended, is a more “revealing” (p. 142) indicator; it allows for a more fine-grained analysis of characteristics of teachers and students and features of the judgment task and how they may be related to judgment accuracy.

A Developmental View of Spelling

Research such as that by Bourassa & Treiman (2001) has shown that three meta-linguistic skills are of particular importance to spelling: a) phonemic awareness, b) morphological awareness, and c) orthographic awareness. Phonemic awareness refers to the realization that words are comprised of single sounds which can be isolated and combined; it is part of a more general awareness of sound called phonological awareness which refers to the ability to notice features of sound at broader levels such as the ability to notice that the speech stream can be segmented into separate, discernable words and that words can be divided into noticeable parts (syllables). Phonological skills develop in a fairly predictable progression (Adams, 1990). Importantly, neither phonemic awareness nor phonological awareness requires any knowledge of written letters. Morphological awareness refers to the understanding that words are also parsable by meaning. The word *clapped*¹, for example, though having only one

¹ For clarity and consistency, examples of printed words and/or names of letters are italicized; sounds are indicated by enclosure between two slashes; examples of nonstandard spellings are indicated by quotation marks.

syllable, can be analyzed as having one part that means striking hands together and another part that signifies time – in this case referring to the past.

Orthographic awareness refers to information stored in long term memory about how spoken words can be represented with written letters/symbols (Apel, 2011). It includes understanding of the alphabetic principle, the idea that in English, for example, *Bb* generally stands for the sound at the beginning of the words *big* and *bark*. The term is also used to describe implicit knowledge of rules that literate persons use but may not notice or be able to describe. For example, although double consonants often appear at the ends of words like *kiss*, *egg* or *fall*, virtually no English words begin with double consonants (Treiman, 1993).

A parallel body of research has explored the extent to which teachers themselves are knowledgeable in these areas (e.g., Binks, 2008; Gibson, 2010; Moats, 1994, 2009; Phelps & Schilling, 2004; Spencer, Schuele, Guillot, & Lee, 2008; Walsh, Glaser, & Wilcox, 2006; Washburn, Joshi & Binks-Cantrell, 2011). The designs of many such studies have depended on measures adopted or adapted from a teacher knowledge survey originally developed by Moats (1994) to assess the knowledge base of experienced language arts teachers in these areas. Test takers were challenged, for example, to count the phonemes in the word *box* (there are 4); they were also asked to do things like count the morphemes in the word *supervisor* (there are 3). Such studies have contributed important descriptive data about teachers' content knowledge for early literacy.

Purpose of the Study

A child's understandings about spelling can provide an important window into his/her spelling and related literacy skills, but according to Invernizzi and Hayes (2004), "the fundamental idea that invented spellings provide a diagnostic cue to a student's current

understanding of how written words work, and that that instruction can be timed and targeted to this understanding, is still, for the most part, overlooked” (p. 217).

As Treiman, Metsala, Ehri (1998) and others have pointed out, the benefits of strong spelling skills are practical, motivational and cognitive (e.g. Moats, 2005; Joshi, Treiman, Carreker & Moats, 2008). In practical terms, Moats (2005) has described how those who spell well may take for granted the role spelling and writing play both in daily life and in high-stakes life events, with poor spelling and writing, for example, likely “dooming” (p. 14) an application for employment.

Researchers have also pointed out that spelling difficulties may affect students’ motivation to write (e.g. Berninger, 1999; Graham, 1999; Graham & Santangelo, 2014). For primary-grade children, the processing demands of encoding speech may cause poor spellers to have trouble retaining in short term memory the ideas he or she began with (Graham, 1999), and they may even influence the words writers choose causing them to avoid words they cannot spell (e.g., Graham & Santangelo, 2014). Beginning writers who have difficulty with physical or mental aspects of transcription processes such as spelling and handwriting may well balk at writing (e.g., Berninger, 1999).

Research has also shown that learning to spell has cognitive effects and is positively correlated with the development of reading (Ehri & Wilce, 1985; Moats, 2005; Paige, et al., 2018; Treiman, 1993). Effective spelling instruction boosts growth in early literacy by enhancing development of children’s phonemic awareness; buttressing understanding of the alphabetic principle, and making it easier for young students to learn sight words (e.g., Ehri & Wilce, 1985; Ehri, 2014). Snow, Griffin & Burns (2005) described the reciprocal role of spelling and reading in similar terms. Spelling and reading are built on the same mental foundation. Knowing the

spelling of a word, they said, makes the representation of it “sturdy and accessible for fluent reading” (p. 86).

Though there is consensus on the reciprocal role of reading and writing skills (see e.g. Ehri, 2000; Moats, 2009; Snow, Griffin, & Burns, 2007; Stotsky, 2006; Treiman, Metsala & Ehri, 1998), I was unable to locate any studies that investigated teachers’ judgement accuracy in the domain of spelling. The primary purpose of the present study was to examine the accuracy of first and second grade teachers in judging the spelling achievement of their students on a developmental spelling task.

A secondary and more exploratory purpose of the present study was to examine a previously unexamined moderator variable that may influence teacher judgment accuracy. Specifically, I explored interactions between accuracy of teacher judgment and teacher knowledge of developmental spelling constructs as manifested by responses during a spelling error analysis task. The error analysis task was designed for this study and was based on the assumption that teachers familiar with research on developmental trajectories in spelling would be more likely to consider such information when analyzing errors.

Although research has often investigated relationships between teacher knowledge and student achievement in reading (e.g., Carlisle, Correnti, Phelps, Zeng, 2009; McCutchen, et al., 2002; Piasta, Connor, Fishman, Morrison, 2009), I was unable to locate any studies with explicit focus on teachers’ knowledge of developmental spelling research. Accordingly, there is a paucity of research on the relationship between teacher knowledge and student achievement in spelling. Treiman (e.g., 1985, 1993, 2017), a researcher who has published extensively in the area of spelling, has described how reading and spelling are often taught as separate subjects with different materials. Since spelling and reading support each other, Treiman asserted that

spelling ought to be taught in a manner sensitive to that developmental course. Without a developmental perspective on spelling, she contended, teachers may be ill-equipped to work with students for whom the acquisition of reading, writing and/or spelling is difficult.

Identification and intervention for students who struggle with reading, writing and/or spelling demands especially accurate teacher judgment. In order to provide accurate feedback and to tailor appropriate instructional designs accordingly, Artelt & Rausch (2014) have argued, teachers need to be able to adequately measure student performance and to be able to judge the difficulty level and demands of class material. Other researchers have used other words to make similar points. Corno & Snow (1986) captured the urgency others have expressed perhaps most emphatically. “If differential treatment of learners is based on invalid assumptions, the teaching system may become maladaptive, even discriminatory in the extreme” (p. 615). Research on the assumptions teachers have about their students’ spelling ability is scarce. The purpose of this study was to address that gap in the literature by exploring the following two research questions. First, how accurately do teachers judge students’ concurrent performance on individual items of a developmental spelling test? Second, what is the relationship between teachers’ accuracy in judging student performance on spelling and their developmental spelling knowledge?

CHAPTER 2

A REVIEW OF THE LITERATURE

This review comprises two parts. I begin with the research on teacher judgment (TJ), reporting on the rationale, methods, types of assessments and key findings of the existing teacher judgment literature related to literacy instruction. I also summarize findings concerning the role of student, teacher, and task variables in influencing TJ, highlighting empirical gaps that provide the basis for my research questions. Second, I discuss key studies on developmental aspects of spelling and also the findings of investigations regarding teachers' understandings of related reading and spelling concepts.

Research on Teacher Judgment of Students' Literacy Skills

In both mainstream media and in scholarly research, there has long been concern and debate about the extent to which teachers are prepared to guide the learning of typical students and to ameliorate the struggles of atypical learners (e.g., Moats, 1994 & 2009; Snow, Burns & Griffin, 1998; Spear-Swerling & Brucker, 2004; Stotsky, 2009). The requisite bodies of declarative and procedural knowledge for teachers have therefore long been the subjects of scholarly inquiry in multiple disciplines.

Shulman (1986) has offered one widely cited conceptual framework for understanding the competencies of effective teachers in various disciplines. Grounded in his earlier experience with the training and assessment of the clinical skills of medical students, Shulman proposed a model of teacher competency that is multi-faceted. Shulman theorized that teachers' competence includes, but is not limited to, such broad categories as domain specific content knowledge, pedagogical knowledge, pedagogical content knowledge, curriculum knowledge, knowledge of

educational contexts, as well as knowledge of learners and their characteristics. Implicit in the last category is a construct similar to accurate teacher judgment.

Similarly, Artelt & Rausch have described teacher judgment (TJ) as an essential capability in teaching and a necessity for adaptive teaching. Accurate diagnostic skills are enormously important for teachers because such information regularly informs the moment-by-moment instructional decisions that teachers make. If a teacher is correct only in sensing that a learner is struggling mightily, but is uninformed regarding the particular source of a student's difficulty, a casually chosen example may only extend confusion (Artelt & Rausch, 2014).

Teacher Judgment Research Design Considerations

In their 1989 meta-analysis of the teacher judgment (TJ) literature, Hoge and Coladarci (1989) examined the results of sixteen studies and reported on the match between teacher estimation of student ability (primarily in reading and math) and student performance on another objective criterion-referenced measure of the same constructs, measured concurrently with another measure given on or about the same date. Their operational definition for accuracy was the correspondence between two sets of values: the teachers' judgments of their students in some area and their students' actual performance on related standardized test. They reported correlations ranging between .28 and .92 and a median correlation of .66.

Artelt and Rausch (2014) emphasized that when looking at individual differences in teacher judgment accuracy, it is important to take into account features of the judgment task. Hoge and Coladarci (1989) distinguished between cases in which teachers were asked to make direct versus indirect judgments. In studies where teachers made direct judgments, they were asked to predict scores on a particular test or even performance on particular items of a test. Direct judgments generated a median correlation coefficient of .69 (ranging from .48 to .92). In

contrast, in studies that involved indirect judgment tasks, teachers were asked, for example, to rank order students overall and then those ratings or rankings were correlated with students' performance on a test with which the teacher may or may not have been familiar. In those studies, the median correlation was .62 and TJ accuracy scores ranged from .28 to .86. Although such a correlation suggests a moderate to strong correspondence between teacher judgment and student achievement, Hoge and Coladarci (1989) were careful to point out that accuracy actually varied widely depending on which teacher judged which students on which tasks. In other words, summary measures such as omnibus correlations, often fail to communicate important details of the data they represent. Hoge and Coladarci explained, for example, that twelve of the sixteen studies used pooled data, lumping all teachers and/or all students together to see if overall, teachers were accurate judges of students' ability / achievement. According to the authors, inherent in this method is the risk of over- or under-estimating what teachers actually know about what their students know.

Other types of statistical procedures, Hoge and Coladarci (1989) suggested, provide more specific and potentially useful information about TJ. One such alternative approach involves the calculation of hit rates. In studies using this approach, researchers have investigated how good teachers predicted individual children's performance on an item-by-item basis. To calculate a hit rate, the total correct judgments made at the item level is divided by the total judgments made. In a recent teacher judgement study, Doore (2010) asked preschool teachers to judge whether their students would respond correctly to individual items on an alphabet subtest of the Test of Early Reading Abilities - Third Edition (Reid, Hresko, & Hammill, 2001). Doore compared teachers' concurrent judgments with students' actual performance on individual items and

calculated a hit rate of 70.1%, meaning that teachers were accurate on approximately 7 out of every 10 judgments.

In a more recent meta-analysis of over seventy teacher judgment studies, Südkamp, Kaiser and Möller (2012) calculated a mean correlation of .63 between TJ and student performance on external measures. Similar to Hoge and Coladarci (1989), Südkamp et al. also categorized teacher judgement studies by design characteristics. For example, like Hoge and Coladarci, they distinguished between designs requiring direct versus indirect judgments but they used slightly different terms than Hoge and Coladarci used in subtly different ways. They considered tasks to be *informed* in studies where teachers judged performance on measures with which they had direct knowledge or experience. In contrast, *uninformed* judgments were more similar in nature to judgment tasks called *indirect* by Hoge and Coladarci (1989). In studies where TJ was uninformed, the mean correlation was .61 but for informed tasks, it was considerably higher ($r = .76$).

As one example of differences between TJ accuracy on informed versus uninformed tasks, Feinberg and Shapiro (2003) asked teachers to make two estimations of students' reading ability. In the first condition, considered more informed, teachers were asked to predict student reading fluency by predicting the number of words one of their students could read correctly in one minute. Importantly, teachers had received a written description of the testing context including audiotaped samples of anonymous students in the test situation. They had also received passages identical to the ones read by their own students. The 150 word reading passages used for this task were selected and analyzed for readability according to Fry's (1969) readability formula. Passages considered not on appropriate grade level were excluded. Without access to the audio recording of their own student collected by the research assistant, teachers

estimated how many words their own student read correctly in one minute. For this TJ task, they reported a substantial correlation between students' reading fluency as measured by the test and teacher judgment of their fluency ($r = .70, p < .001$).

On a second task, considered less informed, teachers were shown several items from a standardized test of reading achievement but rather than asking how students would do on the items, teachers were asked to estimate how well their student would perform on a 1 to 5 scale (1 = far below grade level expectations; 5 = far above grade level expectations). Their estimate was then compared, by the researchers, with students' actual scores. On this task, the researchers reported a statistically significant, positive correlation, which was, however, somewhat smaller than the correlation for the informed task ($r = .62, p < .001$).

Other teacher judgment research (e.g. Bates & Nettelbeck, 2001; Bennett, Gottesman, Rock & Cerullo, 1993) has focused on the role of moderator variables such as student characteristics, teacher characteristics and/or task characteristics and their relationship to judgment accuracy. The mixed findings of such research are reported below.

Student Characteristics as Moderator Variables

In the introduction to their meta-analysis cited above, Südkamp et al. (2012) made the important point that “on the one hand, the combined results of the last 30 years of TJ research “may be interpreted as indicating that teachers' judgments are quite accurate; on the other hand, their judgments are evidently far from perfect, and more than two thirds of the variance in teachers' judgments cannot be explained by student performance” (p. 744). Efforts to understand such variance have been the explicit focus of many TJ studies. The designs of such studies have involved estimates based on correlation but have also included statistical techniques

aimed at addressing what Cunningham, Stanovich, and Maul (2011) called the “third-variable problem” (p. 51).

When a correlation between two variables varies widely (e.g. TJ accuracy and student achievement), a natural response among those attempting to explain such variance is to look for third variables. Cunningham et al. (2011) explained how regression techniques can help researchers compute correlations between variables after other variables are “factored” or “partialled” out (p. 53). Third variables hypothesized as possibly influencing teacher judgment accuracy related to literacy have included students’ behavior, gender, and general abilities relative to their peers. Bennett et al., (1993), for example, reported that teachers predicted lower academic performance for students they considered to have poor behavior relative to their peers, actual academic skills notwithstanding. This finding was especially true, they reported, for boys. Shaywitz, Shaywitz, Fletcher & Escobar (1990) noted a similar gender bias in referrals for reading disabilities. Bates and Nettelbeck (2001), on the other hand, investigated whether behavioral problems of students affected their teachers' ability to accurately judge their reading abilities and found no statistically significant evidence of such bias.

Findings from investigations of a relationship between student gender and TJ accuracy have been mixed. According to Hoge and Coladarci’s meta-analysis (1989), student gender was not significantly related to judgment accuracy. Hinnant, O’Brien and Ghazarian (2009), on the other hand, found a “marginally significant interaction” ($p = .08$) between first grade teacher expectations and third grade reading achievement and students’ gender. In their study, teachers tended to underestimate the later reading achievement of minority males more than girls. In an earlier study by Beswick, Willms and Sloat (2005), student gender was also found to be significantly related to teacher judgment accuracy. In that study, teachers underestimated the

performance of boys, despite there being no actual difference between the performance of the boys or girls in the study.

Begeny, Krouse, Brown, and Mann (2011) also examined teacher judgment across student ability levels. In their study, 27 first- through fifth-grade teachers made indirect, concurrent estimates of the abilities of eight of their students on a broad range of literacy skills by responding to items on a 5-point scale ranging from consistently poor to consistently successful. Sample items included “Please rate the student’s level of fluency during oral reading.” The researchers found that teachers in their study “judged low- and average-performing readers less accurately than high-performing readers” (p. 34). This finding supported that of Hoge and Coladarci (1989) who also observed that students’ general academic ability might influence the accuracy with which teachers judge student achievement. In a study focusing specifically on teacher judgment of reading, Coladarci (1992) found that teachers were markedly less accurate in judging the performance of lower ability students, defined as students reading one year below grade level, than they were in estimating the performance of higher ability students, defined as students reading one year above grade level (62% accurate for low ability students versus 85% accurate for higher ability students).

Conflicting results, though, were reported in a dissertation by Martin (2005). In her study investigating the influence of student ability level on TJ, she found that teachers actually made more accurate judgments for lower achieving students. Martin conjectured that passage of the No Child Left Behind (NCLB) law in 2002 had perhaps encouraged early elementary teachers to be more sensitive to the needs of low achieving readers.

In 2003, Hamilton and Shinn investigated what Nathan and Stanovich (1991) termed the “red herring” (p. 177) of reading research literature – the existence of word callers – students

who purportedly read accurately but are weak in reading comprehension. Studies of word callers have made a notable contribution to the TJ literature by focusing on the distinction between word-level reading accuracy versus satisfactory reading comprehension and the extent to which conflation of the two constructs has muddied the interpretation of results of TJ studies. Hamilton and Shinn (2003) asked teachers to make concurrent direct judgments about student reading by first nominating word callers and similarly fluent peers for inclusion in their study. They found that although teachers perceived all the students they selected for the study as having similar oral reading fluency, in actuality, teacher-identified word callers “read significantly less fluently than students that teachers judged to have equal oral reading skills” (p. 238). Such findings suggest that some teachers may have difficulty accurately diagnosing and therefore treating the actual problems of some of their struggling readers. This particular finding is worrisome given the consensus among researchers that for children who experience literacy difficulties in the early grades (K-3), considerable problems exist at the word level (e.g., Juel, 1988; Washburn, Joshi, Binks-Cantrell, 2011).

Teacher Variables Mediating TJ Accuracy

Variability in judgment acumen has also led researchers to explore whether particular teacher characteristics are related to judgment accuracy. Educational level and years of experience are among the differences that have been explored most frequently by a handful of researchers. Bates and Nettelbeck (2001), for example, investigated the possible mediating role of years of experience but concluded that experience did not seem to make teachers better or poorer judges of students' reading accuracy or comprehension. Similarly, Begeny, Krouse, Brown, and Mann (2011) used chi-square tests to explore the relationship between judgment accuracy and years of experience, grade taught, and level of education (M.A. or not) and found

no significant differences across any of those teacher variables. Valdez (2013) also examined teacher experience as a possible moderator of the concurrent relationship between teachers' judgment of reading skill and students' performance on standardized tests of reading performance. Valdez found that neither years of experience nor educational attainment were a significant moderator variable.

One explanation for such findings may relate to the distal nature of such variables. In other words, neither education level nor years of experience reveal what may be important differences in knowledge that different teachers have been exposed to. In their text, *Knowledge to Support the Teaching of Reading*, Snow, Griffin and Burns (2005) emphasized that "the quantity and complexity of the declarative and practical knowledge teachers need ... is so great that it simply cannot be mastered adequately in the brief time available during a pre-service program" and as a result, novice teachers can at best be expected to "do no harm," (p. 9). As for more experienced teachers, Snow has also asserted that while there is recognition that ongoing learning is as important for teachers as it is for medical doctors or car mechanics, the reality is that "nobody worries about what courses [teachers are] taking. . . the fact of the matter is, you just have to show that you've taken a course" (C. Snow, personal communication, January 2, 2016). In other words, although it is now common practice to require teachers to earn a master's degree or take post-graduate credits, it may well be that even more experienced and more educated teachers have not been exposed to the particulars of a knowledge base that might inform and possibly improve their diagnostic perspicacity.

Snow, Griffin and Burns (2005) were careful to emphasize the importance of what they termed "usable knowledge" (p. 11) and to distinguish it from more esoteric concepts from various linguistic subdomains like sociolinguistics, psycholinguistics, neurolinguistics,

computational linguistics, and applied linguistics. Their conception of usable knowledge bears striking resemblance to a component within Shulman's model of teacher competency (1986), pedagogical content knowledge (PCK).

Shulman (1987) discussed what he saw as an important distinction between subject knowledge of a declarative, factual type – which he termed content knowledge – and pedagogical content knowledge that includes “ways of representing and formulating subject matter that make it comprehensible to others,” (p. 9). Such an ability necessarily presumes mastery of a great deal of relevant content knowledge (CK), but as Shulman put it, pedagogical content knowledge includes understanding what a pupil understands and requires “a deep grasp of both the material to be taught and the processes of learning” (1987, p. 19).

