

NOTE TO USERS

This reproduction is the best copy available.

UMI[®]

**The Validity Of DIBELS As An Indicator Of
Early Literacy Achievement**

Theodore Zarowny

B. Ed., University of Alberta, 1985

Thesis Submitted in Partial Fulfillment Of
The Requirements For The Degree Of
Master Of Education
in
Curriculum And Instruction

The University of Northern British Columbia

December, 2007

© Theodore Zarowny, 2007



Library and
Archives Canada

Bibliothèque et
Archives Canada

Published Heritage
Branch

Direction du
Patrimoine de l'édition

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

ISBN: 978-0-494-48832-4

Our file Notre référence

ISBN: 978-0-494-48832-4

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

NLC Abstract

This study examines the concurrent validity of the Dynamic Indicators of Early Literacy Skills (DIBELS) using Kindergarten and Grade 1 teacher-assigned year-end scores and reading Curriculum Based Measurement (WRC) as the criterion variables. The study also establishes DIBELS benchmarks, or cut-off points, for Kindergarten and Grade 1 using a Northern British Columbia sample. A correlational analysis between the criterion and predictive variables did not confirm concurrent validity of the various DIBELS measures at the Kindergarten level; this was due to a lack of variability in the criterion measure. However, significant correlation coefficients were produced between Grade 1 DIBELS measures and WRC. From the benchmarks, a set of risk factors were established. These risk factors provide an indication to educators the degree to which a student may be at risk in their early literacy development. This study supports the use of using DIBELS as part of a teacher's assessment regime.

TABLE OF CONTENTS

Abstract	ii
Table of Contents	iii
List of Tables	v
List of Figures	vi
Acknowledgements	vii
Chapter One: Introduction	1
Chapter Two: Literature Review	6
Early Literacy Skills : Phonological Awareness	6
Early Literacy Skills : Alphabetic Principle	11
Early Literacy Skills : Reading Fluency	13
Dynamic Indicators of Basic Early Literacy Skills	14
Dynamic Indicators of Basic Early Literacy Skills in School District No. 57	17
Purpose of the Study	18
Research Questions	20
Chapter Three: Method	22
Sample	22
Subject Selection	22
Instruments	23
Predictive Variables	
DIBELS	23
Criterion Variable	25
Language Arts Scores	25
Reading CBM	27
Procedure	28
Data Collection	28
Data Cleaning and Screening	29
Sample Size	29
Collapsing Language Arts Categories	31
Data Analysis	31
Chapter Four: Results	35
Concurrent Validity	35
Descriptive Statistics	35
Patterns in Convergent Validity	37
Benchmarks	40

Chapter Five: Discussion		51
Concurrent Validity		51
Benchmarks		55
Limitations		57
Implications		58
Implications for Education		58
Implications for Future Research		62
References		64
Appendix A	Sample DIBELS Measure	70
Appendix B	Sample Reading CBM	75
Appendix C	Letter of Approval From School District No.57	77
Appendix D	Graphs Used to Establish Benchmarks	79

LIST OF TABLES

Table 1: Descriptive Statistics For Kindergarten and Grade 1	36
Table 2: DIBELS Measures Correlations For Kindergarten and Grade 1	38
Table 3: Biserial Correlations for Kindergarten and Grade 1 with Language Arts Marks	40
Table 4: Benchmarks for Kindergarten and Grade 1	44
Table 5: Frequency of Scores Below the Benchmark	47
Table 6: Risk Factor for the Number of Tests Scored Below the Benchmark	48
Table 7: Pearson Chi-Square Values Between Students Meeting Expectations in Language Arts and Students Meeting Expectations on DIBELS Measures	50

LIST OF FIGURES

Figure 1: Frequency for kindergarten NWF & Language Arts final using a score category of 5.	41
Figure 2: Sample benchmark identification process using grade 1 NWF scores (category = 9)	42
Figure 3: Frequency Distribution for kindergarten ISF & Language Arts Final using a score category of 2.5	43

ACKNOWLEDGMENTS

My heartfelt thanks go to Dr. Peter MacMillan for his amazing patience, direction, and ongoing support throughout all the phases of this document. From the start, Dr. MacMillan was responsive and always made himself available for feedback, discussions and problem solving sessions.

I must acknowledge my parents who, despite life's circumstances denying them a formal education, instilled in me the value of education, perseverance, and life-long learning.

Throughout this process, my wife and best friend Crystal has been unwavering in her support and patience. She believed in my aspirations from day one, and remained the constant through to the completion of this work. It is to her I owe my greatest thanks.

CHAPTER 1: INTRODUCTION

Educators throughout School District No. 57 (Prince George) and throughout the province of British Columbia are increasingly expected to use valid data in making educational decisions, a process that has aptly become known as data-driven decision-making. Data-driven decision-making involves taking what exists and measuring it in some way, usually resulting in numerical data. Based upon that data, objectives are identified and appropriate strategies are chosen to increase, maintain, or lessen the observed behavior.

The reasons for the recent trend in data-driven decision-making in education are many. Popham (2004) identifies public accountability as one of the drivers behind data-driven decision-making. Pressure has been placed on education systems worldwide, and “accountability systems are imposed from outside because the public has doubts about that profession's quality of service.” He adds that “once issued, demands for accountability rarely disappear” (p. 1).

The need to address education's public accountability concern is evident in British Columbia. School districts across the province are expected to submit to the Ministry of Education annual Accountability Contracts. These contracts include a District Plan for Student Success that identifies how the District will address and maintain standards in several broad-based goal areas such as technology, numeracy, aboriginal education, social responsibility, and grade to grade transitions. Each goal has a series of more specific objectives. For example, an objective of the numeracy goal may be to increase student mathematical problem solving skills. Accompanying each objective is a series of strategies that identify specific actions to be taken, timelines, and resources needed to reach that objective. The expectation, of course, is that the goals and objectives are based upon measurable objectives – objectives for which data can be

collected in order to show the current status, and to monitor and observe the effectiveness of the selected intervention strategies.

Provincial Foundation Skills Assessment (FSA) for grades 4, 7 and 10¹ (Ministry of Education, 2005a), Student Satisfaction Surveys (Ministry of Education, 2005c), recently-introduced Grade 10 provincial examinations (Ministry of Education, 2005b) and Parent Satisfaction Surveys (Ministry of Education 2005c) are further examples of the province's reliance on data for accountability purposes. A district's Plan for Student Success must reflect data collected by the district as well as the goals and objectives in each school's School Plan for Student Success.

Each school in British Columbia is required to submit an SPSS. Selected goals, objectives, and strategies within the SPSS are also data driven. Schools use a variety of measures or performance indicators. Some of these are Ministry-based such as the Foundation Skills Assessment (Ministry of Education, 2005a) and Satisfaction Surveys (Ministry of Education 2005c). Others are school-based assessments such as reading running records, performance standard rubrics, standardized tests, and teacher-made rubrics.

Another reason for the increased use of data driven decision making is related to the first reason. A reliance on data-driven decision-making is increasingly necessary for economic reasons. Financial resources for education are stressed and, as Fullan, Bertani and Quinn (2004) state, "last year's success makes possible next year's new money" (p. 45). School jurisdictions and governments use high stakes assessments as evidence that the financial investment into education is paying off. In their School Plans for Student Success, schools make resource priorities and allocations based on goals that are decided upon available data. Accordingly,

¹ FSA scores are not calculated into a student's final course grade. In 2005, Grade 10 FSAs were replaced with provincial final examinations in Math and English and constitute a percentage of a student's final course grade.

School District No. 57 makes similar decisions based on the data collected by schools. Special Education funding, a scarce resource, also relies upon reliable and valid data for the identification of students with special needs. Funding to supplement special education services in schools require assessments that are able to identify at risk learners and monitor their progress.

Besides public and financial accountability, educational researchers identify another reason, and perhaps the most important reason for the increasing use and reliance on data-driven decision-making based on meaningful assessment – to enhance student learning. Popham (2004) makes the case that carefully selected assessment tools that provide accurate data are essential for reducing the achievement gap – the discrepancy between skills of high-performing and low-performing students. He states, “anyone who is working to reduce achievement gaps must become assessment literate - at least with respect to the qualities of achievement tests that will or won't reveal genuine differences between what upper-income and lower-income students learn” (p. 2). Similarly, Fullan, *et al* (2004) point out that education systems need a “collective moral purpose [that] makes explicit the goal of raising the bar and closing the gap for all individuals and schools” (p. 43).

Further, after an extensive review of over 250 studies and journal articles, Black and Wiliam (1998a) found that formative assessment practices in the classroom produced substantial learning gains, especially for low-achieving students and students with learning disabilities. Black and Wiliam (1998b) use “the general term *assessment* to refer to all those activities undertaken by teachers – and by their students in assessing themselves – that provide information to be used as feedback to modify teaching and learning activities. Such assessment becomes *formative assessment* when the evidence is actually used to adapt the teaching to meet student needs” (p. 2). Indeed, the work of such researchers as Black and Wiliam, along with Clarke,

Owens and Sutton (2006) are contributing to the growing “assessment for learning” reform occurring in various school jurisdictions across Canada.