Although intuitively appealing, Park and Oliver (2008) have described how the “amorphic nature” of PCK (p. 262) has created some difficulty in operationalizing it as a tool for research, but such difficulty has not inhibited study of how particular content and pedagogical content knowledge might be related to student achievement outcomes (e.g., Piasta et al., 2009).

Within the field of literacy, there is consensus that such usable teacher knowledge includes – but is not limited to – understanding of a documented continuum of phonological development and the orthography of English and the ways it reflects and encodes speech (e.g. Moats, 2009; Snow, Griffin, & Burns, 2007; Stotsky, 2006). Knowledge in these areas is considered important to effective general early literacy instruction and necessary for teachers in diagnosing and intervening effectively with reading, writing, and spelling difficulties. To clarify, although previous research has attempted to explain student achievement outcomes by assessing teachers' content knowledge and/or pedagogical content knowledge, the only studies that investigated a link between such knowledge and judgment accuracy have used proxy measures

(such as highest degree, or years of experience) rather than more direct probes to explore a relationship between teacher knowledge and teacher judgment accuracy.

Item-Related Variables That May Moderate Judgment Accuracy

Doore (2010) explored whether sensitivity to task differences such as item difficulty might be related to teacher judgment accuracy and found that the pre-school teachers in his study were less accurate in judging student performance on more difficult items on the Test of Early Reading Abilities - Third Edition Alphabet subtest (TERA-3, Reid, Hresko & Hammill, 2001) as compared with judging performance on items that were easier. As an example, he described how, in the aggregate, teachers judged that more than 70% of their pre-school students would be successful in locating the lower-case version of a three-word title presented in all capital letters. In reality, Doore reported that only 9% of students could answer that item successfully. Doore calculated an effect size of nearly +1.00 SD when judging easier vs. more difficult items on the TERA-3 as determined by frequency of correct answers by students.

Empirical Gaps and Questions for Further Research

In summary, although accurate teacher judgment is an important component of teacher competency, research has raised questions about whether teachers' estimation of students is always accurate enough to inform effective instruction. In drawing conclusions from their analyses, Madelaine and Wheldall (2009) reminded the reader to consider the purpose of teacher judgment. It is important, they stressed, to think about what it is we want teachers to judge and why; understanding what students understand is essential to planning effective literacy instruction for all and for planning targeted interventions for those who struggle.

Südkamp et al. (2012) conjectured that teacher judgment accuracy may be associated with what they termed teachers' expert knowledge. Expert knowledge has been described as

including the ability to perceive meaningful patterns of information that are not noticed by others. In the case of spelling, researchers have offered a powerful lens through which to view student spelling errors. The proposed study will examine such a relationship in the context of spelling error analysis. Knowledge of developmental spelling theory allows teachers to discern important patterns of information which may drive more effective instruction and, in turn, may increase student achievement. I was unable to locate any studies that explored teachers' understanding of the particular complexities of spelling or the possible direct or indirect impact on judgment accuracy or on student achievement in spelling.

Developmental Aspects of Spelling

What constitutes even a basic knowledge base for the teaching of early literacy skills is impossible to summarize concisely. Forerunners in early literacy research have filled volumes with their opinions on pedagogical aspects of reading and their findings on physiological and psychological aspects of reading. Discussion of the mechanisms, meaning, and purpose of reading and writing have been presented at conferences and published in journals and books authored by such pioneers as Huey (1908), Chall (1967; 1983), Gough and Hillinger (1984), Gough and Tunmer (1986), and Adams (1990) – to name but a few.

Current understandings of a developmental course in spelling skills are informed by what has been called the Rosetta Stone of spelling research (Templeton, 1992) – the seminal work by Charles Read (1971). Over time, Read's findings have been replicated, refined, and embellished (e.g., Beers & Henderson, 1977; Ganske, 1999; Gentry & Henderson, 1978; Read, 1971 & 1975; Treiman, 1993). Read observed that although individual children take somewhat different routes to becoming literate adults, they nevertheless pass predictable and recognizable landmarks along their journeys, sometimes called phases or stages. There is now consensus that American

English speakers learning to spell, both children and adults, including those identified as learning disabled, pass through predictable phases/stages in quite similar sequences as they gradually move toward mastery of conventional spelling.

In the earliest stages, distinctions between children's drawing and writing may be difficult for adult observers to discern. Henderson and Templeton (1986) termed the earliest stage *preliterate*; in Ehri's research (e.g. 2014), this phase is termed *prealphabetic*; Ganske (2013) has called it *emergent spelling*. Regardless of the label, what characterizes this earliest phase is that prior to using any true understanding of letter sound correspondence, young children can and do frequently refer to different parts of their own writing as letters or words rather than pictures or drawings. Treiman (2017) pointed out that even young children usually have had many chances to look at examples of writing and to observe its unique characteristics. Text features that children refer to as writing are different than those they refer to as drawings. Writing, as Treiman noted is more likely to be "small and dark;" whereas drawings are "large and more colorful" (p. 265). In this early stage, children may use a combination of scribbles, mock letters and even numbers to produce writing that is interpretable only by its author.

Later, children begin to use their knowledge of letters and letter sounds to write more conventional text. In this stage, beginning and ending consonant sounds are far more likely to be captured (rightly or wrongly) than interior sounds. Their attempts to map speech to print reveal their nascent understanding of the alphabetic principle. They often spell using letters whose names include the sound they are trying to represent (Henderson and Templeton, 1986). Henderson and Templeton termed this the Letter Name (LN) stage. According to Ganske (2013), Letter Name spellers are common in first grade and the strategies used in this stage may

well last for as long two years. Even middle and high school students having trouble with spelling and/or reading often have not mastered all of the skills in this early stage.

In many cases, children's use of phonetic properties of letter names explains what might otherwise perplex fully literate adults. Consider, for example, the misspelling, in the present study, by several students of the word *yes* as "ues." Though children were not available for questioning, research has pointed out the logic of children using the /y/ sound at the beginning of the spoken name of the letter *u* as the logical basis for such reasoning (e.g. Bear, Invernizzi, Templeton & Johnston, 2012; Block & Duke, 2015; Henderson & Templeton, 1986).

Likewise, using *h* to spell the /ch/ sound at the beginning of a word like *chip* (as in "hip") is logical if one's theory of spelling development is grounded heavily in the idea of spelling as memorized visual representations, but makes equal sense from a Letter Name / phonetic spelling strategy. Considering that the sound of the letter *h* actually contains two distinct sounds: long /ā/ followed by a /ch/ sound and considering that the latter is the first sound in the word *chip*, such an error is actually quite well reasoned. Similarly, *white* rendered "yit" also makes sense.

Patricia Kuhl (2010) has described children's linguistic perception as "nothing short of rocket science." Indeed, children's early spelling attempts are often evidence of quite sophisticated analysis and acknowledgment of very real differences that many adult native speakers of English have unwittingly learned to ignore. Research has shown that as young children attend to sound (though not exclusively to sound) in order to spell, they are quite attentive to aspects of articulation such as voicing, relative airflow restriction, and nasalization as well as coarticulation (the effects of adjacent sounds upon each other).

When administered by knowledgeable users, developmental spelling inventories offer a glimpse into children's advanced phonological intelligence even at this relatively rudimentary

level. Children in this early stage, for example, are very sensitive to an organizing feature of speech called voicing. There are a number of sounds in English which are identical in place and manner of articulation with the exception of the absence or presence of vibration of the vocal folds and vocal cords (Fromkin & Rodman, 2018). Such attunement prepares them to subtly alter the pronunciation of the plural morpheme commonly spelled with an *s*. As native speakers do, when speaking, they closely attend to and accurately produce the last sound in *dogs* as /z/; they hear and pronounce the last sound in *cats*, however, as being an /s/ sound (Gleason, 1978). When attempting to spell the two words, young children are apt to use different letters at the end of the two words, demonstrating their primary focus on spelling as an act of 1:1 phonetic transcription. Similarly, children in the early stages of encoding regularly have understandable difficulty spelling words whose oral production includes no discernable vowel sound, such as in the case in the words *girl* and *hurt* (e.g., Beer & Henderson, 1977; Gentry & Henderson, 1978; Ganske, 2013).

Children notice that speech sounds affect and even change each other appreciably when articulated in particular sequences. Unlike the adults who might worry over implied speech deficits suggested by spelling *truck* as “chruk,” children realize, intuitively, that mental and motor anticipation of the /r/ requires that native speakers use subtly different points and manner of articulation (affrication) when pronouncing the first sound in *truck* versus the first sound in *tuck*. They recognize without resistance the pleasing alliteration in the phrase *choo-choo train*. As Treiman (1985) has cautioned, a child who gives the “wrong” answer to a question such as “Does *truck* begin with *t*?” does not necessarily lack metalinguistic skill. On the contrary, as explained earlier, children are often aware of phonetic details that may be inaccessible to adults,

but children have “simply not yet learned which features of sounds are represented in the English spelling system” (p. 200).

A developmental spelling assessment anticipates which aspects of the phonological system the child might be expected to master earlier rather than later and accordingly, asks teachers to do a thorough analysis of types of errors they make. Initial and final consonants are the first aspects with which early spellers gain competences (Ganske, 2013). Not surprisingly they do better at spelling sounds for which the letter-sound correspondence is relatively straight forward. Next children begin to be able to represent initial (but not yet final) consonant blends and digraphs. The consistent representation of interior short vowels sounds tends to show up only after the child has demonstrated the ability to use many consonants. The inclusion of vowels, early on, is characterized by many substitutions. Because vowels are sounds produced with little or no contact between the tongue, teeth, lips or other parts of the mouth, the features which distinguish them are subtle, but for children (and for English language learners), intuition usually results in a substitution of a sound produced in the most similar position possible. The short vowels /i/, /e/, and /a/ for example are considered front vowels based on the relative position of the tongue during articulation compared with the relative back (in the mouth) position of the tongue during the production of /u/ and /o/ (Fromkin & Rodman, 2018). Children, therefore, rarely substitute *o* for *i*, but often substitute *i* for *e* (or vice versa) in words like *sit* and *set*. In some American English dialects, such words are legitimate homonyms produced in exactly the same manner and place (e.g., Ehri, et al., 2001; Read, 1975).

Children’s acute sensitivity to such articulation aspects also explains the significant differences in their ability to spell words of similar look and length such as *cap* versus *can*. During speech, most of the consonant and vowel sounds of English include exhalation of air

either relatively gradually or abruptly through the mouth. In other cases, (i.e. /n/, /m/, and /ng/) air leaves the lungs via the nasal passages, bypassing the oral cavity (Fromkin & Rodman, 2018). Native speaker intuition prepares children and adults to unconsciously anticipate such occasions and perform accordingly. When attempting to spell words whose articulation includes nasal exhalation, children may be more attuned to this feature than many adults. In spelling a word like *can* for example, they notice that the nasalization actually affects the adjacent vowel, significantly altering its prototypical short /a/ quality in a way that often makes it unrecognizable to a child who has not been taught and may never be taught the letter that makes the nasalized /ã/ sound.

Late in the LN stage, children gain competence in spelling words with final consonant clusters and digraphs. Difficulty with particular blends is compounded by the difficulty with nasality described earlier. Children who seem capable of spelling words with final consonant clusters like *last* may have great difficulty spelling final clusters that include nasals such as *land* probably because when you say the latter, the flap to the nasal cavity opens when the vowel sound begins and so the vowel and nasal are articulated simultaneously (Ehri, 1987). Thus, though in the same position, the /n/ in *land* is presumably more difficult to spell than the /s/ in *last*. Ganske (2013) and others (e.g. Schreiber and Read, 1980) have observed that inclusion of the preconsonantal nasal, in fact, often marks the transition between the qualitatively different strategies of phonetic or Letter Name spelling to that of using rules and visual patterns to encode words. In summary, phonetic spelling is, in and of itself, a complex task, in part because children are capable of perceiving a variety of contrasts in the speech stream which, though not necessarily relevant in their native language, are very much real (e.g., Burnham, 2003).

Scharer and Zutell (2003) described students in the LN stage as still moving left-to-right across words, not attending to chunks and not seemingly capable of thinking about the vowel in the context of rules regarding what sort of vowel exists or what follows the vowel. As a practical example, they are, in this stage, oblivious to the notion of a *magic e* spelling rule that when applied can render vowel sounds long. Perhaps coincidentally or perhaps not, according to developmental spelling research, this shift in strategies tends to occur when children are theorized, according to Piaget (1964), to be transitioning from a preoperational stage to a concrete operational stage. The former, Piaget wrote, was characterized as a period (ages 2-7) during which children are capable of symbolic thinking but also egocentric and apt to struggle with the idea of constancy and abstractions. In the concrete operational stage (ages 7-11), children begin to think logically and become capable of reasoning from specific information to a general principle, but still struggle with abstractions.

In the Within Word stage of development, spellers gradually come to understand that some spellings are based on rules or visual patterns that do not correspond directly to phonemic sequences. Such a “new perspective on words requires a degree of cognitive maturity,” explained Henderson & Templeton (1986, p. 309). A child may notice, for example, that long vowel sounds are often encoded with a word final *e* and apply this discovery by overextending the rule. They also begin to be capable of spelling by analogy but with many limitations. Scharer and Zutell (2003) detailed, for example, a study involving first- through fifth-grade children being taught to use analogy as a strategy for spelling. Notably, they emphasized, first graders in the study were largely unable to learn many of the anchor reference words even with instruction. Only among students in second grade was a significant positive effect observed, but what was striking, they emphasized, was not that second graders could be taught to use analogies

to spell but “how much effort the researchers had to make to get them to that point” (p. 15), a point they may well have reached with or without intensive instruction, given more time. Such an observation is compatible with Piaget’s (1964) idea that development precedes learning and calls into question the practice of whole class spelling lists that do not take into account an individual child’s current abilities or capabilities.

It isn’t until learners reach the next Syllable Juncture stage (often in the intermediate grades) that students come to understand how to coordinate multiple aspects of sound, meaning (e.g. tense and plurality) and visual spelling rules to accurately write presumably familiar words. Consider for example, the layers of knowledge required to spell the word *clapped* accurately. In the last stage, which Henderson and Templeton (1986) described as characterized as understanding of derivational relationships, learners come to realize that despite variations in sound across related words like *sign* and *signature*, spelling is visual means of preserving and highlighting relationships among words.

Related Theories and Findings Regarding Spelling Development

Research by Bourassa & Treiman (2001) has identified three meta-linguistic skills that are of particular importance to spelling: a) phonological awareness (including phonemic awareness), b) morphological awareness, and c) orthographic awareness. Phonological awareness refers to the metalinguistic ability to notice and manipulate sound units in *spoken* language – to notice for example that the speech stream can be segmented into separate, discernable words, to notice whether two words rhyme, or that words can be divided into noticeable parts called syllables (e.g., Stahl & Murray, 1994). Phonological skills develop in a fairly predictable progression, beginning with the ability to reflect on larger units such as rhyming words. Gradually, children become more able to analyze smaller units such as single

sounds (Adams, 1990). Phonemic awareness refers to more sophisticated ability to attend to and to manipulate single sound units in words. It includes, for example, the ability to produce, in isolation, the first sound in the word *ship*, or to segment and count the sounds that combine to make the oral word *shape*. Importantly, neither phonological nor phonemic awareness requires any knowledge of written letters. The ability to spell *hors d'oeuvres*, for example is unrelated to the ability to easily analyze it as containing two syllables.

Morphological awareness refers to the understanding that words may also be parsable by meaning. The word *clapped* for example, though having only one syllable, can be analyzed as having one part that means striking hands together and another part that signifies time – in this case referring to the past. This is also sometimes referred to as structural analysis. Orthographic awareness refers to information stored in long term memory about how spoken words can be represented with written letters/symbols (Apel, 2011). It includes understanding of the alphabetic principle, the idea that in English, for example, *Dd* generally stands for the sound at the beginning of the words *dog* and *dark*. The term is also used to describe implicit knowledge of rules that literate persons use but may not notice or be able to describe. For example, although *ck* is one way to spell the /k/ sound, the /k/ sound can never be spelled that way at the beginning of a word.

A separate but complementary four-phase model of reading and spelling has been proposed by Ehri (2005). In what Ehri termed a pre-alphabetic phase, children attend to visual features like font or color to read environmental print like McDonald's or Taco Bell signs. This phase can be likened to what Frith (1986) described as a logographic strategy. The same children when presented with the same words presented out of context would likely find the

words unrecognizable. In this phase, a child may look at a tube of Crest® in the bathroom and conclude that toothpaste is spelled CREST (Tolman, 2010).

Ehri & Wilce (1985) termed the next phase an early alphabetic phase and described it as being characterized by partial phonemic awareness. Word recognition in that stage, she explained, is constrained by the child's inability to successfully segment words into all the phonemes they contain. Instead, the child may rely on first or last letters alone as primary cues. The same limitations, they asserted, affect spelling ability. In the early alphabetic stage, a child may write letters for the dominant sounds but not all sounds in words. Furthermore, letters selected to represent sounds may be tied closely to the letter's name rather than the letter's sound. Children in this phase may use the letter *y* to spell the /w/ sound, for example, because the name of the letter *y* begins with the /w/ sound.

A later, more secure alphabetic phase, Ehri (2005) posited, is characterized by complete phonemic awareness as well as growing understanding of the morphophonemic nature of English orthography. In other words, the understanding that words' spellings contain both pronunciation and grammatical cues. For example, children may notice and begin to consistently spell regular plural and regular past tense with *~s* and *~ed*, respectively, despite variants in pronunciation. The later phase is also characterized by a growing sight-word vocabulary.

In the last, consolidated alphabetic phase, Ehri held, readers become both accurate and fluent. In explicating her choice of the word *phase* to label her model, Ehri explained that the use of the term *stage* may denote a strict view of development in which one type of word reading occurs at each stage, and in which mastery is seen as a prerequisite for movement to the next stage. Neither such stage models nor her own phase theory model, she clarified, are actually so rigid.

Gentry (2000) presented a similar classification system for characterizing the qualitative changes over time in what children apparently know and do not know about spelling. In the pre-communicative stage, spellings appear to be random. In the semi-phonetic stage, spellers provide partial letter-sound mappings. In the phonetic stage, a letter is provided for every sound but not necessarily the letter used in conventional spelling. Transitional spellers adhere more closely to conventional spellings, including vowels in every syllable and possibly including letters they know are part of a conventional spelling but in the wrong order (e.g., writing “huose” for *house*). In the last correct or conventional stage, basic knowledge of English orthography rules is securely in place. In summary, research on the developmental aspects of both reading and spelling is well established. Researchers studying the spellings of young spellers have consistently observed a gradual shift in the knowledge sources called upon by children as they put words on paper.

Henderson and Templeton (1986) described these various sources as three “ordering” (p. 306) principles of English spelling that influence children’s spellings in predictable sequence. Beginning with alphabetic or letter name spelling, children progress to noticing and extending within-word spelling patterns and finally they incorporate meaning as a strategy when spelling. It is important to clarify that although researchers have recorded the age ranges and grade spans at which a majority of children can use different spelling strategies, such data is descriptive only; different children proceed at different rates along the same paths.

Research on Teacher Knowledge

Direct Measures of Teachers’ Knowledge

For more than two decades, researchers have maintained that teachers of early literacy need high levels of understanding of the linguistic foundations of early reading and other literacy

related content knowledge (Moats, 2009). Whether and to what extent such knowledge constitutes simple declarative knowledge or pedagogical content knowledge is a matter of debate; however, given the contribution of the alphabetic principle to successful reading and the links among phonology, orthography, and meaning, particularly in the beginning stages of literacy, teachers' own knowledge of the alphabetic principle and of the regular and irregular mappings between language and print have frequently been the subject of reading research. Such knowledge, Moats emphasized, is not acquired casually and is not a natural consequence of mature reading ability. In the present study, I use the term *teacher knowledge* as an umbrella term to reference what in some studies is called the knowledge base, subject matter knowledge, content knowledge, or pedagogical content knowledge. As Shulman himself asserted in regard to one of his own papers, "while far more can be said regarding categories of a knowledge base for teaching, elucidation of them is not a central purpose of this paper" (1987, p. 8).

Measures of teacher knowledge of language structure have been used in many studies whose aim was to explore a relationship between teacher knowledge and student achievement. Intuitively, it was expected that greater teacher knowledge might result in greater student achievement. Some studies have reported modest positive relationships between teacher knowledge and student achievement (e.g., Bos, Mather, Narr, and Babur, 1999; McCutchen, et al., 2002; and Spear-Swerling and Bruker, 2004). These positive results notwithstanding, other studies have resulted in mixed findings (e.g. Carlisle, Correnti, Phelps, & Zeng, 2009 and Duggar, 2016). Such mixed results should not be construed as implying that teacher knowledge is irrelevant to student achievement; more likely, as Spear-Swerling and Cheesman (2012) asserted, such findings are evidence that student academic performance outcomes are the result of various influences interacting in complex ways.

Piasta, Connor, Fishman, and Morrison (2009) were among the first to hypothesize that mixed findings on the relationship between teacher knowledge and achievement might be the result of over-simplistic conceptual models of the relationship between teacher knowledge and student outcomes. Piasta et al. hypothesized that “students’ literacy skill gains would not be predicted by teacher knowledge alone but by teacher knowledge as it informed classroom practices” (p. 228). A strength of the Piasta et al. study was that the research design included measures of how knowledge was enacted in the classroom. The results of their study confirmed that indeed, there were important interactions among the variables in the study. Specifically, the students who made gains in word reading skills were the students of teachers who had both higher knowledge scores and also spent more time specifically focused on teaching decoding skills. For students of teachers with low knowledge scores, the more time the teacher spent on decoding instruction, the more poorly the students scored on word reading tests in the spring. In their conclusion, the authors call for further research to investigate the tangled nature of relationships among these variables.

Indirect Assessments of Teachers’ Knowledge

Other studies have explored teacher knowledge more indirectly, relying, for example, on document analysis (e.g., syllabi from university courses for pre-service teachers; required textbooks for the same courses) to make inferences about the extent to which pre-service teachers had been exposed to topics like phonics and phonological awareness (Hess, Rotherham & Walsh, 2004; Walsh et al., 2006). After comparing the contents of hundreds of syllabi and course texts with recommendations of the National Reading Panel's report (2000), Walsh et al. (2006) concluded that pre-service teachers did not appear to have received preparation in key areas, and in many cases, were taught philosophies of teaching reading which stood in direct

contrast to the results of scientifically based reading research. The conclusions reached by Walsh et al. have been controversial due, in part, to the data collection process that invited cooperation from teacher preparation programs, but then, in cases where the invitation was declined, the researchers went to court for the right to access course syllabi that some had declined to furnish. Other investigators have reached similar conclusions using other methods (e.g., Rigden, 2006). Rigden was commissioned by the National Council for Accreditation of Teacher Education (NCATE) to investigate how well its expectations for teacher knowledge and skills were aligned with the research base. In her study, Rigden examined state licensure tests and their coverage of the most important insights gleaned from two decades of reading research and concluded that “it is quite possible – maybe even probable – that candidates can be licensed to teach elementary students without demonstrating their knowledge of essential components of effective reading instruction derived from research” (p. 6).

Assessment of Teachers’ Knowledge of Spelling

Few studies have targeted (directly or indirectly) teachers’ knowledge base as it specifically relates to spelling. One of the earliest tests was carried out by Moats (1994). Moats developed a knowledge survey designed to tap spelling- and reading-related knowledge of more than fifty teachers enrolled in a course she was teaching. Though Moats referred to such a knowledge base as “content knowledge,” Shulman’s conceptualization of PCK resonates in her detailed description of the “advantages of a good knowledge base” (p. 95). Those include, Moats wrote “being able to pick the best examples for teaching decoding and spelling,” and “being able to use knowledge of morphology to explain spelling” (p. 96). She discovered what she described as “insufficiently developed concepts about language and pervasive conceptual weaknesses in the very skills that are needed for direct, language-focused reading instruction” (p. 91). As

examples, she cited that just 30% could explain the *y* to *i* spelling rule or when to use *ck*; only 20% could explain the rule for doubling *m*, and only 10% could correctly identify the third speech sound in *thank*.

In a more recent study by Moats (2009), she reported on the results of an administration of a newly designed knowledge survey administered to more than 100 primary level teachers in Utah and Florida. An appendix to that study provided multiple examples to support Moat's conclusion that there were widespread and surprising gaps in the groups' overall understanding of oral and written language concepts. For example, only 52% of test takers could select the correct answer (indicated by italics) to the following question:

The /k/ sounds in *lake* and *lack* are spelled differently. Why is *lack* spelled with *ck*?

- a. The /k/ sound ends the word.
- b. The word is a verb.
- c. *ck is used immediately after a short vowel.*
- d. c and k produce the same sound.
- e. There is no principle or rule to explain this.

In one spelling-focused study, Carreker, Joshi, and Boulware-Gooden (2010) considered potential relationships between what they termed "literacy-related content knowledge" and ability to analyze students' spelling errors and select appropriate instructional activities based on such analysis. Content knowledge was measured by a 30-item test that asked participants to count syllables, phonemes, and morphemes. The spelling instruction assessment (SIA) was developed by Carreker to assess participants' ability to use spelling errors to identify a student's underlying difficulty and plan appropriate instruction. The measure consisted of 12 items which assessed whether participants could choose from among alternatives the one instructional activity

that directly targeted a demonstrated spelling problem. Teachers' performance on measures of phoneme, syllable, and morpheme counting were used as independent variables to predict the outcomes on the SIA. They found that participants who demonstrated the greatest knowledge of phonemes, syllables, and morphemes were better able to identify the most appropriate activities. An acknowledged limitation of the study, however, was that reliability for the SIA measure was only moderate (Cronbach's $\alpha = .64$).

McNeil and Kirk (2014) asked teachers in New Zealand to evaluate their preparedness to teach spelling effectively and reported that the majority of teachers who participated in the survey (69%) felt that they had not received adequate preparation to teach spelling as part of their teacher training program. The authors suggested that such results may imply a lack of focus on teaching skills that underlie spelling success in initial teacher education. Teacher preparation programs, they implied, do not put emphasis on teacher candidates' knowledge of the structure of language.

Spear-Swerling and Cheesman (2012) investigated what they termed "teachers' knowledge base about English word structure" (p. 1692) using a multiple-choice knowledge survey patterned after a Massachusetts' teacher licensure exam. In the introduction to their study, they explained how it built upon previous studies of teachers' pedagogical content knowledge in reading, including Moats' survey (1994). Unlike the many iterations of the Moats' survey, they noted that their assessment tool included a majority of items (67%) situated in classroom contexts to assess both knowledge and application of current theory and practice. Overall, they reported, there was wide variability in participants' performance; mean accuracy ranged from about 58%–65% correct. Participants obtained the highest scores on the subscale involving fluency, vocabulary, and comprehension items ($M = 65\%$ correct). Performance on items related

to phonemic awareness and phonics was slightly lower with participants answering 61% of the items correctly. The authors noted that participants had considerable difficulty with items that required them to select the best examples of words to use to introduce phoneme blending; for example, only 26.6% could effectively consider distinctions involving number of phonemes and physical manner of articulation to select the right answer (indicated by italics) to the following question:

A second-grade teacher is working with a student who knows sounds for all single consonants and short vowels, as well as for the letter patterns *sh* and *ch*. However, the student has great difficulty blending sounds. Which of the following words would likely be easiest for the student to blend?

- a. *mash*
- b. tug
- c. slip
- d. chin
- e. I don't know

Synthesizing related research, Washburn, Joshi, & Binks-Cantrell (2011) reported that two-thirds of pre-service and in-service early elementary teachers scored below 60% correct on a comparable teacher knowledge measure by Bos et al. (2001). Similarly, Cunningham et al. (2004) reported that 20% of K-3 teachers they survey were unable to correctly identify the number of phonemes in any of the 11 words on their phonological awareness task. In Washburn, et al.'s own study (2011), the mean percent correct for all alphabetic principle and phonics knowledge and skill items on their measure was just 52%, and the mean percent correct for morpheme identification was approximately 54%.

Alatalo (2015) examined knowledge for teaching spelling of over 250 literacy teachers in grades 1-3 in Sweden. The author described such knowledge in terms of knowledge of code concepts and knowledge of language structures. The teacher knowledge survey (TKS) used was based on Moats' 1994 survey but adapted to take into account differences in Swedish language structure and orthography. The survey included 43 total items with 11 related to spelling rules and conventions. Similar to the Moats' study, participants were asked, for example, to explain the rule (in Swedish) for doubling consonants. Alatalo described the aggregate results as generally low; half the teachers received just partial credit and another 18% either responded incorrectly or wrote "I do not know (p. 11)." Just 32% received full points.

In summary, researchers have found wide variation among teachers in foundational knowledge deemed essential for early literacy instruction by many in the field. Moats (1994) suggested that teachers who are well-versed in foundations are more apt to understand why students make the errors that they make. Such knowledge may enable teachers to judge not only what a particular student knows but also what he or she needs to know next about the relationship between speech and the printed word. As Carreker, Joshi, and Boulware-Gooden, (2010) argued in their study of spelling related teacher knowledge, such domain specific content knowledge is important for teachers of all students, but especially for teachers of students with dyslexia or other language-based learning disabilities.

In conclusion, there is consensus in the extant literature about the importance of reasonably accurate teacher judgment as well as about the importance of a sufficient knowledge base for effective teaching, but no studies have explored the relationship between such knowledge and teacher judgment accuracy. The present study attempted to explore such a

relationship in the specific context of teacher knowledge of developmental spelling and teacher judgment of spelling.

CHAPTER 3

METHOD

The primary purpose of the present study was to investigate the accuracy of teacher judgments of students' spelling. As noted previously, I defined teacher judgment (TJ) as teachers' precision in pinpointing their students' spelling proficiency. The study adds to prior research regarding TJ with its unique focus on the estimation of students' spelling skills and exploration of variables related to differences in judgment acumen among teachers. The study explored two research questions: a) how accurately do teachers judge students' concurrent performance on individual items of a developmental spelling test? and b) what is the relationship between teachers' accuracy in judging student performance on spelling and their developmental spelling knowledge? To address the first question, I asked teachers to group administer a brief spelling inventory to students in their class and then, without reviewing the results, to tell me for each item whether six students who I selected to represent lower, middle and high scorers would have spelled the item correctly. I then compared teachers' judgments to students' actual performance on the same items. To explore the second question, I presented teachers with a spelling error analysis task that was designed to assess knowledge of developmental spelling, and then examined the relationship between that knowledge and their performance on the teacher judgment task.

Procedures

Sample

I recruited participants for the study using a description of the study that I emailed to 320 first and second grade teachers in the State of Maine. The recruitment letter noted that there were two components to participation: group administration of a brief spelling inventory to students in

their class and completion of a brief interview on the topic of spelling (see Appendix A). I limited recruitment to a particular geographic region due to the need to travel to school sites to interview teachers face-to-face. As McMillan and Schumacher noted (2010), response rates can be substantially increased by offering monetary rewards for participation. To that end, I offered all volunteers a \$25 Amazon gift card for their participation. Thirty-six teachers representing 23 schools and 18 districts agreed to participate. Due to a health issue, one participant had to drop out before completing all parts of the study, resulting in a final sample of 35 teachers, 17 who taught Grade 1 and 18 who taught Grade 2. Additional teacher characteristics are included in the next chapter.

Data Collection

Data collection for the present study proceeded in two phases. In the first phase, teachers group-administered a spelling inventory to students in their class. In the second phase, teachers participated in a face-to-face interview with me in which they provided information about their professional background and completed two primary tasks: estimating student performance on a spelling inventory and analyzing a set of spelling errors.

Spelling test administration. Although research designs vary, in TJ studies, teachers have typically been asked to estimate their students' achievement on a dependent measure that has been previously administered to students without the teacher having access to students' actual results (e.g., Doore, 2010). Students' actual performance on the measure has then been compared to teachers' judgments about how their students had performed on the measure. Often, the researcher or a research assistant has been present for the administration of the measure or the measure has been administered by someone other than participating classroom teachers to ensure that at the time they were interviewed, teachers would have been unaware of students'

actual performance on the measure (e.g., Beswick, Willms & Sloat, 2005; Demaray & Elliot, 1998; Feinberg & Shapiro, 2003). In the present study, the dependent measure was a brief spelling inventory that I adapted from Ganske's (2013) Developmental Spelling Assessment (DSA); however, I was unable to oversee the group administration of the inventory myself. Instead, I employed two strategies to reduce the likelihood of teachers in the present study reviewing results for individual students prior to my interview with them. First, the initial recruitment email was intentionally quite general with respect to the purpose of the study. I did not tell teachers in advance that I would ask them to estimate the performance of their students on the spelling inventory. Second, the instructions for group test administration of the spelling test included an acknowledgment that it was critical to the integrity of the study that they not look at the results or discuss the results with anyone before the interview.

Once teachers indicated willingness to participate, I sent a self-addressed, pre-paid return envelope to their school address. After notifying me of receipt of the return envelope, I provided, by email, a packet in pdf format that included a 12-item developmental spelling inventory, a blank student answer sheet, and a detailed protocol for administration of the assessment that I adapted from Ganske's (2013) DSA. The original DSA comprises two parallel forms, A and B, containing 25 words that capture information on students' ability to spell various features. Some items, for example, assess whether students can spell words with consonant blends at the beginning (e.g. *slid*), while others assess whether students can spell a similar feature when it is present at the end of a word (e.g. *nest*). Other items represent theoretically different and later stages of spelling development. For example, the Within Word Feature Inventory assesses whether students can spell words with long vowels using the VCVe pattern (also known as the silent *e* pattern – e.g. *ripe* or *frame*). To reduce time for

administration, and also to reduce the likelihood of teachers using prior familiarity with Form A or B of the original inventory, the twelve items were drawn from both lists. The spelling inventory and instructions for administration are included as Appendix B. Immediately after group-administering the test, teachers chose a student or other adult to collect the student papers and put them in the mailing envelope to return to me. All participants administered and returned the spelling inventory between June 1st and June 16th 2017.

Student selection. As soon as I received student test papers, I hand-scored each one by assigning 1 point for each correctly spelled item; no partial credit was assigned to any individual items. I then used a stratified sampling strategy to select six students from each teacher's classroom. First, to ensure that teachers would have the opportunity to judge students representing a broad range of spelling ability relative to their same classroom peers, I selected students from each teacher's classroom whose scores represented different levels of achievement on the 12-item test. Initially, I intended to select two students who had scores between 1 and 4 as examples of low achievers, another two students with scores between 5 and 8 items to represent midrange students, and two students with scores between 9 and 12 to represent high scoring students. Because I did not have access to all student scores before beginning to schedule interviews, I selected these values by simply dividing the possible score range by three. Because mean performance of second grade students was higher than performance of first grade students, and because teachers varied in number of students scoring within each range, selecting an equal number of students from high, middle and low score bands from each teacher and grade was, in the end, not possible. Accordingly, I adjusted the score bands downward for categorizing Grade 1 students (and for one Grade 2 class with no high performers) categorizing scores between 1 and 3 as low, scores between 4 and 7 as midrange, and scores between 8 and 12 as

high. Second, I was also looking for equal gender distribution within each score band, so the next selection variable was gender, although exact gender balance was not possible (males, $n = 108$; females, $n = 102$). No information regarding students' race, SES, or disability status was available, so those variables were not selection factors. For statistical analysis purposes, student performance data on each item was coded using a simple system where 0 indicated a wrong spelling and 1 indicated a correct spelling. Using SPSS, the data was summarized and compared with teacher judgment data described below.

Teacher interviews. After scoring the spelling tests, I then invited teachers to schedule a face-to-face interview with me. All interviews were conducted between June 12th and July 13th 2017. The interview comprised three components. The first component included a restatement of the general purpose of the study and confirmation of each teacher's willingness to participate. I then asked interviewees to summarize their professional background by describing their undergraduate major, highest degree attained, years of teaching experience, and areas of certification. I also asked them to self-assess their familiarity with various commercial spelling programs on a scale from 0 to 5 where 0 indicated "never heard of; never used," and 5 indicated "very familiar; extensive experience." The interview protocol is included as Appendix C. I included a list of seven programs, some grounded in developmental spelling research and some developed according to different conceptual frameworks. I selected these programs based on my own reading, conversations with teachers, and internet searches regarding adoption of spelling programs by large districts in states like California, Texas, Florida and New York. I also invited participants to add to my list any other programs with which they were familiar and to rate themselves using the same scale. In alphabetical order by title, the programs about which I inquired were: *Making Words* (Cunningham & Hall, 2009); *Saxon Phonics & Spelling*

(Simmons, 2006); *Sitton Spelling and Word Skills Sourcebook for Teachers Level 1* (Sitton, 2006); *Sitton Spelling and Word Skills Sourcebook for Teachers Level 2* (Sitton, 2006); *Spelling Connections 1* (Gentry, 2016); *Spelling Connections 2* (Gentry, 2016); *Word Journeys* (Ganske, 2013); *Word Matters* (Pinnell & Fountas, 1998); and *Words Their Way* (Bear, Invernizzi, Templeton & Johnston, 2012).

The second part of the interview consisted of the teacher judgment task. I began by telling teachers that I had selected six students from their class, and that I would ask them to make judgments for each of the six students about their success on each of the 12 items on the spelling inventory they had recently administered. Procedurally, they judged all six students on each item before moving on to the next item. I began, in other words, by asking each teacher to first consider the word *hurt* and to indicate by marking *X* for no or *O* for yes whether they thought each of the students selected had spelled that word correctly. We then moved on to the word *frame* and continued until teachers had completed the same procedure for each of the twelve words on the assessment. According to Coladarci (1986), such a procedure may reduce a response bias based more on student characteristics than on task characteristics. I used this data to calculate a judgment accuracy score based on the number of times a teacher's judgment about a student's performance on a particular item was either correct or incorrect. When a teacher was correct (either because she had predicted a student would spell an item correctly and the student did spell it correctly or, conversely, the teacher assumed the student would not be correct and the student was in fact unable to spell an item), I coded the instance as a "hit" (entered as a 1 in SPSS) and when a teacher was incorrect, I coded it as a "miss" (entered as a 0 in SPSS). To calculate hit rate, I used SPSS to divide the number of each teacher's hits by the total number of judgments made.

Finally, in the third and last part of the interview, teachers completed an error analysis task. The list of errors examined is included as Appendix D. Specifically, I introduced a set of misspellings as being representative of errors commonly encountered by researchers who have studied spelling development (e.g. Read & Treiman, 2013) and asked them to tell me what the errors indicated about what the student knew and did not know about spelling. When the teacher had finished commenting on the individual items, I then asked her to comment on the set as a whole. Comments for this part of the interview were recorded for later transcription and analysis. Scores on the task were entered in SPSS and analyzed in relation to judgment accuracy.

Measures

An Overview of Ganske's (2013) Developmental Spelling Assessment

The 12-item spelling inventory that I developed for use in the present study was a modified version of Ganske's (2013) Developmental Spelling Assessment (DSA), a measure that was based on developmental spelling theory. The typical protocol for administering the DSA includes two phases. An initial screening inventory is used to determine a child's current stage of development (which I will refer to as the stage inventory) and, once that has been established, a feature inventory highlights strengths and weaknesses in using specific orthographic features within each stage. The words on the inventories were designed to elicit specific types of spelling knowledge that researchers have shown typically unfold in a predictable developmental sequence (e.g. Masterson & Apel, 2000).

According to Ganske (2013), items for the DSA were selected from a list of sight words compiled by Harris and Jacobson (1972). Their *Basic Elementary Reading Vocabularies* encompasses words most frequently encountered in pre-primer through eighth grade instructional

texts. The 20-item list for Ganske's DSA stage inventory contains four sets of words that become progressively more difficult. Scores are calculated based on the number of words spelled correctly in each subset. When a student has spelled only 0 or 1 word correctly in a subset, testing stops. Items selected for the feature inventories were chosen on the basis of prior research which has described predictable stages of spelling development through which learners pass as they progress toward more conventional spellings (e.g. Read, 1971; Beers & Henderson, 1977; Henderson & Templeton, 1986). To compare the relative difficulties of the words selected for inclusion in the DSA feature inventories, Ganske also consulted a Houghton Mifflin spelling series and Henderson's *Teaching Spelling* (1990). To avoid confounding factors of syllable-related issues, Ganske only included single syllable words for the Letter Name and Within Word pattern lists.

The formal DSA protocol assumes that once a student's stage of development has been established based on the DSA stage inventory, a follow up 25-item feature inventory will also be administered. Each of five features characteristic of each stage of development were tested by five different items to allow generous opportunity for chance errors. Appendix E includes a complete listing of the features with exemplars from each stage. Scores on the stage and feature inventories range from 0 to 25. Based on her research, Ganske proposed that scores between 12-21 indicated that the level was within the child's instructional range. Scores above 21 suggested no need for further instruction on features at that stage, and scores below 12 suggested too many confusions for instruction at that stage to be useful.

The psychometric properties of the assessment were investigated by Ganske (1999) using a sample of just over one thousand students from two central Virginia school districts in Grades 1 to 8. With respect to reliability, the consistency, or dependability of a measurement, Ganske

reported coefficients of internal consistency of .91 for the younger students (Grades 1 to 4). Regarding test-retest reliability, the coefficient for Grades 1 to 4 ($n=137$) was .94. For most research purposes, reliability coefficients of .80 are considered within acceptable limits (Binks-Cantrell, Joshi & Washburn, 2012).