Such attempts to use data are evident in School District No. 57 (Prince George). Kindergarten and Grade 1 educators in School District No. 57 use the University of Oregon’s Dynamic Indicators of Basic Early Literacy Skills (Dynamic Indicators of Basic Early Literacy Skills, 2000-2004) as a key tool for assessing early literacy skills. Each measure included in the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) is comprised of probes, short one-minute to two-minute tests that assess one skill. The probes are used in the fall, winter and spring and provide educators with data in five indicators of literacy: letter naming fluency, initial sound fluency, phonemic segmentation fluency, nonsense word fluency, and oral reading fluency. Students who score considerably lower than do their peers are considered at risk, and intervention strategies are administered by teachers to increase student skills, thereby reducing the gap.

As Kaminsky and Good (1998) point out, effective early literacy intervention has encountered two difficulties. One difficulty is a question of what literacy skills should be taught. The second difficulty is that while the effects of certain instructional strategies improve group outcomes, some individual students often remain deficient and may require individualized intervention strategies. Kaminsky and Good suggest that although many educators rely upon a “teach-and-hope” approach whereby early literacy instruction is based on beliefs about reading acquisition and instructional strategies, what is really required are effective assessment strategies that identify specific areas of concern for individual students and direct the use of specific instructional strategies which address the deficits identified by the assessment. Kaminsky and Good argue that the use of DIBELS helps fill those specific needs.

There is an ever-increasing reliance and demand upon assessment data. Public accountability, economic accountability, and an increased understanding of the link between assessment and student achievement, all create need for valid assessment techniques. The Dynamic Indicators of Basic Early Literacy Skills (DIBELS) is one assessment that school jurisdictions across North America are using to meet that demand.

CHAPTER 2: LITERATURE REVIEW

Early Literacy Skills: Phonological Awareness

Three questions should be considered when deciding upon effective assessment practices, whether it be assessment for learning or assessment of learning. First, what should be assessed? One would want to be certain that the data that are obtained accurately reflect what one wants to measure. In the case of early reading, therefore, the question becomes, “What are the aspects of early literacy that are meaningful indicators of reading ability.” A second consideration is whether or not the skill or concept being assessed is teachable. Again, in the case of early literacy, if one assesses a student on a certain aspect of reading, can that aspect be affected through instruction? And third, what tool should be used to make that assessment? Is the tool reliable and valid under the conditions in which it is used?

In deciding what should be assessed in early literacy, one key ability is phonological awareness. However, the term phonological awareness is not entirely clear. Stanovich (1994) defines “phonological awareness” as the ability to deal explicitly and segmentally with sound units smaller than the syllable, and includes a spectrum of possible activities. For example, Snow Burns, and Griffen (1998) distinguish phonemic segmentation as one of the more precise phonological abilities in that spectrum. Similarly, Adams (1990) identifies five levels of what are identified as “phonemic awareness” activities, and Chard and Dickson (1999) provide a similar spectrum that ranges from less complex initial rhyming and rhyming songs to the most complex activity - phonemic segmentation. Goswami and Bryant (1992) also make the case that phonological awareness involves a range of activities.

Stanovich (1991), Mann (1993), Kaminsky and Good (1996), and Liberman, et al. (1989) make the case that phonemic awareness, or the lack thereof in early readers is a moderate

to strong predictor of future reading ability. Liberman et al. (1989) refer to several studies done in the 1970s and 1980s that demonstrate not only a correlation between phonological awareness and reading ability, but a causal relationship as well. A causal relationship would mean that improving the phonological abilities of readers would increase their reading abilities.

Phonological processing weaknesses are characteristic of a broad spectrum of students who have difficulty acquiring early word recognition skills. The failure to acquire adequate early word recognition skills has dire consequences. Students who are not correctly assessed and remediated lag further behind in the development of critical early word reading skills. These students

... will receive less practice in reading than will other children, they will miss opportunities to develop reading comprehension strategies, and they will acquire negative attitudes about reading itself. Furthermore, if children do not acquire good word reading skills early in elementary school, they will be cut off from the rich knowledge sources available in print, and this may be particularly unfortunate for children who are already weak in general verbal knowledge and ability. (Torgesen, 2000, p. 4)

Liberman and Shankweiler (1987) surveyed the research and make the case that being a skilled reader, either as an adult or child, is related to one's phonological understanding of words. Weak readers, they argue, have difficulty identifying the phonemic structure of words, and this weakness in phonological awareness has other implications. For instance, weaknesses in phonological awareness and processing leads to difficulties in developing word recognition strategies. Another weakness demonstrated by readers weak in phonemic awareness is the ability to name things; poor readers often have difficulty accessing the correct words in their lexicon and this causes them to misname objects - a problem the researchers say results from not being able to access the phonological properties of the words (Liberman & Shankweiler, 1987; Fowler, 2004).