Details of the Modified Spelling Inventory Used in the Present Study

For purposes of this study, I created a shortened test using items from both Form A and B of Ganske's (1999) original inventory. The words tested were: *hurt, frame, shallow, chop, trample, clapped, with, paint, drive, bridge, yawn* and *bump*. I included items from the Letter Name (LN), Within Word (WW) and Syllable Juncture (SJ) feature inventories to ensure that students would be unlikely to spell all words correctly. This decision was based on Ganske's field testing which indicated that early elementary students (grades 1-2) were unlikely to test beyond the Within Word (WW) stage. In the Ganske (1999) study, she reported that 69% of first graders tested at the LN stage, with another 21% at a pre-LN stage and only 9% at the WW stage and 1% at the SJ stage. Among second grade students, 44% tested at the LN stage and 47% tested at the WW stage, leaving 10% scoring at Syllable Juncture stage or above (Ganske, 1999). LN spellers tend to be beginning readers, and most beginning readers are LN spellers. Letter Name spellers are common in first grade and in the first half of second grade, and in some cases, kindergarten. Older students who struggle with reading may also have significant gaps in understanding of concepts in the LN stage (Ganske, 1999). In the present study, a hybrid iteration of the test was created in order to streamline test administration, but care was taken to sample knowledge of most of the features within the LN stage; all of the features within the WW stage and some of the features of the SJ stage.

Error Analysis Task

As Ganske (2013) has noted, there is an important difference between recognizing an error and interpreting it for appropriate instruction. Subject matter knowledge for literacy instruction (including spelling) encompasses multiple domains. In the past, many researchers have focused on teacher content knowledge of the structures in spoken and written language as key elements for effective reading and spelling instruction (e.g. Alatalo, 2015; Masterson & Apel, 2000; Malatesha Joshi, et al., 2009; Moats, 1994, 1999; Moats & Lyon, 1996). Such studies have often been conducted using measures that focus heavily on teacher's meta-linguistic knowledge. Items in a frequently adapted Moats survey (1994) ask participants, for example, to count the phonemes or morphemes in spoken words.

Critics, however, have called for measures of broader scopes of teacher knowledge, particularly for items that assess teacher's ability to apply knowledge to the context of teaching (e.g., Carlisle, Kelcey, Rowan, & Phelps, 2011; Foorman & Moats, 2004; Phelps & Schilling, 2004). The present study included a brief, exploratory measure of developmental spelling knowledge: a spelling error analysis task developed for use in this study. Procedures for developing, administering, and scoring the task are described below.

Development of the error analysis task. The task was inspired, in part by the methodology of researchers in mathematics education (e.g. Jüttner & Neuhaus, 2012) and also by previous literacy-related teacher knowledge assessment research (e.g., Carreker, Joshi, and Boulware-Gooden, 2010; Ediger and McCormack, 2015; and Spear-Swerling & Cheesman, 2012). In the latter study, the researchers described the creation of their multiple-choice survey as intentionally patterned on a teacher licensure exam that they described as “focused on pedagogical content knowledge” (p. 1699) and as including many application items “situated in

classroom contexts” (p. 1691). One survey item, for example, displayed spelling errors by various children and asked test takers to use the misspellings to decide which child’s spelling revealed the weakest phonemic awareness. The purpose of the error analysis task in the present study was to look for similarities and differences in the way teachers interpreted spelling errors. To do so, I selected ten misspelled words to represent misspellings of various features assessed by the DSA inventories at the LN, WW, and SJ stages. Seven of the misspellings represented actual errors made by Grades K-6 students of six teachers who assisted me in spring 2016 by piloting a spelling inventory very similar to the one used for the present study. The other three misspellings had been discussed in previous research (Ganske, 2013). A word processing error not discovered until after the first three interviews resulted in my not being able to use one word. I had intended to present *moment* misspelled as “momment,” but did not detect that it had been auto-corrected until a participant pointed it out. The nine words and misspellings analyzed by all participants were: *yes* spelled as “ues;” *shop* spelled as “sop;” *rip* spelled as rep; *van* spelled as “ven;” *drop* spelled as “jrop;” *hunt* spelled as “hut;” *girl* spelled as “gril;” *seat* spelled as “ct;” and *trotted* spelled as “chrotid.”

The development of the task was also based on a conceptual framework for “expert noticing” put forth by Ross and Gibson (2010) and also by the observations of Leinhardt (1983). Ross and Gibson’s analysis of different teachers’ comments while watching a video-taped literacy lesson confirmed, they asserted, important differences between the observations of literacy experts versus novices. Experts responses, they contended, included “more detailed hypothesizing and more breadth and depth of elaboration” (p. 175). The comments of experts also demonstrated knowledge and showed their ability to recognize meaningful patterns in data. Consider the following extended example. As second grader, Cherie [pseudonym], made the

word *rest* during a phonics-focused portion of a videotaped lesson, the comments of one novice observer exemplified what Ross and Gibson described as unembellished statements of behaviors. “The student has to make the word *rest*. She’s selecting the tiles, and she did it correctly” (p. 185). In contrast, consider the comments of an expert viewing the same segment from the lesson. “Cherie sounds out each letter, /r/ /e/ /s/ /t/, carefully. So, she has the letter boundaries, the phonemic awareness of these letters in her head” (p. 185). Later, as the teacher prompts Cherie to make the word *tigers*, one expert observed:

The child first repeated the word. The child, when putting [the word] together, seemed to be articulating. Some letters could be heard, others seemed to be subvocalized to put together *tigers*. She left out the *E*, and after the teacher said there was one letter missing and that it was an *E*, the first comment from the child was that it would go at the end.

Such observations, Ross and Gibson (2010) maintained, would for expert teachers, represent clear evidence of a particular stage of spelling development and likely motivate the teacher to plan instruction to help move the student from that (alphabetic) stage into the next (pattern-based) spelling stage (Bear, Invernizzi, Templeton, & Johnston, 2012). Implicit in such an inference is that expert teachers know and can apply information about qualitative differences in spelling strategies and also that they understand the sequence of spelling strategies typically employed by young learners as outlined in prior research.

Leinhardt (1983) also reported salient differences between novices and experts while engaged in a literacy-related judgment task. She noted that experts’ predictions about word reading were more likely to be grounded in observations about a particular word’s structure or knowledge of a scope and sequence of the curriculum, while novices seemed to focus more on whole words or personality traits as reasons for responses. As an example, while commenting on

the ability of students to match an orally presented definition to the written word *bake*, one more novice participant noted that the task would be easier for girls because “they’ve made a cake with their mother,” and more difficult for a male student because “I don’t think he’s interested in making a cake” (p. 12). Teachers, on the other hand, Leinhardt wrote, referenced phonics patterns or elements they knew they had or had not taught. Of *bake*, one said, for example, “we didn’t do the ~ake cluster” (p. 12).

Administration of the error analysis task. The misspelled words were presented on index cards that I shuffled before presentation, and I invited participants to work through the set in whatever order they chose. Before the task, I read from the following script:

The last thing I’m going to ask you to do today is look at some misspellings. None of these are items from the spelling test you gave, but they are typical of early elementary spellers, and what I’d like you to do is to put on your diagnostic hat and think aloud for me about what is going on with a child who makes such an error. What do these misspellings tell you about what a child who makes such error already knows or still needs to learn about spelling? This is intentionally a very open ended task so you can talk about the words in whatever order you’d like or come back and make additional comments at any time. You may notice at some point that as you talk, I’ll make a comment for the audio record like “she’s talking about ‘shop.’”

When participants indicated they were finished, I read from this script to probe for more information:

So now that you’ve thought about these words, I wonder if you could comment on them as a set; in other words, how representative does it seem to you of the kinds of errors you actually see? Does anything stand out that seems like it doesn’t belong, and on the flip side, can you think of something you see a lot of that’s missing?

Scoring the error analysis task – Step one. Previous research such as that by Leinhardt (1983) revealed a tendency for some teachers to attribute errors to student work habits. In a pilot for the present study, for example, one teacher’s explanation for spelling *hunt* as “hut,” captured the gist of similar comments by many. “They [were] probably just hurrying. They need to slow

down.” Although teachers in the present study also hypothesized about student characteristics as a source of difficulty in explaining a particular misspelling, I did not assign points to such explanations. Rather, the intent was to investigate the extent to which the explanations reflected knowledge of spelling from a developmental perspective, as well as the extent to which explanations reflected additional characteristics of “expert noticing” described earlier. Therefore, the rubric developed for analyzing this task was based on references to distinctions in word structure and developmental sequences well documented by developmental spelling researchers and based on statements that revealed recognition of meaningful patterns or categories.

Responses to each misspelling earned points for appropriate reference to one or more of the following findings – all of which represent issues of consensus in developmental spelling research. The list was developed also based, in part, on the system offered by Ganske for scoring the Developmental Spelling Assessment and associated Feature Inventories. Explanations that involved more than one of the issues described below received multiple points.

1. Letter names can provide misleading cues to spelling. For example, the letter name associated with the character *Uu* may help explain the misspelling of *yes* as “ues.” (e.g., Beers & Henderson, 1977; Block & Duke, 2015; Bourassa & Treiman, 2001).
2. The position of a sound within a word impacts success in spelling it. For example, children can more easily represent the first and last sounds in the word *fast* than the vowel or the interior /s/ sound (e.g. Ganske, 1999; Henderson & Templeton, 1986; Invernizzi & Hayes, 2004).

3. Spelling blends and digraphs is qualitatively different than spelling single consonant sounds. For example, spelling *which* represents a greater challenge than spelling *win* (e.g., Cunningham, Zibulsky & Callahan, 2009; Ganske, 1999; Moats, 1994).
4. The sounds of the letters *n* and *m* belong to a special class of consonants (nasals) that affect the articulation of nearby consonants and vowels; *ng* also belongs in this category although no words were presented with this sound or orthographic feature. For example, though both are spelled with an *a*, the vowel in the word *cap* is quite different from the vowel sound in the word *man* which is affected both by the nasal that precedes and the one that follows it (e.g., Beers & Henderson, 1977; Ganske, 1999; Invernizzi & Hayes, 2004; Phillips, Clancy & Lonigan, 2008; Read, 1975).
5. Some words are influenced by an /r/ sound that significantly obscures any vowel sound. For example, though spelled differently in each case, the sound that follows /g/ in the word *girl* is the same as the sound in the middle of the words *hurt* and *her* (e.g., Block & Duke, 2015; Ganske, 1999; Hurford, et al., 2016).
6. The blends *tr* and *dr* belong to a unique class of blends. In these blends, articulation of the *t* and *d* is routinely affricated by adults and children. Such legitimate pronunciation does *not* imply a speech deficit to be corrected. For example, spelling *dragon* as “jargon” is both common and logical (e.g., Henderson & Templeton, 1986; Read, 1971; Scharer & Zutell, 2003; Treiman, 1985; Werfel & Schuele, 2012; Worthy & Invernizzi, 1990).
7. A predictable rule for doubling consonants in spelling exists. For example, the correct spelling of *clapped* is predictable based on the short vowel in the base word *clap* (e.g., Alatalo, 2016; Invernizzi, Abouzeid & Gill, 1994; Moats, 2005).

8. Success in spelling particular features of words can be expected before others. For example, the ability to spell the short vowel in the word *fat* is likely to occur long before the same child can spell the long vowel in the word *frame* (e.g., Beers & Henderson, 1977; Henderson & Templeton, 1986; Invernizzi & Hayes, 2004; Treiman, 2017).
9. Misspellings of short vowels are rarely random; instead they are characterized by substitutions that are predictable according to similar articulation features; namely, short vowels classified as front (e.g. short i and short e and short a) are more likely to be confused with each other than with those classified as back (e.g. short o). Regional dialects substitutions are similarly predictable on the same basis. For example, words like “tin” and “ten” are often homonyms in some dialects, but words like “pot” and “pet” are never homonyms in the dialects of native English speakers (Ehri, Wilce & Taylor, 1987; Mann, Tobin, Wilson, 1987; Read, 1971).

After analyzing responses based on reference to the nine specific points listed above, I again reviewed explanations for the errors presented and also examined teachers’ responses to my follow up question looking for evidence of distinctions between novices and experts as observed by prior researchers (e.g., Leinhardt, 1983; Ross & Gibson, (2010). Specifically, I assigned additional points for expertise based on previous research which has shown that categorical thinking is one characteristic that distinguishes novices from more expert practitioners (e.g. Bransford, Brown & Cocking, 2000; Chi, Feltovich, Glaser, 1981; Leinhardt, 1983). Specifically, points were assigned for responses that demonstrated either or both of the following two characteristics.

10. The ability to quickly and accurately categorize errors based on similar word structure represents perceptual awareness in recognizing meaningful patterns (Ross & Gibson, 2010).
11. The ability to quickly and accurately notice features missing from the set is indicative of thinking which is categorical or pattern based in nature and also implies greater depth and breadth of elaboration (e.g. Bransford, Brown & Cocking, 2000; Chi, Feltovich, Glaser, 1981; Leinhardt, 1983; Ross and Gibson, 2010).

Scoring examples. To illustrate, I begin by presenting several examples of teachers' reactions to spelling *yes* as "ues" which were awarded no points and conclude with an example of a response that did receive a point. One teacher offered, "I think that . . . I notice that with my struggling readers, y is hard sound for them to keep in their head." Because the response included neither direct nor indirect reference to any of the nine points described above, it was awarded no points. Another teacher remarked "That one makes me sad . . . because it's such a common word, you know. My kids should know how to spell *yes*. But y is tricky because it can make different sounds; sometimes it's a vowel, sometimes not." Another response also referenced the word's relative high frequency. "That's another word that we have all around the room and they see a lot of the time . . . I think again it's recognizing those vowel sounds." Someone else replied simply, "I'm surprised. I don't see anybody that uses a *u* for the *y*." In contrast to the kinds of responses just described, another teacher considered the substitution of "u" for "y" and announced "Oh! well it makes the *u* sound /y/, /y/, /yu/, so I could see that in terms of the sound that *u* makes." Although this teacher did not use the label Letter Name strategy in her description, she voiced recognition that letter names might constitute a rationale for the misspelling. According to the rubric developed for this study, only the last explanation

was awarded credit for addressing the first point listed above regarding the role of letter names as a spelling strategy.

Scoring the error analysis task – Step two. In a second step, I reviewed responses for categorization of errors and awarded points for such accurate observations. Several teachers noted, for example, that both *rip* and *van* involved the misspelling of short vowels or that *drop* and *trotted* constituted a set of words that begin with consonant blends. I also looked for mention of missing features (i.e., those commonly assessed according to a developmental spelling framework but not included in the set presented for analysis). As an example, several teachers noted that none of the words misspelled involved understanding of the silent *e* spelling pattern. A few noticed that none of the words had pluralization. Such reflections were awarded points for each feature mentioned. In cases where participants confirmed that features they saw in the set were common among their own students but offered no examples of features missing, no additional points were given. One teacher concluded, for example, by saying “You have the ‘tr’ as ‘chr.’ I didn’t mention it when I did that card, [“chrotid”] but that ... *truck* ... with ‘chr’ I see that a lot.” I did not assign points for this observation.

Procedurally, I first coded each teacher’s responses to the nine misspellings and I then examined teachers’ responses across the task as a whole to determine whether comments reflected categorization of errors and/or noted features taught within a developmental spelling framework but missing from the set. I awarded additional points accordingly and calculated a final score for the error analysis task. Because teachers were free to offer as many observations as they wanted to about the individual errors and about the set as a whole, there was no maximum or perfect score for the task. A transcribed example of one teacher’s responses has been included as Appendix F.

Reliability of Scoring. I worked with a faculty member in the College of Education and Human Development with knowledge of developmental spelling research to review the scoring rubric for content validity and then, to estimate the reliability of scoring. Specifically, the faculty member double-scored a sample of seven teacher transcripts (20% of the total), and I then calculated inter-rater agreement by counting the number of times we were in absolute agreement on points assigned on each of the nine error analysis items, as well as on the tenth scoring dimension described above, and then dividing that number by the total number of ratings. We were in absolute agreement 90% of the time (i.e., 63 of 70 ratings were in agreement) which suggested that the rubric provided a reliable framework for evaluating the kind of information described by teachers in their responses to each of the misspellings and to the set as a whole.

Data Analysis

Using SPSS, I computed descriptive statistics relating to judgment accuracy. I also used correlational analyses to explore whether judgment accuracy was related to teacher, student, and/or item variables and to explore whether judgment accuracy was related to scores on the error analysis task. The results of those analyses are described in the next chapter.

CHAPTER 4

RESULTS

This chapter describes the volunteer sample and the results of statistical analyses related to two research questions: a) how accurately do teachers judge students' concurrent performance on individual items of a developmental spelling test? and b) what is the relationship between teachers' accuracy in judging student performance on spelling and their developmental spelling knowledge?

I begin with a description of the sample of teachers and their students followed by analyses that investigated my first research question regarding teacher judgment (TJ) accuracy. These analyses addressed the relationship between TJ accuracy and (a) teacher characteristics such as education level and years of experience (e.g. Doore, 2010) and self-assessed familiarity with various spelling programs, (b) student factors such as gender and ability level (e.g. Begeny et al., 2011; Beswick et al., 2005; Hoge & Coladarci, 1989; Südkamp et al., 2012), and (c) item factors such as percentage of students able to answer an item correctly (e.g. Doore, 2010; Eckert, Dunn, Coddington, Begeny, & Kleinmann, 2006). I end the chapter with the results of analyses related to my second question. These analyses explored the relationship between teacher judgment accuracy and pedagogical content knowledge based on a spelling error analysis task.

Sample

Teacher Characteristics

As I mentioned above, the sample for this study included 35 teachers from 18 school districts and 23 schools. Seventeen taught grade one, and 18 taught grade two. Class sizes ranged from 10 to 21 students, $M = 14.83$, $SD = 2.81$. The demographics of the schools in which participants taught also varied. The free and reduced lunch rates of the schools ranged from 9-

78% eligibility ($M = 46.97, SD = 17.66$). School size ranged from a low of 91 students to a high of 616 students.

Education and experience. Highest level of education varied with 19 teachers reporting that the highest degree earned was a bachelor's degrees and 16 teachers reporting that they had earned a master's degree or beyond. Participants' years of professional experience ranged from three to 37 years ($M = 18$ years, $SD = 10$ years). The majority of teachers were certified as general elementary education teachers ($n = 19$); the others had elementary education endorsements plus one or more additional endorsements: special education ($n = 6$); early childhood, $n = 5$; secondary ($n = 1$); other ($n = 4$).

Familiarity with spelling programs. I also asked teachers to self-assess their familiarity and experience with seven spelling programs on a Likert scale ranging from 0 (*Never heard of; never used*), to 5 (*very familiar; extensive experience*). I also invited teachers to share other programs with which they had experience and to rate their familiarity with them using the same scale. Table 4.1 summarizes teachers' self-reports regarding experience with spelling programs from the list I showed them as well as other programs that they added to the list. As can be seen, teachers reported being most familiar with *Words Their Way* (Bear et al., 2012), followed by Pinnell & Fountas' (1998) *Word Matters*, and Cunningham & Hall's (2009) *Making Words*, M familiarity = 1.86, $SD = 2.00$. They reported being least familiar with Gentry's (2016) *Spelling Connections*, Saxon *Spelling* (Saxon, 2006) and Ganske's (2013) *Word Journeys*.

Because the items for the spelling test in the current study were drawn from Forms A and B of a developmental spelling assessment included in Ganske's (2013) *Word Journeys*, I was interested in participants' experience with that program. Twenty-six of the teachers described themselves as having "never heard of; never used the program," (indicated by a rating of 0). Of

the nine teachers who indicated some familiarity (i.e., a score of 1 or above), only one teacher rated her familiarity as 5 (i.e., *very familiar; extensive experience*). Two reported their experience as 4; another two rated their experience as 3, and four characterized their familiarity with *Word Journeys* as 1.

Table 4.1.

Self-reported Familiarity with Published Spelling Programs

Program	Teachers with no familiarity (<i>n</i>)	Teachers with at least some familiarity (<i>n</i>)	Familiarity Rating <i>M</i> (<i>SD</i>)
<i>Making Words</i> (Cunningham & Hall, 2009)	16	19	1.83 (2.02)
<i>Saxon Spelling</i> (Saxon, 2006)	24	11	0.57 (1.09)
<i>Sitton Spelling</i> (Sitton, 2006)	13	22	1.63 (1.83)
<i>Spelling Connections</i> (Gentry, 2016)	27	8	0.43 (1.04)
<i>Word Journeys</i> (Ganske, 2013)	26	9	0.66 (1.37)
<i>Word Matters</i> (Pinnell & Fountas, 1998)	10	25	2.31 (1.95)
<i>Words Their Way</i> (Bear, Invernizzi, Templeton & Johnston, 2012)	2	33	3.34 (1.80)
Other:			3.00 (2.00)
<i>Cast-A-Spell</i> (Fontenault & Salter, 1993)		3	
<i>Foundations</i> (Wilson, 2013)		6	3.67 (1.63)
<i>Phonics Lessons</i> (Fountas & Pinnell, 2002)		9	3.56 (1.33)

Student Characteristics

As I mentioned in chapter three, I selected six students from each teacher’s classroom based on a combination of performance on the spelling inventory and gender. Information

regarding students' race, SES, and disability status was not available. The raw scores of selected students on the twelve-item test ranged between 1 and 11 out of a possible 12 points. Grade 1 students ($n = 102$) spelled an average of 4.62 ($SD = 2.51$) words correctly, while Grade 2 students ($n = 108$) spelled an average of 6.32 ($SD = 2.95$) words correctly. The results of an independent samples t test indicated that the difference in the mean scores was significant, $t(208) = 4.50$, $p = <.001$), with second graders outperforming first graders.