Liberman and Shankweiler (1987) make another connection between phonemic awareness and reading. They conclude that short term memory is essential for reading comprehension. Because the nature of short term memory allows its capacity to be reached quickly, the faster the rate that information can be put into it, the more that can be remembered. Weak readers have comprehension difficulties as the rate at which they can put words into short term memory is decreased by their slow word recognition. Liberman and Shankweiler (1987) conclude that “reading comprehension difficulties may reflect processing limitations originating in the phonology, and not necessarily absence or malformation of the higher level structures of the sentence grammar” (p. 219).

Recent studies (e.g. Gray & McCutcheon, 2006), however, question the degree to which phonology plays a role in reading comprehension. They confirm that phonological processing indeed plays an important role in word recognition, but while phonological awareness plays a role in comprehension, uncertainty remains as to whether a different kind of phonological processing is involved in comprehension, or whether the contribution of phonology is of a different degree.

Of all the phonological activities, phonemic awareness appears to be the most difficult one to develop. Phonemic awareness involves manipulation the individual sounds, or phonemes, that comprise a word. Liberman and Shankweiler (1987) explain that a word is formed by a certain phonological structure, or its combination of individual phonemes. They provide the example of the syllable /ba/. The consonant /b/ has a “buh” sound, and the vowel /a/ has the “ah” sound. Yet we do not produce the sound “buh-ah” when we pronounce that syllable. Instead, because of the coarticulation of speech sounds, we produce the sound /ba/. Understanding the

phonological structure of a word means being able to break a word down to its individual phonemes - this is phonemic awareness.

While phonemic awareness has a primary role in early literacy, Torgesen (2000) argues that phonemic awareness is a particularly difficult reading skill to master, but should, nonetheless, be taught as “difficulties in learning to ‘sound out’ words, or to use phonetic cues to help decipher them, limit the ability of children to read independently and accurately throughout first grade, and these difficulties usually extend into late elementary school and adulthood” (p. 3). Studies such as Uhry (2002) suggest that even reading strategies such as finger-point reading require phonemic awareness. Kaminsky and Good (1998) add that “by ensuring that all children have adequate phonological awareness skills when they enter first grade, we may be able to mitigate the effects of many differences in the initial skills required for successful reading instruction and prevent many cases of reading failure” (p. 115). Perfetti (1999) further argues that the relationship between phonemic awareness and reading is a reciprocal one “in which phonemic ability is first promoted through literacy acquisition and which then enables further gains in literacy” (p. 52).

While the research shows that phonological awareness and, more specifically phonemic awareness, is indeed an important early literacy skill, another consideration is whether phonemic awareness is something that can be taught, or whether it is a product of biology and genetics and therefore difficult to affect. This is an important consideration as the implications for education would differ greatly if phonemic awareness was not affected by instruction. Hence the question - despite its integral role in early literacy, can phonemic awareness be taught?

To begin answering this question, one might assume that the ability to segment words into phonemes might be related to one’s analytic ability which implies a deficit in the cognitive

domain if phonemic awareness abilities were weak. However, as Liberman and Shankweiler (1987) point out, weak readers are just as competent as strong readers with non-linguistic analysis tasks, but are weak with phonemic segmentation, suggesting that the ability to segment words into phonemes is not a deficit in the analytical domain, but in the linguistic domain.

Moreover, illiterate adults and adults who had reading instruction but remain poor readers are unable to perform well on tasks that require phonemic awareness (Liberman et al., 1989). Such information suggests that phonological awareness is not something that develops spontaneously, but rather a skill that requires direct instruction.

The literature also tells us that, despite its important function in early literacy, young children have difficulty identifying phonemic segments (Adams, 1990; Chard & Dickson, 1999, Liberman & Shankweiler, 1987; Perfetti 1999; Treiman, 1992). For instance, Treiman notes that early readers have greater success with larger sound constructs such as intrasyllabic units. *Intrasyllabic units* are word components that are larger than a single phoneme, but smaller than a syllable. An example of an intrasyllabic unit from the word “children” would be /en/ which is composed of two phonemes (/e/ and /n/) but smaller than the final syllable /ren/.