Descriptively, the mean score of male students ($n = 108$, $M = 5.62$, $SD = .28$) was slightly higher than that of female students ($n = 102$, $M = 5.36$, $SD = .29$); however, an independent samples t test revealed no significant difference between the mean scores of the two groups, $t(208) = 0.65$, $p = .52$.

Teacher Judgment Accuracy

A primary purpose of the study was to investigate teacher judgment accuracy in the domain of spelling. To that end, as described in chapter 3, I calculated a judgment accuracy score (i.e., hit rate) for each teacher based on the number of times a teacher's judgment about a student's performance on a particular item was either correct or incorrect. Each teacher in the study judged the performance of six students on 12 items; thus, the total number of judgments for each teacher was 72. The average hit rate across the six students judged by each teacher ranged from .64 to .88 with a mean hit rate of .77, $SD = 5.5$, $Mdn = 76.40$.

I also explored the extent to which teachers over or underestimated their students' achievement. Table 4.2 depicts item level results for first grade teachers, showing the prevalence of over and underestimation for each of the 12 items on the inventory. Items coded as underestimates indicated that the teachers believed students would *not* spell the item correctly, but students did in fact spell the item correctly. Conversely, items coded as overestimates

represented the opposite scenario. Namely, teachers predicted students would spell the item correctly, but students made errors on the item. To calculate these percentages, I divided the number of overestimates by the total number of misses. In the current study, of 289 errors in judgment made by first grade teachers, 177 of the errors (61%) were overestimates, while 112 errors represented underestimates (39%). Descriptively, Grade 1 teachers were more likely to overestimate than underestimate student performance on nine of the twelve words and more likely to underestimate than overestimate performance on three words. Results of a paired t-test, however, indicated that the difference in percentage of over- versus underestimates was not statistically significant, $t(16) = -1.91, p = .07$.

Table 4.2.

Percentage of Grade 1 Teachers' Hits, Over and Underestimations

Item #	Hits	Underestimates	Overestimates
1 hurt	71	06	24
2 frame	69	15	17
3 shallow	84	10	09
4 chop	78	10	12
5 trample	79	13	08
6 clapped	75	03	23
7 with	74	07	20
8 paint	77	11	12
9 drive	75	12	13
10 bridge	89	06	05
11 yawn	77	11	12
12 bump	71	08	22

The outcome for second grade teachers was similar. Descriptively, overestimates also outnumbered underestimates (see Table 4.3). Specifically, 160 of 294 misses (55%) were overestimates, and 134 of the 294 misses (45%) were underestimates. Within the pool of 12 items, Grade 2 teachers were more likely to overestimate than underestimate performance on seven items and they were more likely to underestimate than overestimate performance on five items. Results of a paired t-test, however, indicated no significant difference in percentage of over versus underestimates, $t(17) = -.61, p = .549$.

Table 4.3.

Percentage of Grade 2 Teachers' Hits, Over and Underestimations

Item	Hits	Underestimates	Overestimates
1 hurt	70	11	19
2 frame	78	13	09
3 shallow	76	14	10
4 chop	88	02	10
5 trample	72	21	06
6 clapped	71	05	24
7 with	88	02	10
8 paint	80	08	12
9 drive	78	17	06
10 bridge	73	07	19
11 yawn	79	15	06
12 bump	75	09	16

The Relationship Between Teacher Characteristics and Judgment Accuracy

Prior research has examined teacher characteristics that may account for variability among teachers in judgment accuracy. In the present study, I examined four teacher characteristics: grade level taught, teaching experience, educational level, and familiarity with published spelling programs. Results of these analyses are described below.

Grade level. To test whether judgment accuracy was different for first versus second grade teachers, I used an independent samples t test. The results indicated that judgment accuracy of first grade teachers ($n = 17$, $M = .76$, $SD = .01$) was comparable to that of second grade teachers ($n = 18$, $M = .77$, $SD = .01$), $t(33) = .49$, $p = .63$.

Experience and education. Next I examined whether teachers' years of experience and education level were related to judgment accuracy. I first examined the hit rate for participants who did ($n = 16$, $M = .78$, $SD = .05$) and did not have a graduate degree ($n = 19$, $M = .76$, $SD = .06$). The results of an independent samples t test indicated that the two groups did not differ in judgment accuracy, $t(33) = -.72$, $p = .47$. As I mentioned above, teacher experience varied greatly, ranging between three and 37 years of experience ($M = 17.80$, $SD = 9.88$). I explored the relationship between experience and judgment accuracy by calculating the correlation between the two variables and found no significant relationship, $r = -.13$, $p = .45$. That is, teachers who were more experienced did not tend to be more accurate than teachers who were less experienced.

Familiarity and experience with spelling programs. I also investigated whether teachers' self-assessed familiarity/experience with various spelling programs was related to judgment accuracy. I conducted two analyses to address this question. First, I computed a spelling program familiarity total score by adding familiarity scores across all of the programs on

my list as well as on programs that teachers added to the list. As I mentioned above, teachers rated their familiarity with several programs on a scale of 0 to 5. Scores of 0 indicated no familiarity and scores of 5 indicated high familiarity. Total familiarity scores ranged between 5 and 22, $M = 12.70$, $SD = 5.61$. I then used those totals to examine whether teachers who had familiarity with a greater number of programs were more accurate in their judgments than those who had less familiarity with commercial spelling programs but found no significant correlation, $r = -.00$, $p = .99$.

Second, I investigated whether greater familiarity with particular spelling programs accounted for variability in judgment accuracy. I included in this analysis only programs that had a familiarity rating of 1 or above by at least half the teachers in the sample due to lack of variability in familiarity ratings for the other three programs (see Table 4.1). To explore whether knowledge of particular programs was related to judgment accuracy, I calculated correlations between overall hit rate and knowledge ratings for each of the four programs rated as at least somewhat familiar (i.e., ratings above 0) to at least half the teachers in the sample: *Making Words* (Cunningham & Hall, 2008), *Sitton Spelling* (Sitton, 2006), *Word Matters* (Pinnell & Fountas, 1998), and *Words Their Way* (Bear, Invernizzi, Templeton & Johnston, 2012). I found a significant relationship between overall judgment accuracy (hit rate) and knowledge for just one of the programs. Teachers who reported higher familiarity with *Word Matters* (Pinnell & Fountas, 1998) tended to have somewhat lower judgment accuracy scores ($r = -.41$, $p = .02$) than teachers who reported low familiarity with that program. Correlations between judgment accuracy and program familiarity were small and insignificant for *Making Words* ($r = -.10$, $p = .55$), *Sitton Spelling* ($r = -.08$, $p = .66$), and for *Words Their Way*, $r = .03$, $p = .86$).

As mentioned above, I was particularly interested in the results related to Ganske's (2013) *Word Journeys*. Because the items for the spelling test in the current study were drawn from that resource, I also used an independent samples *t* test to compare the mean hit rate for teachers who rated their familiarity with that program as 0 ($M = .77, SD = .01$) to those who rated it as 1 or above ($M = .78, SD = .01$) and found no difference in judgment accuracy.

The Relationship Between Student Characteristics and Teacher Judgment Accuracy

Based on previous research, I also explored relationships between teacher judgment accuracy and two student characteristics: gender and spelling achievement. I first conducted a paired samples *t* test to compare the accuracy of teachers in judging male ($M = 77.58, SD = 0.08$) versus female ($M = 76.96, SD = .08$) students. The results indicated no significant difference in judgment accuracy $t(33) 0.34, p = .74$. That is, teachers were just as accurate in judging male as female students.

Next, I investigated the relationship between judgment accuracy and student achievement level as measured by their raw scores on the spelling inventory. Because Grade 2 students scored significantly higher than Grade 1 students on the inventory, I conducted separate correlational analyses for first and second graders. The results indicated that teacher judgment accuracy did not vary by student spelling score for first graders ($r = -.09, p = .32$) or for second graders, ($r = .02, p = .81$).

The Relationship Between Item Characteristics and Judgment Accuracy

To explore differences in teacher judgment related to item characteristics (i.e. item difficulty), I calculated a difficulty score for each item by dividing the number of students who spelled each item correctly by the total number of students who attempted each item. Since Grade 2 students outperformed Grade 1 students, a separate difficulty score was calculated for

each of the two grades. Table 4.4 summarizes item difficulty by grade level. Items are listed from least frequently spelled correctly (i.e., most difficult) to most frequently spelled correctly (i.e., least difficult).

Table 4.4.

Percentage of Students Spelling Each Item Correctly (Item Stage and Feature)

Item	Grade 1	Item	Grade 2
clapped (SJ K)	3	clapped (SJ K)	13
bridge (WW I)	7	bridge (WW I)	17
shallow (SJ L)	12	shallow (SJ L)	31
trample (SJ O)	16	hurt (WW G)	38
yawn (WW J)	21	trample (SJ O)	38
hurt (WW G)	23	yawn (WW J)	38
paint (WW H)	27	paint (WW H)	48
frame (WW F)	59	frame (WW F)	67
drive (WW F)	63	bump (LN E)	79
bump (LN E)	69	drive (WW F)	82
with (LN E)	79	chop (LN D)	87
chop (LN D)	80	with (LN E)	87

Note. LN D = Letter Name stage, affricate feature; LN E = Letter Name stage, final blends and digraphs feature; SJ K = Syllable Juncture stage, doubling and final e-drop with ~ed and ~ing; SJ L = Syllable Juncture stage, syllable juncture doubling feature; SJ O = unstressed syllable vowel patterns; WW F = Within Word stage, long vowel silent e pattern; WW G = Within Word stage, r-controlled vowels; WW H = Within Word stage, other common long vowel patterns; WW I = Within Word stage, complex consonants; WW J = Within Word stage, abstract vowels.

Descriptively, students were most accurate in spelling features within the Letter Name (LN) phase; less accurate at handling Within Word (WW) features and least accurate at spelling words including Syllable Juncture (SJ) features. For the three LN items: *chop*, *with* and *bump*, on average three-quarters ($M = .76$) of first graders ($n = 102$) were able to spell them correctly. For the six WW items: *drive*, *frame*, *paint*, *hurt*, *yawn* and *bridge*, the average fell to .33, and for the three SJ words: *trample*, *shallow* and *clapped*, just .10 of first graders were able to spell those items accurately. Trends for second grade students ($n = 108$) were similar. LN stage items were spelled correctly by 84% of second grade students, WW items were spelled right by 48% of the same students, and SJ items were spelled accurately by just 27% of students. Letters in parentheses to the right of each item indicate the spelling stage and feature of each item according to the Developmental Spelling Assessment (DSA). As explained in the literature review, although the sequence of mastery is by no means fixed, research has shown that items coded as feature A (initial and final consonants) are generally mastered earlier than items coded, for example, as feature D (affricates). Item difficulty analyses in the present study were consistent with this sequence.

After computing an item difficulty index for each item, I computed the correlation between teacher judgment accuracy and item difficulty. For first grade teachers, I found a moderate negative correlation between the two variables, $r = -.48$, $p = 0.05$. This finding indicated that Grade 1 teachers tended to be more accurate in judging items that were more difficult for Grade 1 students to spell. That is, items that fewer students answered correctly were more accurately judged than items that more students answered correctly. For second grade teachers, in contrast, there was a strong positive correlation, $r = .74$, $p < .01$. Second grade

teachers, in other words, were more accurate judging items that were easier for their students to spell than items that were more difficult for students.

Error Analysis Task

The second question I explored was whether teacher knowledge of developmental spelling predicted teacher judgment accuracy. This section begins with a description of teachers' performance on the error analysis task, including the relationship between teacher knowledge and teacher characteristics. Next, I report on the results of correlational analyses of two variables, knowledge score and teacher judgment accuracy (hit rate). As described earlier, responses were scored first item-by-item based on consideration of nine specific findings from developmental spelling research (primarily related to word structure) and then reviewed based on the extent to which explanations revealed two additional characteristics of expert noticing as described by Leinhardt (1983) and by Ross and Gibson (2010), namely categorization of errors and identification of features missing from the set of misspellings presented for analysis.

The Relationship Between Teacher Characteristics and Developmental Spelling Knowledge

Across the 35 teachers who participated in the study, total scores on the error analysis task ranged between 1 and 13 points, $M = 6.34$, $SD = 3.04$ (see description of scoring rubric for the task in chapter 3). The scores for Grade 1 teachers ranged between 1 and 13 points ($M = 6.76$, $SD = 3.63$), while the scores of Grade 2 teachers ranged between 1 and 9 points, $M = 5.94$, $SD = 2.39$. An independent samples t test showed no significant difference between the two groups, $t(33) = .79$, $p = .43$.

Next, I examined the relationship between three teacher characteristics and developmental spelling knowledge: length of teaching experience, highest degree earned, and spelling program familiarity. As was the case in the analysis of the relationship between teaching

experience and hit rate, I found no relationship between developmental spelling knowledge and years of teaching experience, $r = -.13, p = .46$. In addition, there was no difference in mean knowledge scores between teachers with and without a graduate degree, ($t(33) = -.39, p = .70$), a finding that also mirrored results for hit rate. Lastly, I calculated the correlation between error analysis scores and overall program familiarity, a score I obtained by adding familiarity ratings across all seven programs that participants considered. The analysis yielded a small positive correlation, $r = .38, p = .03$, indicating that teachers who had greater familiarity with published spelling programs tended to have somewhat higher error analysis scores than teachers with less familiarity. I then examined the relationship between error analysis scores and specific program familiarity ratings for the four spelling programs that at least half of the teachers rated above a score of 0. As was the case in analyses of the relationship between judgment accuracy and program familiarity, I did not investigate the relationship between error analysis scores and program familiarity for the other three programs due to lack of variability in familiarity ratings (see Table 4.1). The results of correlational analyses indicated that there was no relationship between error analysis scores and familiarity ratings for *Making Words* ($r = .19, p = .28$), *Sitton Spelling* ($r = -.28, p = .10$), *Word Matters* ($r = .00, p = .99$), or *Words Their Way*, $r = .03, p = .86$).

The Relationship Between Error Analysis Scores and Hit Rate

Finally, I assessed the relationship between scores on the error analysis task and teacher judgment accuracy (i.e., hit rate). The results of that analysis indicated a moderate correlation between the two variables, $r = .42, p = .013$. Although modest, this result indicated that teachers who scored higher on the error analysis task tended to be more accurate in estimating student performance on the spelling inventory than teachers who scored lower on the error analysis task.

CHAPTER 5

DISCUSSION

Because literacy skills are so foundational to all of education, teachers' effectiveness in guiding students towards proficiency in this area is critical. One component of effective instruction is accurate teacher judgment. Both formal and informal estimations of students' particular strengths and weaknesses routinely guide teachers' practice. Teacher judgment informs instructional grouping decisions; teachers use estimations of students' current skills to respond to student questions and to choose examples for instruction. Judgment accuracy has the potential to influence adaptive word-level literacy instruction and is critical to setting and monitoring appropriate individual learning goals. Furthermore, as suggested in the first chapter, knowledge of typical and atypical developmental learning sequences may positively impact teacher judgment accuracy and also impact decisions about whether an individual student's struggles are atypical enough to warrant referral for special education services (e.g., Artelt & Rausch, 2014; Bates & Nettlebeck, 2001; Coladarci, 1992). To date, however, researchers have yet to investigate teacher judgment accuracy in the domain of spelling, nor have they considered a relationship between teachers' content knowledge in this area and judgment accuracy. The present study was designed to fill these gaps by addressing two questions. First, how accurately do teachers judge students' concurrent performance on individual items of a developmental spelling test? Second, what is the relationship between teachers' accuracy in judging student performance on spelling and their developmental spelling knowledge? In this chapter, I summarize the findings related to these questions and consider how they relate to previous research. I acknowledge limitations and make suggestions for future research and practice in this area.

Teacher Judgment Accuracy

The Grade 1 and Grade 2 teachers in the study judged, at the individual item level, the performance of six of their students on a 12 item spelling test adapted from a developmental spelling inventory by Ganske (2013). Two findings from the present study are particularly noteworthy. First, the results of this study corroborated the findings of previous research showing that teachers' judgment of their students' literacy skills are fairly accurate (e.g., Coladarci, 1986; Doore, 2010; Hoge & Coladarci, 1989, Südkamp, Kaiser and Moller, 2012). As described in chapter 2, many researchers have used correlational analyses to compare teacher judgments with student outcomes (e.g. Hoge and Coladarci, 1989; Südkamp et al., 2012). Only a small number of recent studies have used percent agreement methods or "hit rates" to investigate teacher judgment accuracy (i.e., Coladarci, 1986; Demaray & Elliot, 1998; Doore, 2010; Leinhardt, 1983). In the Demaray & Elliot study, for example, the researchers compared teachers' predictions of how students would perform on each item of a norm-referenced achievement test to students' actual performance on the test's reading, spelling and math items and found that hit rates ranged from .58 to .91; $M = .79$. In Doore's study of pre-school teachers' judgment of student performance on alphabetic knowledge items, the hit rates ranged between .56 and .83; $M = .70$. Hit rates in this study were comparable, ranging from .64 to .88 with a mean hit rate of .77 ($SD = 5.5$, $Mdn = 76.40$), indicating that overall, teachers were accurate in forecasting student performance on approximately three out of every four items.

Second, previous research has revealed a tendency to over rather than underestimate student performance on various measures of academic achievement (e.g., Begeny, Eckert, Montarello & Storie, 2008); Begeny, Krouse, Brown & Mann, 2001; Coladarci, 1986; Doore, 2010; Feinberg & Shapiro, 2003; Hamilton & Shinn, 2003). In the present study, the difference

in the prevalence of overestimates versus underestimates was not statistically significant for either first or second grade teachers, but the direction of the difference was in line with prior results, showing a greater tendency to overestimate by assuming that students would spell items correctly than to underestimate by assuming that students would not spell items correctly.

The four items most frequently overestimated by Grade 1 teachers were *clapped* (89% of misses on the item were overestimates); *hurt* (80%); *with* (74%); and *bump* (73%). The items most frequently overestimated by Grade 2 teachers were *clapped* (84% of misses were overestimates); *bridge* (72% of errors were overestimates); *hurt* (63%) and *bump* (also 63%). Notably, the items most commonly over-estimated by Grade 1 and Grade 2 teachers overlapped three out of four times. Each of the three words represents a different and qualitatively more advanced developmental stage of spelling.

The word *clapped* is from Ganske's (2013) Syllable Juncture stage inventory – a stage characterized as understanding spelling patterns that involve where syllables meet and understanding conventions for spelling meaning units such as prefixes and suffixes (including inflections). Theoretically, such aspects are generally mastered after spelling patterns for one syllable words, which are themselves mastered after a stage or phase which is characterized by a spelling strategy based primarily on sound. Though spelling researchers consider Syllable Juncture spellers as mastering such concepts gradually sometime between third and eighth grades, the percentage of overestimates by Grade 1 and Grade 2 teachers seemed to imply that teachers were unaware that such words represent such an advanced stage of spelling development. The word *hurt* was taken from Ganske's Within Word stage inventory and represents a word difficult to spell based on sound alone, and while many of the comments during the error analysis task showed that teachers understood there were special rules for

spelling ‘r-controlled’ vowels, other teachers’ comments revealed confusion about the source of difficulty, namely, teachers believed that children, in many cases, were hurrying and needed to slow down in speaking and spelling in order to attend properly to what they seemed to consider equally salient phonetic detail. Finally, the tendency to over-estimate their students’ ability to spell *bump* is not surprising given that so few teachers expressed awareness of the confounding role of nasality during the error analysis task.

Factors Impacting Judgment Accuracy

In the past, researchers have attempted to determine whether third variables such as teacher, student or item level characteristics were related to judgment accuracy (e.g. Bates & Nettelbeck, 2001; Begeny, Krouse, Brown, & Mann, 2011; Beswick, Willms & Sloat, 2005; Hoge & Coladarci, 1989). In this section, I first summarize key findings related to each of those factors and offer possible explanations for the findings.