Perfetti (1999) agrees that phonemes have an “invisibility” to them and that there is “little outside of literacy contexts that can serve to draw the existence of phonemes” (p. 51). He maintains, however, that phonemic awareness in the pre-reading stage and at the Grade 1 level is essential for further gains in literacy, and that “good literacy instruction makes phonemes more visible...” (p. 52). However, will instruction increase a reader’s phonemic processing abilities?

Much evidence exists that phonological awareness can be taught. Liberman and Shankweiler (1987), Liberman et al. (1989) and Torgesen (2000) cite several studies that show training in phonological awareness helps early learners to read. Byrne (1991) argues that

phonological awareness needs to be taught – that it does not happen by learning to read holistically. Based on this discussion therefore, phonological awareness and, more specifically phonemic awareness, is a skill that, when assessed, can provide important information to teachers as they provide reading instruction to their students.

Early Literacy Skills: Alphabetic Principle

Having phonemic awareness skills alone are not enough in order to learn how to read. In other words phonemic awareness is a necessary but not sufficient aspect of a skilled reader (Phillips & Torgesen, 2006). Learning to read involves using phonological awareness to match the phonological structure of words to their alphabetical transcriptions (Byrne, 1992; Gough, 1992; Liberman et al., 1989; Perfetti, 1999). In other words, learning to read alphabetic writing systems includes the awareness that the individual letters that make up words merely represent sounds of spoken language. Understanding that written words are made up of individual letters and applying their letter-sound knowledge to these letters is necessary for children to analyze and to reliably and accurately read. This is called alphabetic understanding (Adams, 1990).

The alphabetic principle has two parts (Big Ideas In Beginning Reading, 2004). First, alphabetic understanding is the knowledge that words are composed of letters that represent sounds or phonemes. The second part of the alphabetic principle is phonological recoding. This skill involves understanding the systematic relationships between letters and phonemes (letter-sound correspondence) to pronounce or spell an unknown printed string.

Phonological recoding consists of two tasks, the first of which is regular word reading (beginning decoding). Regular word reading requires that the word be read from left to right, that sounds for all letters can be generated, and sounds can be blended into recognizable words. A

second phonological recoding task is reading irregular words. Irregular words are those like “laugh” and “their” which do not conform to regular decoding rules. Irregular words are more difficult to read because the sounds of the letters are unique to that word or a few words, or the student has not yet learned the letter-sound correspondences in the word (Carnine, Silbert & Kame'enui, 1997). The third phonological recoding skill is advanced word analysis.

The system that maps print to speech - the alphabetic principle- requires a strength in phonological awareness, and Liberman et al. (1989) state that “although both reading and speech require some degree of mastery of language, reading requires, in addition, a mastery of the alphabetic principle” (p. 5).

According to Moats (1999) children who lack understanding of the alphabetic principle face many difficulties. Such children have difficulties understanding that words are composed of letters, cannot associate a letter of the alphabet with its corresponding phoneme or sound, will have difficulty identifying a word based on a sequence of letter-sound correspondences (e.g., that “ba” is made up of two letter-sound correspondences /b/ /a/). Other difficulties will be experienced when blending letter-sound correspondences to identify decodable words, using knowledge of letter-sound correspondences to identify words in which letters represent their most common sound, identifying and manipulating letter-sound correspondences within words, and reading pseudowords (e.g., “nug”) with reasonable speed.

Mastering the alphabetic principle along with phonological processing abilities is central to learning to read and should be a primary focus in early reading instruction. Perfetti (1999) argues that reading education should have a general goal of “making sure they learn the alphabetic principle, something that requires some attention to fostering students’ phonemic awareness” (p. 54).

As with phonological awareness, whether or not instruction in alphabetic awareness is possible and has an effect on reading ability is an important consideration. Studies such as Foorman, Fletcher, Schatschneider and Mehta (1998) demonstrate that when given direct instruction in the alphabetic principle, student word recognition skills significantly improve.

Early Literacy Skills : Reading Fluency

A third area of the literature argues that another key indicator of early literacy is reading fluency. Collins, Good, Knutson, Shinn and Tilly (1992) define reading fluency as “the speed and accuracy with which a student reads words” (p. 460). More recently, Pikulski and Chard (2005) refer to a more recent definition:

Reading fluency refers to efficient, effective word recognition skills that permit a reader to construct the meaning of text. Fluency is manifested in accurate, rapid, expressive oral reading and is applied during, and makes possible, silent reading comprehension (p. 510).