Teacher Related Factors

As summarized in chapter four, few teacher characteristics appeared related to judgment accuracy. Overall, Grade 1 teachers were as accurate as Grade 2 teachers; teachers with less teaching experience were as accurate as teachers with greater experience, and teachers without graduate degrees were as accurate as teachers with graduate degrees. These results are in line with similar findings of previous teacher judgment research (e.g. Bates and Nettelbeck, 2001; Begeny, Krouse, Brown, & Mann, 2011; Valdez, 2013). In the present study, teachers who reported familiarity with a greater number of spelling programs were as accurate in their judgments as those who reported familiarity with fewer programs. I also examined the relationship between judgment accuracy and self-reported experience with specific programs that were at least somewhat familiar to at least 50% of the teachers: *Making Words* (Cunningham &

Hall, 2008); *Sitton Spelling* (Sitton, 2006); *Word Matters* (Pinnell & Fountas, 1998); and *Words Their Way* (Bear, Invernizzi, Templeton & Johnston, 2012). I found no correlation between judgment accuracy and familiarity with *Making Words*, *Word Matters*, or *Words Their Way*, but there was a moderate negative correlation ($r = -.41$) between familiarity with the resource titled *Word Matters: Teaching Phonics and Spelling in the Reading/Writing Classroom* by Pinnell and Fountas (1998).

Because this was an unexpected finding, I can only hypothesize about the meaning of the negative relationship. It may perhaps be related to the perspective of the program's authors regarding the continuum of learning in writing. Pinnell and Fountas (1998) explained that "early word solvers," for example, are characterized by the "use of simple words they know to derive new words by association or analogy" (p. 255). Such a description contrasts with that of Ehri and Robbins (1992) who wrote an article characterizing beginning readers and writers as unlikely to be capable of the analytic task of solving by analogy unless they have requisite decoding skills. Young children operating in what Ehri termed a *partial alphabetic phase* write some letters accurately but have difficulty remembering and writing words completely correctly because they have not yet achieved mastery of major letter-sound correspondences. Pinnell and Fountas (1998) hypothesize that spelling by analogy precedes a *later* phase in which children "hear most sounds, including the harder-to-hear vowels" (p. 256). Also, although the authors strongly suggest using a developmental spelling test as an indicator of children's growing ability to analyze words, the system of categorizing errors is quite general compared, for example, with that of resources like *Word Journeys* (2013) and/or *Words Their Way*, (2012). Column headings in a table included in the Pinnell and Fountas resource for recording and analyzing the errors of children (p. 121) seem to imply that high frequency words are to be instructionally targeted first,

followed by consonants including initial, final, clusters, and digraphs. In a third column heading, *Vowels: Long, short, vowel patterns, and y as a vowel*, vowel patterns are lumped together in a way that seems to imply that teaching might be simultaneously directed toward any or all of those features. Such an implication is a sharp departure from the much more fine-grained analysis suggested by Bear et al. (2012) and by Ganske (2013).

Pinnell and Fountas (1998) have argued against a predetermined sequence for decoding and/or encoding. Such skills should be taught instead, they have asserted, primarily in the context of students generating their own meaningful sentences and stories. To be sure, this interpretation of a relationship between judgment accuracy and experience with the Pinnell and Fountas resource represents an arguable level of conjecture. It is perhaps worth noting anecdotally, however, that when asked to suggest other programs with which they had experience, a quarter of the teachers named another Fountas and Pinnell resource. It remains possible, therefore, that teachers familiar with *Word Matters* might have been less likely to consider the relative difficulty of the features tapped by the items and consider instead factors such as how often that word might appear in stories, and so be less accurate in their judgments. At the same time, however, teachers with greater familiarity with spelling programs that did focus on a developmental progression of difficulty based on features (e.g. blends and digraphs in initial versus final positions) were not more accurate than teachers with lower familiarity scores, so an explanation for this finding remains speculative.

Student Related Factors

Previously, some researchers have reported a curvilinear relationship between student achievement and teacher judgment accuracy (e.g., Doore, 2010), a phenomenon for which a plausible explanation has been that teachers' strategies for predicting student achievement may

unwittingly be more related to a global perception of a student's ability than to achievement on particular items. In other words, teachers may judge students they perceive as high achievers as high performers on all items of a judgment task and judge those they perceive as low achievers as doing poorly on all items of a judgment task. For especially high and low achieving students, such a response strategy may be sufficient, whereas for students in the middle, consideration of additional factors (e.g. item difficulty) might play a more important role.

In the present study, teacher accuracy did not appear to be related to student achievement. This finding represents somewhat of a discrepancy when compared with Doore's finding and with previous research in which teachers were better at judging the performance of high achieving students than judging the performance of their low achieving students (e.g., Begeny, Krouse & Mann, 2011; Coladarci, 1986). A possible explanation may lie in a limitation in this study, namely relying on student performance on a brief spelling test as the sole measure of achievement level. In some teacher judgment studies (e.g., Hoge and Butcher, 1984), separate, broader measures of student achievement have been collected and used in explorations of teacher judgment accuracy. In the present study, I did not have access to any independent assessments of students' broader literacy skills.

Item Related Factors

As described in chapter 4, I found a moderate negative correlation between item difficulty and judgment accuracy for Grade 1 teachers. First grade teachers tended to be more accurate in judging items that were more difficult for Grade 1 students to spell. That is, items that fewer students answered correctly were more accurately judged than items that more students answered correctly. Second grade teachers, in contrast, were more accurate judging items that were easier for their students to spell than items that were more difficult for students.

A plausible explanation for these findings may be the curricular relevance of the judgment task. Since most teachers indicated that spelling tests were part of their curriculum, it seems likely that Grade 1 teachers would recognize words or word features that were quite difficult and that they rarely, if ever, asked Grade 1 students to spell. Although they could distinguish items likely to be of great difficulty, they may have been less sensitive to features of ostensibly easier words that also impact difficulty. Similarly, while second grade teachers could recognize words and features they expected to be easier, it might have been more challenging to apprehend distinctions in levels of difficulty because a greater number of items on the test were grade appropriate for second graders than for first graders. Of the 12 items on the spelling test used in the present study, three represented items from the Letter Name stage and six were items from the Within Word stage. Three items represented the Syllable Juncture stage. As reported in chapter 2, in Ganske's (1999) field study, 69% of first graders tested at the LN stage; 21% scored at a pre-LN stage, 9% scored at the WW stage and only 1% had scores within the SJ stage,. Among second grade students, 44% tested at the LN stage; 47% tested at the WW stage, and 10% scored at Syllable Juncture stage or above.

Teacher Knowledge and Teacher Judgment

A second purpose of this study was to explore whether and to what extent judgment accuracy might be related to developmental spelling knowledge as estimated by performance on an error analysis task. As Carreker, Joshi, & Boulware-Gooden, (2010) argued in their study of spelling-related teacher knowledge, such domain-specific content knowledge is important for teachers of all students, but especially for teachers of students with dyslexia or other language-based learning disabilities. Although the knowledge assessment measure was exploratory, the findings of the error analysis represent a unique contribution to the teacher judgment literature.

As described in the methodology chapter, the purpose of the error analysis task in the present study was to explore and compare differences in the way teachers conceptualized hypothetical spelling errors. The intent was to investigate the extent to which respondents might consider word structure and task characteristics from a developmental perspective or might, instead, hypothesize about student characteristics as a source of difficulty in explaining a particular misspelling. Teachers thought aloud about why *yes* had commonly been misspelled as “ues;” *shop* as “sop;” *rip* as “rep;” *van* as “ven;” *drop* as “jrop;” *hunt* as “hut;” *girl* as “gril;” *seat* as “ct;” and *trotted* as “chrotid.” The items were selected to represent misspellings of features assessed explicitly by the DSA (2013) inventories at the LN, WW, and SJ stages. Leinhardt (1983) pointed out how a think-aloud procedure allows teachers to call upon multiple, distinct sources of information.

In the present study, some teachers’ reasoning revealed knowledge of a typical developmental sequence of orthographic understanding, while others gave explanations based on understanding of individual items’ sources of difficulty. Some shared their assumptions that general student work habits affect achievement in general and also in spelling. Finally, many teachers expressed genuine puzzlement regarding the misspellings they were asked to analyze. Considered alongside teacher judgment accuracy for the 12 items on the spelling inventory given to their students, these comments provide a potential rationale for low hit rates on particular items.

Among the presumably easier to spell words from the Letter Name stage inventory, Grade 1 and Grade 2 teachers were least accurate in judging whether students would spell *bump* correctly. Grade 1 ($n = 17$) teachers were accurate on the item 71% of the time, and Grade 2 teachers ($n = 18$) were accurate 75% of the time. The lower accuracy for the item may be

explained, in part, by examining remarks made by teachers in Grades 1 and 2 during the error analysis task when asked to consider why *hunt* would be commonly misspelled as “hut.”

Explanations for the omission of *n* from *hunt* were characterized largely as reflective of work habits. “It’s like they just didn’t stretch it out enough.” “They probably hear the /n/ but they just keep going; they’re not slowing down.” “They just didn’t tap out that ending carefully enough.” “They didn’t say it carefully or listen to themselves carefully to the /n/ sound.” “I guess they just didn’t listen.” In some cases, though, teachers referenced the location of the blend at the end of the word as adding to the difficulty in spelling it. “They hear the beginning; they hear the end, but they miss what’s in the middle.” “That’s a common one...a lot of kids leave out ending blends; they just put the last sound.” Other teachers reported the absence of the *n* but had difficulty understanding why it might be missed. “That one’s very ... I don’t know.” Only a few teachers explicitly stated that the nasal in the blend might contribute to children’s difficulty in spelling it. One revisited her explanation about misspelling *van*, pointing and announcing “Again, it’s that ... I can’t remember the speech word for it, but it’s that *n*!” Two used the term “nasal” in their explanation, though one expressed some uncertainty about the label. “Yeah, some of the nasal sounds can be tricky sometimes... the *n* and the *m*... well, I don’t know if *m* is considered a nasal sound?” The other teacher returned to *hunt* late in the spelling error analysis task as she reflected on the set as a whole, looked up for several seconds, tapped conspicuously on the table twice and then declared “It’s that pre-consonantal nasal.” It is speculative, but certainly conceivable that understanding of the unique quality of nasals might contribute to more accuracy in predicting the correct spelling of words containing the feature.

Similarly, of the test items selected from the Within Word stage, Grade 1 and Grade 2 teacher accuracy indices for the word *hurt* were among the lowest for the twelve words (71% and

70% respectively). Some of the rationales offered for misspelling *girl* as “GRIL” suggest a possible explanation for the lower accuracy in judging the similar spelling inventory item *hurt*. Although a majority of teachers accurately used terms like *r-controlled vowels* to describe an understandable source of difficulty, for some teachers, their own competence as literate adults may have actually interfered with their ability to recognize that what is heard when listening to words with r-controlled vowels is not the issue. Although the word is *not* spelled as it sounds, some explanations implied that it is and contended that the issue was likely due, therefore, to work habits. “/gr-il/... that’s... I think another one where you need to really listen to what you’re saying when you say it.” Another teacher explained the difficulty similarly: “They are not saying the word really thoughtfully in their mind...I would talk [to them] about slowing it down and saying the word inside their mind or out loud and noticing that the /i/ is before the /r/.” If teachers are themselves unaware of differences between spoken and written words, it is understandable that they might be less accurate judges of the demands of the task of spelling such words and therefore less accurate judges of their students’ ability to do so.

Lastly, as mentioned earlier, Grade 1 and Grade 2 teachers were also relatively inaccurate (compared with their judgment accuracy for the other 11 items on the spelling inventory) at predicting whether students would correctly spell the word *clapped* (75% and 71% respectively). Because the word is structurally similar to the error analysis task item *trotted*, comments about the misspelling of that word as “chrotid” are perhaps relevant to explaining this finding. Many comments contained a combination of remarks about the worrisome implications of misspelling the initial blend along with the assertion that the misspelling of the past tense was perfectly normal. “I don’t think they are pronouncing [*trotted*] correctly and therefore they are spelling it as it sounds.” “So ‘chrotid’ is so common for the kiddo that kind of struggles speech wise.”

“Again, with *trotted* it’s trying to pronounce it carefully.” “Kids classically make this [“id” for “ed”] mistake until they get some hard core instruction around the suffix *ed*. We spend a lot of time with it because that is how it sounds – very phonetic.” “I get a lot of “id” for the suffix *ed*.” It sounds like /id/ but it’s not spelled that way.” “They just haven’t learned those three sounds that the *ed* suffix can make.” Only a minority called explicit attention to the word being multisyllabic or to the idea that its spelling also depended on mastery of a rule for doubling consonants. Just one person verbalized the rule for doubling: “The double consonant after a short vowel before adding a suffix ... that’s really hard for kids to remember,” and only one person referenced both the inflection feature and the doubling rule in terms of an expected developmental sequence: “I think about my *Words My Way* chart. Those [*ed* suffix sounds] are very late or beginning second grade, and double consonants are late second grade.” It seems reasonable to assume that teachers who understood the multiple considerations involved in spelling words like *trotted* and *clapped* correctly might also have been more accurate in judging whether their students were up to those demands.

In summary, in teachers’ explanations for spelling mistakes and confusions, common themes arose. In some cases, teachers made comparisons between errors and typical sequences of development, noting that certain features were not expected until later or seemed to characterize early attempts. Many teachers discussed the persistent tendency of children to spell phonetically even after other rules and strategies for conventional spelling had been taught. In some cases, however, children’s speech and auditory perception were described as a legitimate basis from which to write, while in others (primarily in the cases of words with affricates and nasals), teachers believed that children were hurrying and needed to slow down in speaking and spelling in order to attend properly to what they seemed to consider equally salient phonetic

detail. Because a considerable body of research exists which might guide teacher's understanding of and instruction relative to their students' spelling errors, further research on teachers' understanding of this body of research is warranted.

Limitations and Implications for Research and Practice

The findings of this study should be considered in light of important limitations related to the sample of teachers who volunteered for the study, the students who teachers judged, and the measure used to assess teacher knowledge, some of which have already been discussed. These considerations are discussed below, along with implications of the results for future research and for practice.

Sample of Teachers

A first limitation of this study was that it relied upon a sample of convenience within one geographic region in one state. As such, it is possible that the participants may have been influenced by common professional development in the area of spelling, and so be more homogeneous in their knowledge of spelling than might be the case in a more diverse sample of teachers. Second, the sample of students within schools served by participating teachers was relatively homogenous, linguistically speaking. It remains uncertain, therefore, whether results would generalize to teachers working with students with native speaker dialect issues or English as a second language issues. Third, teachers knew that the study would address spelling, so it is possible that teachers who volunteered to participate had greater interest or expertise in teaching spelling than might be the case in the general teacher population.

Fourth, given that seven teachers indicated some familiarity with Ganske's (2013) program *Word Journeys* and given that the resource contains the specific items used on the 12-item spelling test in the present study, it is possible that item specific knowledge may have

influenced the results in some way. Although familiarity with *Word Journeys* was not associated with higher mean judgment accuracy when compared to teachers who reported no familiarity, future researchers should consider using a spelling assessment that is not part of an established program.

Finally, no information about race, disability status, or academic achievement (in domains outside of spelling) of teachers' students was collected. Access to that information would likely add breadth and depth to the conclusions that can be drawn in the absence of such information. In particular, because some previous research has shown a tendency for teachers to be less accurate judges of students of lower performing students (e.g., Begeny, Eckert, Montarello & Storie, 2008), it is important to consider with caution the finding that teachers in the present study were equally accurate at judging the performance of students at both ends of the ability spectrum. It remains unclear whether students with low spelling scores in the present study were low achievers in all literacy areas or just on the items selected for the measure used in this study. Future research should aim to include and collect information regarding classroom diversity factors and to explore additional student achievement indicators beyond the spelling measure used in the teacher judgment task.

Assessment of Teacher Knowledge

A key limitation of the methodology of the study is the experimental nature of the error analysis task, and so the finding of a relationship between scores on the error analysis task and judgment accuracy should be considered tentative. In the absence of a valid and reliable measure of spelling related content knowledge, I developed a brief measure based on existing spelling research (e.g. by Scharer & Zutell, 2003) and on methodology often used by researchers interested in other disciplines such as mathematics education (e.g. Jüttner & Neuhaus, 2012) to

create a task I hoped would shed light on a topic not addressed by existing research. Likewise, although the rubric for scoring responses to the task was grounded in the findings of developmental spelling research, it was not finalized until after the task was administered. Had the rubric been finalized earlier, the interview protocol could have been adjusted to include probes and follow up questions more specifically aligned with the rubric. Finally, neither the error analysis nor the scoring rubric were subjected to validation outside the study, and so their reliability and validity require further investigation. Replication studies using this rubric might shed light on its reliability and validity. In addition, future studies should also focus on the refinement of the tool and the development of a more comprehensive assessment to measure knowledge of developmental spelling.

Finally, given that teachers with greater knowledge of developmental spelling on the error analysis task were more accurate judges of students' spelling skills, the present study suggests an important implication for practice – namely that teachers may benefit from more explicit exposure to the structure of the English language and the relationship between spoken language and written language as presented within a developmental spelling framework. Children's spellings provide insights into what they understand about language and how they interpret the literacy instruction to which they are exposed. A logical next step for future research is to examine whether teachers who are more knowledgeable about what their students know are more likely to have a positive impact on student achievement.

Conclusions

As many researchers have noted, having a clear sense of what students know is essential to planning effective instruction for all learners and to planning targeted interventions for those who struggle (e.g. Artelt & Rausch, 2014, Begenly et al., 2011 and Gresham, Mac-Millan, &

Bocian, 1997). This study contributes to the existing body of teacher judgment research with its two unique findings. First, in the aggregate, the early elementary teachers in the present study were found to be at least as accurate in judging the spelling achievement of their students as previous researchers have found similarly educated and experienced teachers to be in judging their students' reading skills. Second, the study also suggests that judgment accuracy may be related to teacher knowledge of developmental spelling. Although I found a modest significant interaction between the two variables, further research in this area is needed to replicate the finding in other samples, and to determine whether greater teacher judgment accuracy has a positive effect on student achievement.