Research shows that reading fluency measures are effective indicators of reading achievement (Deno, 1985; Good & Jefferson, 1998; Shinn, 1989). Criterion related validity strength of oral reading CBMs in relation to standardized tests and teacher judgments has been identified by Shinn et al. (1992). Collins, et al. (1992) also conclude that “oral reading fluency fits current theoretical models of reading well and can be validated as a measure of general reading achievement, including comprehension” (p. 476).

Phonological awareness, knowledge of the alphabetic principle, and reading fluency, therefore, are three aspects of early literacy that play an important role in reading ability. They are measurable, and can be affected by appropriate interventions. There is a recognition that effective assessment tools are required for appropriate intervention, and that current assessment

techniques are necessary (Ysseldyke & Christensen, 1988). D'Angiulli and Siegel (2003) make a similar argument when they conclude that measures such as the WISC-R do not necessarily identify students with reading disabilities, and that measures other than IQ are needed to accurately identify students with learning disabilities.

Good, Simmons, and Smith (1998) identify the procedures needed for an effective assessment tool. The assessment must identify students who are deficit in early literacy skills, provide ongoing feedback, evaluate the efficacy of interventions, identify accurately students with serious learning problems, and evaluate the overall effectiveness of intervention strategies.

Dynamic Indicators of Basic Early Literacy Skills

The Dynamic Indicators of Basic Early Literacy Skills (DIBELS) is a set of brief standardized measures designed for monitoring progress and for early identification of children with reading difficulties (DIBELS, 2004). Designed by the University of Oregon to be short fluency measures, DIBELS provide assessment information on a set of early literacy skills that are identified in the literature as being directly related to later reading success – phonological awareness, the alphabetic principle and reading fluency. The Phoneme Segmentation Fluency (PSF) measures provide assessment data for phonological awareness while Initial Sounds Fluency (ISF), Letter Naming Fluency (LNF) and Nonsense Word Fluency (NWF) provide indicators of a child's ability with the alphabetic principle. Finally, the Oral Reading Fluency (ORF) measure is the indicator of reading fluency.

Recent research shows that DIBELS measures are a valid indicator of a child's progress towards the acquisition of early literacy skills. For instance, Kaminsky and Good (1996) found evidence of criterion-related validity in the areas of phonological awareness, vocabulary

development, and letter-naming fluency. Results indicate a significant correlation between Letter Naming Fluency (LNF) and Phonemic Segmentation Fluency (PSF), and the criterion measures which included the Stanford Diagnostic Reading Test and The McCarthy Scales of Children's Abilities. Coefficients ranged between .58 and .90 for LNF and between .63 and .73 for PSF. Weaker but still significant relationships were identified between the criterion variable and Picture Naming Fluency, another DIBELS measure.

Since the Kaminsky and Good (1996) study, replications have been conducted using the various DIBELS tasks and a variety of criterion measures with similar results. For instance, Barger (2003) found a strong relationship between the DIBELS Oral Reading Fluency (ORF) measure and the North Carolina End of Grade Test. In his study, Barger took the median score of three ORF readings of 38 Grade 1 students from North Carolina. A strong correlation of .73 was found to exist between the North Carolina End of Grade Test and DIBELS ORF. Similar results were found by Buck and Torgesen (2003) when 1,102 third grade students were administered the ORF measure and the results then correlated with the Florida Comprehensive Assessment Test - Sunshine State Standards. A coefficient of .70 was obtained between the reading comprehension component and the ORF scores. Other studies (Shaw & Shaw, 2002; Vandre Meer, Lentz, and Stollar, 2005; Wilson, 2005) provide similar results.

Strong correlation between both standardized achievement levels to DIBELS and teacher ratings of achievement to DIBELS were found also found by Elliot, Lee, and Tollefson (2001). A modified battery of DIBELS measures was used and included fluency measures (LNF and Sound Naming Fluency) and ability measures (Initial Phoneme Ability, and Phoneme Segmentation Ability). The ability measures relate to the DIBELS subtests of Initial Sound Fluency and Phoneme Segmentation Fluency. They were renamed with minor alterations for the purpose of

their particular study. Using 75 Kindergarten students from one district in the mid-western United States, the researchers found high correlations between DIBELS and four criterion measures: the Woodcock-Johnson PsychEducational Achievement Battery - Revised Broad Reading and Skills, the Test of Phonological Awareness, The Developing Skills Checklist, the Kaufman Brief Intelligence Test, and a Teacher Rating Questionnaire (TRQ). The TRQ was a five point teacher rating scale with which teachers rated student reading at the end of the year. The scale ranged from “well below average” to “well above average.” Concurrent validity between the DIBELS measures and the achievement measures were found as correlations ranged from .68 to .75.

In the same study, Elliot et al. (2001) also confirmed the predictive ability of DIBELS using a hierarchical regression analysis. In one analysis significant standardized beta weights were reported for the Woodcock-Johnson Broad Reading and Skill and for the Teacher Rating Questionnaire. Their results indicate significant correlations ranging from .56 to .70.

Hinze, Ryan, and Stoner (2003) also conclude that DIBELS has strong concurrent validity. The researchers used a standardized test, The Comprehensive Test of Phonological Processing (CTOPP,) as the criterion measure. In their study 86 Kindergarten students from Massachusetts were given three of the DIBELS measures: LNF, ISF and PSF; these measures were correlated to CTOPP. The CTOPP is also composed of various measures, and while some correlations between the subtests of both measures were low (.08 for instance), the correlations between DIBELS and the composite scores were significant, with 8 of the 9 possible correlations ranging from .20 to .60.

The literature also supports the reliability of DIBELS. Kaminsky and Good (1996) established significant alternate form reliability coefficients. The coefficient for Phonemic

Segmentation Fluency was .88 and .93 for Letter Naming Fluency. These results were mirrored by Elliot et al. (2001) who used interrater reliability, test-retest, and equivalent forms to establish reliability estimates. The reliability coefficients for Letter Naming Fluency, Sound Naming Fluency, Initial Phoneme Ability, and Phonemic Segmentation Ability ranged from .64 to .94 while the values for the composite scores ranged between .89 and .91.

Despite being able to establish significant relationships between DIBELS and other measures, Kaminsky and Good (1998) recognize the inherent difficulties in assessing the early literacy performance of young readers. While only a year separates a Kindergarten student and a Grade 1 student, rapid changes and growth adds to the variability in performance over that year. Kaminsky and Good contend that DIBELS measures are not as predictive of future performance as a CBM may be with older students, and that “more data points may be needed to obtain the same amount of confidence in a performance estimates” than with older students (p. 121).

Dynamic Indicators of Basic Early Literacy Skills in School District No. 57

As jurisdictions were required to develop District Plans for Student Success, School District 57 began searching for an early literacy assessment. DIBELS was seen as a viable choice. School District No. 57 (Prince George) adopted the use of DIBELS as its early literacy assessment tool to serve two primary functions. First, DIBELS provides Kindergarten and Grade 1 teachers in School District No. 57 assessments to assist in the identification of students who have early literacy skill deficits. Such knowledge guides teachers as they choose appropriate instructional and intervention strategies to correct deficits that are identified by the assessment.

Besides acting as a feedback mechanism for educators, DIBELS would help track early literacy trends in the schools and in the district. From 2003 to 2004, School District No. 57

developed a data matrix – a table and timeline of the types of data each school in the district is expected to collect in order to facilitate school-based decision-making. Once collected, student scores from the variety of measures are entered into a district-wide data bank.

As educators use assessment data such as DIBELS in developing intervention plans, information is required about how the target student compares to other students in the local population (Kaminsky & Good, 1998). Therefore, in the winter of 2002, local district DIBELS norms were established for School District #57. All District analyses were performed under the supervision of Dr. Peter MacMillan, a professor at the University of Northern British Columbia. By the winter of 2005, all elementary schools were expected to use DIBELS as their primary early literacy assessment.

Each January, Kindergarten and Grade 1 DIBELS scores are collected by the elementary schools in the District. This data is used by District staff in the development of the District Plan For Student Success, by School Planning Councils for their respective School Plan for Student Success, and by classroom and learning assistance teachers as they plan their instruction.

The creation of goals, the making of recommendations and the implementation of interventions and support programs from the District to the classroom level is, therefore, partially based partially on DIBELS data.

Purpose of the Study

The DIBELS research field is lacking research on two levels: research with a Canadian sample, and research using teacher-assigned grades as the criterion variable. Most of the literature uses a standardized measure to determine the validity of the DIBELS measures. In the early primary grades, year-end assessments are not produced by standardized tests. Instead, teacher assigned grades based on a rich assortment of records, anecdotal notes, and assessments

contribute to the assigning of a year end grade for students. How DIBELS relates to teacher-assigned grades is an area that lacks research.