REFERENCES

- Adams, M. J. (1990). *Beginning to read: Learning and thinking about print*. Cambridge, MA: MIT Press.
- Alatalo, T. (2016). Professional content knowledge of grades one—three teachers in Sweden for reading and writing instruction: language structures, code concepts, and spelling rules. *Scandinavian Journal of Educational Research*, 60(5), 477–499. <https://doi.org/10.1080/00313831.2015.1024734>
- Apel, K. (2011). What is orthographic knowledge. *Language Speech and Hearing Services in Schools*, 42, 592–603.
- Artelt, C., & Rausch, T. (2014). Accuracy of teacher judgments. In *teachers' professional development* (pp. 27–43). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-94-6209-536-6_3
- Ashton, P. T. (1975). Cross-cultural Piagetian research: An experimental perspective. *Harvard Educational Review*, 45(4), 475–506. <https://doi.org/10.17763/haer.45.4.kpq1578t13tk0541>
- Bates, C., & Nettlebeck, T. (2001). Primary school teachers' judgements of reading achievement. *Educational Psychology*, 21(2), 177–187. Retrieved from http://www.academia.edu/download/30556325/Book_review__Risk_and_resilience_in_childhood-_An_ecological_perspective.pdf
- Bear, D., Invernizzi, M., Templeton, S., & Johnston, F. (2012). *Words their way: Word study for phonics, vocabulary, and spelling instruction*. Upper Saddle River, NJ: Pearson Education.
- Beers, J. W., & Henderson, E. H. (1977). A study of developing orthographic concepts among first graders. *Research in the Teaching of English*, 11(2), 133–148. Retrieved from <http://www.jstor.org/stable/40170666>
- Begeny, J. C., Eckert, T. L., Montarello, S. A., & Storie, M. S. (2008). Teachers' perceptions of students' reading abilities: An examination of the relationship between teachers' judgments and students' performance across a continuum of rating methods. *School Psychology Quarterly*, 23(1), 43.
- Begeny, J. C., Krouse, H. E., Brown, K. G., & Mann, C. M. (2011). Teacher judgments of students' reading abilities across a continuum of rating methods and achievement measures. *School Psychology Review*, 40(1), 23–38.
- Bennett, R. E., Gottesman, R. L., Rock, D. A., & Cerullo, F. (1993). Influence of behavior perceptions and gender on teachers' judgments of students' academic skill. *Journal of Educational Psychology*, 85(2), 347–356. <https://doi.org/10.1037/0022-0663.85.2.347>

- Berninger, V. W. (1999). Coordinating transcription and text generation in working memory during composing: automatic and constructive processes. *Learning Disability Quarterly*, 22(2), 99–112. <https://doi.org/10.2307/1511269>
- Beswick, J., Willms, D., & Sloat, E. (2005). A comparative study of teacher ratings of emergent literacy skills. *Education*, 126(1), 116–137.
- Binks, E. (2008). *An Assessment of University Instructors and their Pre-Service Teachers' Knowledge of Basic Language Constructs Before and After University Instructor Professional Development*. Doctoral dissertation, Texas A&M University.
- Binks-Cantrell, E., Joshi, R. M., & Washburn, E. K. (2012). Validation of an instrument for assessing teacher knowledge of basic language constructs of literacy. *Annals of Dyslexia*, 62(3), 153–171. <https://doi.org/10.1007/s11881-012-0070-8>
- Block, M. K., & Duke, N. K. (2015). Letter names can cause confusion and other things to know about letter-sound relationships. *Young Children*, 70(1), 84-91. Retrieved from <http://search.proquest.com/openview/4fd74aceb31a7f4a78e404fe178476c5/1?pq-origsite=gscholar&cbl=27755>
- Bos, C, Mather, N., Narr, R., & Babur, N. (1999). Interactive, collaborative professional development in early literacy instruction: Supporting the balancing act. *Learning Disabilities Research & Practice*, 14(4), 227–238.
- Bos, Candace, Mather, N., Dickson, S., Podhajski, B., & Chard, D. (2001). Perceptions and knowledge of preservice and inservice educators about early reading instruction. *Annals of Dyslexia*, 51(1), 97–120. <https://doi.org/10.1007/s11881-001-0007-0>
- Bourassa, D. C., & Treiman, R. (2001). Spelling development and disability: The importance of linguistic factors. *Language, Speech, and Hearing Services in Schools*, 32(3), 172–181. Retrieved from <http://pubs.asha.org/article.aspx?articleid=1780259>
- Bransford, J., Brown, A., & Cocking, R. (2000). How experts differ from novices. In *How people learn: Mind, brain, experience, and school*. (pp. 31–50). Washington, DC: National Academy Press.
- Burnham, D. (2003). Language specific speech perception and the onset of reading. *Reading and Writing: An Interdisciplinary Journal*, 16, 573–609.
- Carlisle, J., Correnti, R., Phelps, G., & Zeng, J. (2009). Exploration of the contribution of teachers' knowledge about reading to their students' improvement in reading. *Reading and Writing*, 22(4), 457–486. <https://doi.org/10.1007/s11145-009-9165-y>

- Carlisle, J. F., Kelcey, B., Rowan, B., & Phelps, G. (2011). Teachers' knowledge about early reading: effects on students' gains in reading achievement. *Journal of Research on Educational Effectiveness*, 4(4), 289–321. <https://doi.org/10.1080/19345747.2010.539297>
- Carreker, S., Joshi, R. M., & Boulware-Gooden, R. (2010). Spelling-related teacher knowledge: The impact of professional development on identifying appropriate instructional activities. *Learning Disability Quarterly*, 33(3), 148–158. Retrieved from <http://ldq.sagepub.com/content/33/3/148.short>
- Chall, J. (1967). *Learning to read: The great debate*. New York, NY: McGraw-Hill.
- Chall, J. (1983). *Stages of reading development*. New York, NY: McGraw-Hill.
- Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5(2), 121–152. https://doi.org/10.1207/s15516709cog0502_2
- Coburn, C. E. (2001). Collective sensemaking about reading: How teachers mediate reading policy in their professional communities. *Educational Evaluation and Policy Analysis*, 23(2), 145–170. Retrieved from <http://epa.sagepub.com/content/23/2/145.short>
- Coladarci, T. (1992). Teachers' knowledge of what students know: The case of reading. *Reading Improvement*, 29(1), 34–39.
- Coladarci, Theodore. (1986). Accuracy of teacher judgments of student responses to standardized test items. *Journal of Educational Psychology*, 78(2), 141. Retrieved from <http://psycnet.apa.org/journals/edu/78/2/141/>
- Corno, L., & Snow, R. E. (1986). Adapting teaching to individual differences among learners. In *Handbook of Research on Teaching* 3(605-629).
- Cunningham, A., Perry, K., Stanovich, K., & Stanovich, P. (2004). Disciplinary knowledge of K-3 teachers and their knowledge calibration in the domain of early literacy. *Annals of Dyslexia*, 54(1), 139–167. Retrieved from <http://www.springerlink.com/index/RWX4K504257X2850.pdf>
- Cunningham, A., Stanovich, K., & Maul, A. (2011). Of correlations and causes: The use of multiple regression modeling in literacy research. In *Literacy Research Methodologies* (pp. 50–69). New York, NY: Guildford Press.
- Cunningham, A., Zibulsky, J., & Callahan, M. (2009). Starting small: Building preschool teacher knowledge that supports early literacy development. *Reading and Writing*, 22(4), 487–510. Retrieved from <http://www.springerlink.com/index/11w13n817j3q7227.pdf>
- Cunningham, P., & Hall, D. (2009). *Making words*. Chicago: Pearson/Allyn.

- Demaray, M. K., & Elliot, S. N. (1998). Teachers' judgments of students' academic functioning: A comparison of actual and predicted performances. *School Psychology Quarterly, 13*(1), 8–24. <https://doi.org/10.1037/h0088969>
- Doore, B. E. (2010). *Preschool Teacher Judgments of Students' Alphabetic Knowledge*. Retrieved from <http://digitalcommons.library.umaine.edu/etd/988/>
- Dottin, E. S. (2009). Professional judgment and dispositions in teacher education. *Teaching and Teacher Education, 25*(1), 83–88. <https://doi.org/10.1016/j.tate.2008.06.005>
- Doust, J. (2012). Over-diagnosis: the view from inside primary care. Retrieved September 16, 2012, from <https://theconversation.com/over-diagnosis-the-view-from-inside-primary-care-8889>
- Duggar, S. W. (2006). *The Relationship of Teacher Knowledge and First-Grade Reading Outcomes in Low-Income Schools*. Florida State. Retrieved from <http://etd.lib.fsu.edu/theses/available/etd-03312006-162445/>
- Eckert, T. L., Dunn, E. K., Coddling, R. S., Begeny, J. C., & Kleinmann, A. E. (2006). Assessment of mathematics and reading performance: An examination of the correspondence between direct assessment of student performance and teacher report. *Psychology in the Schools, 43*(3), 247–265. <https://doi.org/10.1002/pits.20147>
- Ediger, A., & McCormack, B. (2015). *Teacher language knowledge as applied to error analysis and design of instruction*. Presented at the Language Teacher Education Conference, Minneapolis, MN
- Ehri, L. (2000). Learning to read and learning to spell: Two sides of a coin. *Topics in Language Disorders, 20*, 19–36.
- Ehri, L. C. (1987). Learning to read and spell words. *Journal of Reading Behavior, 19*(1), 5–31. <https://doi.org/10.1080/10862968709547585>
- Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading, 9*(2), 167–188. https://doi.org/10.1207/s1532799xssr0902_4
- Ehri, L. C. (2014). Orthographic mapping in the acquisition of sight word reading, spelling memory, and vocabulary learning. *Scientific Studies of Reading, 18*(1), 5–21. <https://doi.org/10.1080/10888438.2013.819356>
- Ehri, L. C., & Wilce, L. S. (1985). Movement into reading: Is the first stage of printed word learning visual or phonetic? *Reading Research Quarterly, 20*(2), 163. <https://doi.org/10.2307/747753>
- Ehri, L. C., Wilce, L. S., & Taylor, B. B. (1987). Children's categorization of short vowels in words and the influence of spellings. *Merrill-Palmer Quarterly, 33*(3), 393–421.

- Ehri, L., Nunes, S., Willows, D., Schuster, B., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly*, 36(3), 250–287. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1598/RRQ.36.3.2/abstract>
- Feinberg, A. B., & Shapiro, E. S. (2003). Accuracy of teacher judgments in predicting oral reading fluency. *School Psychology Quarterly*, 18(1), 52. Retrieved from <http://psycnet.apa.org/journals/spq/18/1/52/>
- Fontenault, D., & Salter, N. (1993). *Cast-a-spell process spelling: A strategy-based spelling program*. East Hampton, CT: EDU-CARE.
- Foorman, B., & Moats, L. (2004). Conditions for sustaining research-based practices in early reading instruction. *Remedial and Special Education*, 25(1), 51–60.
- Fountas, I., & Pinnell, G. (2002a). *Phonics lessons: Letters, words, and how they work, Grade 1*. Portsmouth, NH: Heinemann.
- Fountas, I., & Pinnell, G. (2002b). *Phonics lessons: Letters, words, and how they work, Grade 2*. Portsmouth, NH: Heinemann.
- Frith, U. (1986). A developmental framework for developmental dyslexia. *Annals of Dyslexia*, 36(1), 67–81. <https://doi.org/10.1007/BF02648022>
- Fromkin, V., Rodman, R., & Hyams, N. (2018). *An introduction to language* (11th ed.). Boston, MA: Cengage Learning.
- Fry, E. (1969). The readability graph validated at primary levels. *The Reading Teacher*, 22(6), 534–538.
- Ganske, K. (1999). The developmental spelling analysis: A measure of orthographic knowledge. *Educational Assessment*, 6(1), 41–70. https://doi.org/10.1207/S15326977EA0601_4
- Ganske, K. (2013). *Word journeys: Assessment-Guided phonics, spelling, and vocabulary instruction* (2nd edition). New York: The Guilford Press.
- Gentry, R. (2000). A Retrospective on invented spelling and a look forward. *The Reading Teacher*, 54(3), 318–332.
- Gentry, R. (2016a). *Spelling Connections 1*. Columbus, OH: Zaner Bloser.
- Gentry, R. (2016b). *Spelling Connections 2*. Columbus, OH: Zaner Bloser.
- Gentry, R., & Henderson, E. (1978). Three steps to teaching beginning readers to spell. *Reading Teacher*, 31(6), 632–637.

- Gersten, R., & Dimino, J. A. (2006). RTI (Response to Intervention): Rethinking special education for students with reading difficulties (yet again). *Reading Research Quarterly*, 41(1), 99–108. <https://doi.org/10.1598/RRQ.41.1.5>
- Gibson, S. (2010). Reading Recovery® teacher expertise: Gaining and structuring content knowledge for early literacy intervention. *Literacy Teaching & Learning: An International Journal of Early Reading & Writing*, 15(1/2), 17–51.
- Gleason, J. B. (1978). The acquisition and dissolution of the English inflectional system. 109, 120. In *Parallels and divergences* (pp. 109–120). Baltimore, MD: Johns Hopkins University Press.
- Gough, P. B., & Hillinger, M. L. (1980). Learning to read: an unnatural act. *Bulletin of the Orton Society*, 30(1), 179–196. <https://doi.org/10.1007/BF02653717>
- Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, 7(1), 6–10. Retrieved from <http://rse.sagepub.com/content/7/1/6.short>
- Graham, S. (1999). Handwriting and spelling instruction for students with learning disabilities: A Review. *Learning Disability Quarterly*, 22(2), 78–98. <https://doi.org/10.2307/1511268>
- Graham, S., & Santangelo, T. (2014). Does spelling instruction make students better spellers, readers, and writers? A meta-analytic review. *Reading and Writing*, 27(9), 1703–1743. <https://doi.org/10.1007/s11145-014-9517-0>
- Gresham, F. M., MacMillan, D. L., & Bocian, K. M. (1997). Teachers as “Tests”: Differential validity of teacher judgments in identifying students at-risk for learning difficulties. *School Psychology Review*, 26(1), 47-60.
- Hamilton, C., & Shinn, M. (2003). Characteristics of word callers: An investigation of the accuracy of teachers’ judgments of reading comprehension and oral reading skills. *School Psychology Review*, 32(2), 228–240. Retrieved from <http://168.144.206.53/documents/wordcaller.pdf>
- Harris, A., & Jacobson, M. (1972). *Basic elementary reading vocabularies*. New York, NY: Macmillan.
- Henderson, E. (1981). *Learning to read and spell: The child’s knowledge of words*. Dekalb, IL: Northern Illinois University Press.
- Henderson, E. (1990). *Teaching spelling*. Chicago: Houghton Mifflin College Division.
- Henderson, E., & Templeton, S. (1986). A developmental perspective of formal spelling instruction through alphabet, pattern, and meaning. *The Elementary School Journal*, 86(3), 304–316.

- Hess, F., Rotherham, A., & Walsh, K. (2004). *A qualified teacher in every classroom? Appraising old answers and new ideas*. Cambridge, MA: Harvard Education Press.
- Hinnant, J. B., O'Brien, M., & Ghazarian, S. R. (2009). The longitudinal relations of teacher expectations to achievement in the early school years. *Journal of Educational Psychology, 101*(3), 662–670. <https://doi.org/10.1037/a0014306>
- Hoge, R. D., & Butcher, R. (1984). Analysis of teacher judgments of pupil achievement levels. *Journal of Educational Psychology, 76*(5), 777-781. Retrieved from <http://psycnet.apa.org/journals/edu/76/5/777/>
- Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research, 59*(3), 297–313. Retrieved from <http://rer.sagepub.com/content/59/3/297.short>
- Hoth, J., Döhrmann, M., Kaiser, G., Busse, A., König, J., & Blömeke, S. (2016). Diagnostic competence of primary school mathematics teachers during classroom situations. *ZDM, 48*(1–2), 41–53. <https://doi.org/10.1007/s11858-016-0759-y>
- Huey, E. (1908). *The psychology and pedagogy of reading*. New York: MacMillan.
- Hurford, D. P., Hurford, J. D., Head, K., Keiper, M., Nitcher, S., & Renner, L. (2016). The Dyslexia dilemma: A history of ignorance, complacency and resistance in colleges of education. *Journal of Childhood & Developmental Disorders, 2*(3). <https://doi.org/10.4172/2472-1786.100034>
- Invernizzi, M, Abouzeid, M., & Gill, J. T. (1994). Using students' invented spellings as a guide for spelling instruction that emphasizes word study. *The Elementary School Journal, 95*(2), 155–167.
- Invernizzi, Marcia, & Hayes, L. (2004). Developmental-spelling research: A systematic imperative. *Reading Research Quarterly, 39*(2), 216–228.
- Joshi, R., Treiman, R., Carreker, S., & Moats, L. C. (2008). How words cast their spell. *American Educator, 32*(4), 6–16. Retrieved from <https://www.aft.org/sites/default/files/periodicals/joshi.pdf>
- Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology, 80*(4), 437–447.
- Jüttner, M., & Neuhaus, B. J. (2012). Development of items for a pedagogical content knowledge test based on empirical analysis of pupils' errors. *International Journal of Science Education, 34*(7), 1125–1143. <https://doi.org/10.1080/09500693.2011.606511>

- Klug, J., Bruder, S., Kelava, A., Spiel, C., & Schmitz, B. (2013). Diagnostic competence of teachers: A process model that accounts for diagnosing learning behavior tested by means of a case scenario. *Teaching and Teacher Education, 30*, 38–46.
<https://doi.org/10.1016/j.tate.2012.10.004>
- Kohn, A. (2000). Standardized testing and its victims. *Education Week, 20*(4), 46–47.
- Kuhl, P. (2010). *The linguistic genius of babies*. Retrieved from
https://www.ted.com/talks/patricia_kuhl_the_linguistic_genius_of_babies.html
- Leinhardt, G. (1983). Novice and expert knowledge of individual student's achievement. *Educational Psychologist, 18*(3), 165–179.
- Lonigan, C. J., & Shanahan, T. (2008). Executive Summary - Developing early literacy: Report of the National Early Literacy Panel.
- Madelaine, & Wheldall. (2009). *Developments in educational psychology*. (K. Wheldall, Ed.) (2nd edition). London; New York: Routledge.
- Malatesha Joshi, R., Binks, E., Hougen, M., Dahlgren, M. E., Ocker-Dean, E., & Smith, D. L. (2009). Why elementary teachers might be inadequately prepared to teach reading. *Journal of Learning Disabilities, 42*(5), 392–402.
<https://doi.org/10.1177/0022219409338736>
- Mann, V. A., Tobin, P., & Wilson, R. (1987). Measuring phonological awareness through the invented spellings of kindergarten children. *Merrill-Palmer Quarterly, 33*(3), 365–391.
- Martin, S. (2005). *Examining the Accuracy of Teacher Judgments of Students' Early Literacy Skills* (Doctoral Dissertation). Lehigh University.
- Masterson, J., & Apel, K. (2000). Spelling assessment: Charting a path to optimal intervention. *Topics in Language Disorders, 20*(3), 50–65.
- McCutchen, D., Harry, D., Cunningham, A., Cox, S., Sidman, S., & Covill, A. (2002). Reading teachers knowledge of children's literature and English phonology. *Annals of Dyslexia, 52*, 207–228.
- McCutchen, Deborah, Abbott, R. D., Green, L. B., Beretvas, S. N., Cox, S., Potter, N. S., ... Gray, A. L. (2002). Beginning literacy: Links among teacher knowledge, teacher practice, and student learning. *Journal of Learning Disabilities, 35*(1), 69–86. Retrieved from <http://journals.sagepub.com/doi/abs/10.1177/002221940203500106>
- McNeill, B., & Kirk, C. (2014). Theoretical beliefs and instructional practices used for teaching spelling in elementary classrooms. *Reading and Writing, 27*(3), 535–554.
<https://doi.org/10.1007/s11145-013-9457-0>

- Moats, L. C. (1994). The missing foundation in teacher education: Knowledge of the structure of spoken and written language. *Annals of Dyslexia*, 44(1), 81–102. Retrieved from <http://www.springerlink.com/index/9510LX8J4572588W.pdf>
- Moats, L. C. (2009). Still wanted: Teachers with knowledge of language. *Journal of Learning Disabilities*. Retrieved from <http://psycnet.apa.org/psycinfo/2009-12677-001>
- Moats, L. C. (1999). Teaching reading is rocket science: What expert teachers of reading should know and be able to do. Retrieved from <http://eric.ed.gov/?id=ED445323>
- Moats, L. C. (2005). How spelling supports reading. *American Educator*, 29(4), 12–22. Retrieved from <http://www.perfectlyspeaking.com/uploads/Moats.pdf>
- Moats, L., & Lyon, G. (1996). Wanted: Teachers with knowledge of language. *Topics in Language Disorders*, 16(2), 73–86.
- Moats, L. C. (2009). Knowledge foundations for teaching reading and spelling. *Reading and Writing*, 22(4), 379–399. <https://doi.org/10.1007/s11145-009-9162-1>
- Nathan, R., & Stanovich, K. (1991). The causes and consequences of differences in reading fluency. *Theory Into Practice*, 30(3), 176–184.
- Paige, D. D., Smith, G. S., Rasinski, T. V., Rupley, W. H., Magpuri-Lavell, T., & Nichols, W. D. (2018). A path analytic model linking foundational skills to Grade 3 state reading achievement. *The Journal of Educational Research*, 1–11. <https://doi.org/10.1080/00220671.2018.1445609>
- Park, S., & Oliver, J. (2008). Revisiting the conceptualisation of pedagogical content knowledge (PCK): PCK as a conceptual tool to understand teachers as professionals. *Research in Science Education*, 38(3), 261–284.
- Perfetti, C. (1985). *Reading ability*. Oxford University Press.
- Perfetti, C. (1989). There are generalized abilities and one of them is reading. In *Knowing, learning, and instruction: Essays in honor of Robert Glaser*, (pp. 307–335).
- Phelps, G., & Schilling, S. (2004). Developing measures of content knowledge for teaching reading. *The Elementary School Journal*, 105(1), 31–48. Retrieved from <http://www.jstor.org/stable/10.1086/428764>
- Phillips, B. M., Clancy-Menchetti, J., & Lonigan, C. J. (2008). Successful phonological awareness instruction with preschool children: Lessons from the classroom. *Topics in Early Childhood Special Education*, 28(1), 3–17. <https://doi.org/10.1177/0271121407313813>

- Piaget, J. (1964). Part I: Cognitive development in children: Piaget development and learning. *Journal of Research in Science Teaching*, 2(3), 176–186. <https://doi.org/10.1002/tea.3660020306>
- Piasta, S. B., Connor, C. M., Fishman, B. J., & Morrison, F. J. (2009). Teachers' knowledge of literacy concepts, classroom practices, and student reading growth. *Scientific Studies of Reading*, 13(3), 224–248. <https://doi.org/10.1080/10888430902851364>
- Pinnell, G., & Fountas, I. (1998). *Word Matters: Teaching Phonics and Spelling in the Reading/Writing Classroom* (1st ed.). Westport, CT: Heinemann.
- Read, C. (1971). Pre-School children's knowledge of English phonology. *Harvard Educational Review*, 41(1), 1–34. <https://doi.org/10.17763/haer.41.1.91367v0h80051573>
- Read, C. (1975). *Children's categorization of speech sounds in English* (No. 17). National Council of Teachers of English.
- Read, C., & Treiman, R. (2013). Children's invented spelling: What we have learned in forty years*. In M. Piattelli-Palmarini & R. C. Berwick (Eds.), *Rich Languages From Poor Inputs* (pp. 197–211). Oxford University Press. Retrieved from <http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199590339.001.0001/acprof-9780199590339-chapter-13>
- Reid, D., Hresko, W., & Hammill, D. (2001). *Test of Early Reading Abilities* (3rd ed.). Austin, TX: Pro Ed.
- Rigden, D. (2006). Report on licensure alignment with the essential components of effective reading instruction. *National Council for Accreditation of Teacher Education, Reading First Teacher Education Network*.
- Ross, P., & Gibson, S. A. (2010). Exploring a conceptual framework for expert noticing during literacy instruction. *Literacy Research and Instruction*, 49(2), 175–193.
- Scales, R. Q., Wolsey, T. D., Lenski, S., Smetana, L., Yoder, K. K., Dobler, E., Grisham, D., Young, J. R. (2018). Are we preparing or training teachers? Developing professional judgment in and beyond teacher preparation programs. *Journal of Teacher Education*, 69(1), 7–21. <https://doi.org/10.1177/0022487117702584>
- Scharer, P., & Zutell, J. (2003). The development of spelling. In *Handbook of Early Childhood Literacy* (pp. 271–286). London: Sage.
- Schreiber, P., & Read, C. (1980). Children's use of phonetic cues in spelling, parsing, and—maybe—reading. *Bulletin of the Orton Society*, 30(1), 209–224.

- Shaywitz, S. E., Shaywitz, B. A., Fletcher, J. M., & Escobar, M. D (1990). Prevalence of reading disability in boys and girls. Results of the Connecticut Longitudinal Study. *JAMA: The Journal of the American Medical Association*, 264(8), 998–1002.
<https://doi.org/10.1001/jama.264.8.998>
- Shulman, L. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14. Retrieved from <http://www.jstor.org/stable/10.2307/1175860>
- Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1–21.
- Simmons, L. (Ed.). (2006). *Saxon phonics and spelling*. Saxon.
- Sitton, R. (2006a). *Sitton spelling and word skills practice book Level 1*. Cambridge, MA: Educators Publishing Service.
- Sitton, R. (2006b). *Sitton spelling and word skills practice book Level 2*. Cambridge, MA: Educators Publishing Service.
- Snow, C., Burns, M., & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
- Snow, C., Griffin, P., & Burns, M. (2005). *Knowledge to support the teaching of reading: Preparing teachers for a changing world*. Hoboken, NJ: Jossey-Bass.
- Spear-Swerling, L., & Brucker, P. O. (2004). Preparing novice teachers to develop basic reading and spelling skills in children. *Annals of Dyslexia*, 54(2), 332–364. Retrieved from <http://link.springer.com/article/10.1007/s11881-004-0016-x>
- Spear-Swerling, L., & Cheesman, E. (2012). Teachers' knowledge base for implementing response-to-intervention models in reading. *Reading and Writing*, 25(7), 1691–1723.
<https://doi.org/10.1007/s11145-011-9338-3>
- Spencer, E., Schuele, C., Guillot, K., & Lee, M. (2008). Phonemic awareness skill of speech-language pathologists and other educators. *Language, Speech, and Hearing Services in Schools*, 39(4), 512.
- Stahl, S. A., & Murray, B. A. (1994). Defining phonological awareness and its relationship to early reading. *Journal of Educational Psychology*, 86(2), 221–234.
- Stotsky, S. (2006). Why American students do not learn to read very well: The unintended consequences of Title II and teacher testing. *Third Education Group Review*, 2(2). Retrieved from <http://tegr.org/Review/Articles/vol2/v2n2.pdf>

- Stotsky, S. (2009). Licensure tests for special education teachers: How well they assess knowledge of reading instruction and mathematics. *Journal of Learning Disabilities*, 42(5), 464–474. <https://doi.org/10.1177/0022219409338740>
- Südkamp, A., Kaiser, J., & Moller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A Meta-Analysis. *Journal of Educational Psychology*, 104(3), 743–762.
- Templeton, S. (1992). New trends in an historical perspective: Old Story, New resolution—sound and meaning in spelling. *Language Arts*, 69(6), 454–463.
- Tolman, C. (2010). *LETRS Module 1, Chapter 4: Chall's stages and Ehri's phases of reading development*. Retrieved from http://sopriswest.http.internapcdn.net/sopriswest_vitalstream_com/Elearning/LETRS/LETRS_Mod1Chp4/HTML/index.html
- Treiman, R. (1993). *Beginning to spell: A study of first grade children*. Oxford University Press on Demand.
- Treiman, Rebecca. (1985). Phonemic awareness and spelling: Children's judgments do not always agree with adults'. *Journal of Experimental Child Psychology*, 39(1), 182–201. [https://doi.org/10.1016/0022-0965\(85\)90035-9](https://doi.org/10.1016/0022-0965(85)90035-9)
- Treiman, Rebecca. (2017). Learning to Spell Words: Findings, theories, and issues. *Scientific Studies of Reading*, 21(4), 265–276. <https://doi.org/10.1080/10888438.2017.1296449>
- Treiman, Rebecca, Metsala, J. L., & Ehri, L. C. (1998). Why spelling? The benefits of incorporating spelling into beginning reading instruction. *Word Recognition in Beginning Literacy*, 289–313. Retrieved from [https://pages.wustl.edu/files/pages/imce/readingandlanguagelab/Treiman%20\(1998\)%20-%20Why%20Spelling.pdf](https://pages.wustl.edu/files/pages/imce/readingandlanguagelab/Treiman%20(1998)%20-%20Why%20Spelling.pdf)
- Valdez, A. (2013). Teacher judgment of reading achievement: Cross-Sectional and longitudinal perspective. *Journal of Education & Learning*, 2(4). Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=19275250&AN=93355435&h=1uzJZAyBvAEIIVLu4%2B6iy7B%2Byf9WvdzumfPk0jjPbvDk4f7Dba3jdf%2BzLM4Xrm%2BNQs0a6yAQEsEhiYwHRy5xIg%3D%3D&crl=c>
- Vygotsky, L. (1978). Mind in society. In *Mind in Society* (pp. 79–91). Cambridge, MA: Harvard University Press.
- Walsh, K., Glaser, D., & Wilcox, D. (2006). *What education schools aren't teaching about reading and what elementary teachers aren't learning*. Washington, DC: National Council on Teacher Quality.

- Washburn, E., Joshi, R., & Binks-Cantrell, E. (2011). Teacher knowledge of basic language concepts and dyslexia. *Dyslexia, 17*(2), 165–183. <https://doi.org/10.1002/dys.426>
- Werfel, K. L., & Schuele, C. M. (2012). Segmentation and representation of consonant blends in kindergarten children's spellings. *Language Speech and Hearing Services in Schools, 43*(3), 292. [https://doi.org/10.1044/0161-1461\(2012/11-0005\)](https://doi.org/10.1044/0161-1461(2012/11-0005))
- Wilson, B. (2013). *Foundations Teacher's Manual Level 1* (2nd ed.). Oxford, MA: Wilson Language Training Corporation.
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*(2), 89–100. <https://doi.org/10.1111/j.1469-7610.1976.tb00381.x>
- Worthy, M. J., & Invernizzi, M. (1990). Spelling errors of normal and disabled students on achievement levels one through four: Instructional implications. *Annals of Dyslexia, 40*(1), 138–151. Retrieved from <http://www.springerlink.com/index/X330218Q34L1736L.pdf>
- Wuttke, E., & Seifried, J. (2013). Diagnostic competence of (prospective) teachers in vocational education. In K. Beck & O. Zlatkin-Troitschanskaia, *From Diagnostics to Learning Success* (pp. 225–240). Rotterdam: Sense Publishers.

APPENDIX A

Recruitment email

My name is April Spencer and I am a doctoral candidate in Literacy Education at the University of Maine in Orono. I am also a kindergarten teacher at Windsor Elementary school. I am writing to invite you to participate in a study on spelling that I am conducting for my dissertation. The purpose of the research is to learn more about teachers' perspectives on spelling development in first and second graders.

Compensation

As a token of appreciation for administering the spelling test and completing the interview, you will receive a \$25 Amazon gift card. Gift cards will be issued electronically within 24 hours of completing the face to face interview.

What will you be asked to do?

If you decide to participate you will be asked to do two things. First you will be asked to group administer a 12-word spelling test to your class by the end of this school year, and to mail the students test records to me in a postage paid Priority Mail envelope which I will provide.

Administration of the spelling test should take about 15 minutes. Once you have administered the test, you will send the student test records to me without reviewing or scoring them. Second, you will be asked to schedule a follow-up interview with me of 30-45 minutes. Although spelling test administration must be completed before this school year is finished, follow up interviews can be scheduled at a time and location of your convenience during the months of June and early July. Interviews will be audio recorded to ensure the accuracy of the collected information and will be transcribed for analysis. At the time of the interview I will ask you to tell me about your experience in education (e.g., years of teaching experience, degree(s))

earned, spelling programs in use in your school), and to reflect on spelling development in first and second graders.

Risks

Except for your time and inconvenience, there are no risks to you from participating in this study.

Benefits

Participation in the study will have no direct benefit for you or your students; however, this research will contribute to the knowledge base on teachers' perspectives about spelling development and in doing so, provide insight to promote use of effective approaches to spelling instruction.

Confidentiality

Your name will not be on any of the data. A code number will be used to protect your identity. A key linking your name to the data will be kept separate from the data in a locked file cabinet in my home office and destroyed by December 2018. Coded data from the interview will be kept on my password protected computer indefinitely. The digital recordings of the interview will be destroyed once I have completed transcription process. Written transcripts will be stored in a locked file cabinet in the office of the researcher until January 2020 and then shredded. Neither your name nor your school's name will be included in any reports, presentations or publications of the study's results.

Voluntary

Participation is voluntary. If you choose to take part in this study, you may withdraw at any time. During the interview, you may skip specific questions you do not wish to answer with

no reduction in compensation; however, you must complete the interview to receive the \$25 gift card.

Contact Information

If you have any questions about this study, please contact me via email at aprilis@maine.edu or by telephone at (207) 860-8630. You may also contact my faculty advisor, Dr. Janet Spector, at spector@maine.edu. If you have questions about your rights as a research participant, you may also contact Gayle Jones, Assistant to the Protection of Human Subjects Review Board, at gayle.jones@umit.maine.edu or (207) 581-1498.

If you are willing to participate, please respond to this email and I will send you further instructions about how and when to administer the spelling test. Please note that I am able to accommodate up to 40 participants in the study and volunteers will be accepted on a first-come, first-served basis. I will also need a mailing address for you so I can send you a self-address stamped envelope for returning the spelling tests to me. Submission of students' spelling tests and willingness to schedule an interview will imply consent to participate. Once you have returned the tests to me, I will contact you again by email to schedule an interview at a time and place of your convenience. Thank you for considering participation in the study.

APPENDIX B

Spelling inventory and instructions for administration

Spelling Test Protocol Before the test: Please print this page and enough copies of the answer sheets for all of your students.

Please make sure that all students have written their first names on the tests before beginning the test, (if there are duplicate names please instruct those students to use first name and last initial). During the test: Do not circulate around the room during the test. The spelling test is meant to be given “cold.” The test is not intended as a teaching or additional practice opportunity. Please do not provide any unwitting hints to spelling. For example, do not alter natural pronunciation. If it helps, imagine your students are taking the test based on a computer delivered model where an anonymous speaker is reading the words and sentences aloud. The point of the study is to find out what students in grades 1-2 can do without any cues, prompts or help. Once you are sure all papers have names on them, begin with the scripted protocol below. Say “I am going to say some words that I want to you spell. There will be 12 words. Write the words in the box beside the number I say. Some of the words may be easy to spell and some may be more difficult. I will say each word, use it in a sentence and then say the word again. If you don’t know how to spell a word, just do the best you can. If you need more time before I go on to the next word, raise your hand.”

After the test: Choose a student or other adult to collect the papers and put them together with this instruction sheet in the mailing envelope and return ASAP using the pre-paid Priority Mail envelope provided. Although you will naturally be curious to see how your students did, it is critical to the integrity of the study that you not look at the results, not intentionally revisit

these words as spelling items on another test or in another setting or discuss the results with anyone before our interview. Please copy the code beside the student answer sheet here:

By returning this page, I acknowledge that I have read the instructions for test administration, that I understand their meaning and intent, and I voluntarily and knowingly agree to comply with the guidelines.

1. hurt The old man fell and *hurt* his back.
2. frame The picture has a wooden *frame*.
3. shallow The *shallow* water was frozen.
4. chop Please *chop* the carrots into pieces.
5. trample The horses will *trample* the flowers if they walk on them.
6. clapped Everyone *clapped* at the end of the play.
7. with My brother will come *with* us.
8. paint The men were getting ready to *paint* the house.
9. drive They will *drive* to the grocery store.
10. bridge The *bridge* had to be fixed.
11. yawn When you're tired, sometimes you *yawn*.
12. hurt The bump on his head *hurt*.

APPENDIX C

Interview protocol as approved by IRB

Part I

I'd like to start by first explaining what I hope to do while we are together today. First I'd like to spend some time getting to know a little more about you and your experience in education and with the teaching of spelling in particular. Then I'm going to ask you to talk about how you think your kids did on the spelling test you gave and finally I'm going to ask you to look at some misspellings (not necessarily those of your students) and talk about what those errors tell you about what the speller understands or doesn't understand about spelling.

- So, first, would you talk to me a little bit about yourself and your background.
- How long you've been teaching?
- Where and when did you get your first degree? What was your major?
- Have you ever studied Latin?
- Are you studying for or have you earned any other degrees since the first?
- When and in what area?
- What's the title of your current position?
- Do you have any other certifications besides those required for your current position?

Next, I'm going to give you the names of some of the spelling programs I'm aware of being used in districts around the state and ask you to rate your level of familiarity with each of the programs on a scale from 0-5 where 0 represents no familiarity and 5 represents extensive familiarity. Also, tell me if you use any of these programs in your classroom or in your school,

Current position (title): _____

(Current or expired) Maine state certifications (please check all that apply)

- General elementary K-8 Middle level
 Literacy Specialist ESL
 Secondary level Special education
 Other (please specify): _____

Years of professional* experience in education

- 0-2 3-5 6-8
 9-11 12-14 15+ (please specify) _____

If you ever studied Latin in high school or college, please specify length/level of study

- No Yes (for how long) _____

Do you use a specific program to teach spelling? If yes, please describe briefly.

Whether or not you currently use the following programs / authors, please rate your familiarity with the following programs for teaching / assessing spelling on a scale from 0 to 5 where 0= never heard of / never used and 5 = very familiar; extensive experience

<i>Words their way</i> – Bear, et al.	0	1	2	3	4	5
<i>Sitton Spelling</i>	0	1	2	3	4	5
<i>Word Journeys</i> – Ganske	0	1	2	3	4	5
<i>Spelling connections: Zaner Bloser</i>	0	1	2	3	4	5
<i>Saxon Spelling</i>	0	1	2	3	4	5
<i>Word Matters</i> – Fountas & Pinnell	0	1	2	3	4	5
<i>Making Words: Patricia Cunningham</i>	0	1	2	3	4	5
Other: _____	0	1	2	3	4	5
Other: _____	0	1	2	3	4	5

APPENDIX D

List of misspellings used in error analysis task

Misspellings for Analysis Listed by Feature

Word	Feature(s) targeted	Misspelling
yes	Letter name stage – feature A: initial and final consonants	<i>ues</i>
shop	Letter name stage – feature B: initial blends and digraphs	<i>sop</i>
rip	Letter name stage – feature C: short vowels	<i>rep</i>
van	Letter name stage – feature C: short vowels	<i>ven</i>
drop	Letter name stage – feature D: affricates	<i>jrop</i>
hunt	Letter name stage – feature E: final blends and digraphs	<i>hut</i>
girl	Within word stage – feature G: r-influenced vowel patterns	<i>gril</i>
seat	Within word stage – feature H: other common long vowels; <i>also letter name spelling</i>	<i>ct</i>
trotted	Syllable juncture stage – feature K = doubling and e-drop with “ed” and “ing”	<i>chrotid</i>

APPENDIX E

Developmental Spelling Assessment

Developmental Spelling Analysis Feature Listing With Exemplars

<i>Stages</i>	<i>Feature</i>	<i>Exemplars</i>
Letter name		
stage		
A	Initial and final consonants	<u>cut</u> , <u>map</u>
B	Initial consonant blends and digraphs	<u>ship</u> , <u>glad</u>
C	Short vowels	<u>hop</u> , <u>wet</u>
D	Affricates	<u>jet</u> , <u>chin</u>
E	Final consonant blends and digraphs	<u>bump</u> , <u>fish</u>
Within-word		
pattern stage		
F	Long vowels: vowel–consonant–e (VCe)	<u>ripe</u> , <u>frame</u>
G	R-influenced vowel patterns	<u>girl</u> , <u>clear</u>
H	Other common long vowels	<u>steep</u> , <u>tight</u>
I	Complex consonant units	<u>bridge</u> , <u>scrub</u>
J	Abstract vowels: digraphs and diphthongs	<u>frown</u> , <u>shook</u>
Syllable juncture		
stage		
K	Doubling and e-drop with “ed” and “ing”	<u>clapped</u> , <u>taping</u>
L	Other syllable juncture doubling	<u>pilot</u> , <u>shallow</u>
M	Long vowel patterns in the stressed syllable	<u>compete</u> , <u>contain</u>
N	R-influenced patterns in the stressed syllable	<u>termite</u> , <u>sturdy</u>
O	Unstressed syllable vowel patterns	<u>solar</u> , <u>fountain</u>
Derivational		
constancy stage		
P	Silent and sounded consonants	<u>condemn</u> , <u>hasten</u>
Q	Consonant changes (alternations)	<u>electrician</u> , <u>disruption</u>
R	Vowel changes (alternations)	<u>inspiration</u> , <u>composition</u>
S	Latin-derived suffixes	<u>grievance</u> , <u>inedible</u>
T	Assimilated (absorbed) prefixes	<u>irresistible</u> , <u>suppress</u>

APPENDIX F

Sample of a transcribed response

to the error analysis task

***Yes* as “ues”**

- *Y* is such a uncommon letter that they use ... that I think they haven't had enough practice with knowing *Y* as a consonant, so I do think that's tricky. When you sound it out, it sounds like a *u*; /yu/ /yu/ it makes that /u/ sound they're used to.

***Shop* as “sop”**

- *Shop*, again I'm hoping at second grade they've really got those digraphs but they could they could end up writing it as “CH.” I still have a couple who mix the “SH” and “CH.”

***Rip* as “rep”**

- I would think they would all get that. I would think they would all get that one right.

***Van* as “ven”**

- It's that *A* before the *N* that makes *A* hard in this one. It doesn't make the true /a/ sound that we, that you know they get taught from pre-K and kindergarten that *A* says /a/ and we don't say /van/, so that one's definitely ... the vowel is tricky. [Looks again at *Rip*] Yeah, the *E* and the *I*, they do commonly...like with the end of *moment*

***Drop* as “jrop”**

- The *D* as a *J* is really common for when ... the *dr* combination it's hard to separate those two letters. The ending's usually fine.

Hunt as “hut”

- Again it’s that ... I can’t remember that speech word for it, but that *N* that’s really the back of your mouth when you sort of ... I call it the swallowed *N*. Hunt [repeats word to self]

Girl as “gril”

- [Laughter] Oh gril. Yeah, we have a lot of grills in my class. They know the letters: they know it from seeing it so often and just don't have that r-controlled pattern down yet. So that happens all the time, mixing “IR” with “RI.”

Seat as “ct”

- That vowel team would be really tricky. If they remember to use a vowel team they would ... they may not know “EE” or “EA.” They may not know if they should use a silent *~e*, so even if they remember that it's a long vowel, and that they needed to do something, they may not remember *what* to do. And because *C* says /see/. It says the letter *C*, I can see how they would make that that mistake, too.

trotted as “chrotid”

- All right *trotted* ... they would probably have a hard time remembering to double the *T* and I still had a couple who had a hard time with the *ED* ending would still write “ID.”

Additional comments on the set as a whole:

- You have the *TR* as “CHR”. I see that often. I didn't mention it when I did that card, but that ... like *truck* with “CHR,” I see that a lot.
- One that surprised me was “CT” for seat. I've never ... I can see why they would spell it that way, but I've never seen that, but I can totally understand why it happens.

BIOGRAPHY OF THE AUTHOR

April Spencer was born and raised in Franklin County, Maine and is a graduate of Jay High School. She attended the University of Maine at Farmington and graduated in 1989 with a Bachelor's degree in English Literature. She earned a Master of Arts in Applied Linguistics in 1991 and a Master of Education in Counseling in 1995, both at Ohio University.

She has been employed as literacy interventionist, a Japanese teacher, an elementary level classroom teacher and also as an English as a Second Language teacher in Kentucky, Ohio and in Japan. April is a candidate for the Doctor of Philosophy degree in Literacy from the University of Maine in December 2018.