Similarly, studies using a Canadian sample, and more specifically a northern British Columbia sample, are also lacking. Socioeconomic, cultural, and ethnic factors create a unique population and diversity. The *District Data Summary 2001/02 - 2005/06: 057 Prince George*. Ministry of Education (2007b) provides some insight into the demographics of School District No. 57 (Prince George). On many levels, the population in the District is economically disadvantaged when compared to the province. Education attainment levels, for instance, indicated that in 2001, 10% of the District population had a University degree while 17% held a University degree at the provincial level. Unemployment rates in the District were higher than the provincial average as was the percent of individuals on income assistance while the percent of single parent families exceeded the provincial average. The District is becoming increasingly culturally diverse. Almost 2% of households speak Punjabi or Chinese as the primary language in the home, and 7% of the students are identified as English as a Second Language students. Students of aboriginal ancestry comprise 20% of the student population in the District.

DIBELS has been developed using an American sample. Using a different sample would broaden the base of validity research and provide some data useful for School District No. 57 .

Besides contributing to DIBELS validation research, this thesis will also produce DIBELS benchmarks for School District No. 57. A benchmark is a reference or measurement point that can be used for comparison purposes. In the case of DIBELS, a series of benchmarks for each of the measures has been created by the University of Oregon Center for Teaching and Learning (2004). Along with the benchmark is a risk indicator. For instance, a score below the

benchmark of 4 for Kindergarten ISF at the beginning of the year indicates the child is at risk with this aspect of reading. A score above the benchmark of 8 indicates low-risk.

Currently, School District No. 57 is using the University of Oregon's benchmarks for identifying DIBELS cut-off points in the identification of at-risk readers. The District is interested in providing locally-based benchmarks (S. Fewster, personal communication, February, 2004). The University of Oregon benchmarks are based on a population and demographic different from the one in School District No. 57. Establishing benchmarks using a local sample can be beneficial in the identification and intervention processes for young at risk readers in the District.

Research Questions

One of the key measures that is mandated to be collected by School District No. 57 in its data matrix is DIBELS scores. Because of the reliance on DIBELS data on decision making at so many levels, educators want to be sure that the instrument is valid. Educators and parents want to be confident that reading deficits are correctly identified and appropriate intervention strategies for correcting those deficits are chosen at such formative stages of learning and language development. Educators, policy makers, and the public want to be confident that scarce educational resources are being directed towards real needs as they identify school improvement goals. In order to address these considerations, this study investigates the following question: What is the concurrent validity of DIBELS and what is strength of the relationship between DIBELS scores and other reading measures used by Kindergarten and Grade 1 teachers?

Besides answering the above question, this study will also establish DIBELS benchmarks using a Northern British Columbia sample. Benchmarks for Kindergarten and Grade 1 will help

teachers in their assessment of the degree to which a student may be at risk at reading using local samples.

CHAPTER 3: METHOD

Sample

Subject Selection

The population for this study is School District No. 57 students who were enrolled in an elementary school in the 2003-2004 school year. The schools range in size from 26 students to over 500 students. While most of the students are registered in Prince George city schools, the district does include several smaller rural schools.

The subjects of this study were students who were registered in Kindergarten and Grade 1 in the 2003-2004 school year. Not all students registered in these grades were subjects in this study. In January, 2004, all teachers of Kindergarten and Grade 1 across the district were required to administer an early literacy assessment to students. Kindergarten teachers had a choice of using either DIBELS as the main performance indicator of early literacy skills or a University of British Columbia early literacy screen. Only scores from those schools who used DIBELS were used in the study. The use of an early literacy screen was optional for Grade 1 teachers. Those who did administer one used DIBELS.

Moreover, while the district recommended that Grade 1 reading CBMs be administered, the use of these measures was a school-based decision. Therefore, the sample size was reduced by those schools not administering Grade 1 CBMs.

All Kindergarten and Grade 1 students were assigned a performance level mark for Language Arts (LA) on the year-end reports. This study used only those schools who could report on all the variables in each of the years.

Other exclusions included special needs students who met the criteria in the *Guidebook for the Use of Curriculum Based Measurement in School District #57* (Prince George School

District No. 57, 1996). Consequently, students who were visually impaired, hearing impaired, multiply disabled, mentally disabled, or had English as a Second Language were not analyzed in this study.

Instruments

Predictive Variables

DIBELS

The DIBELS measures assess the following areas of early literacy: phonological awareness, alphabetic principle, and fluency with connected text (DIBELS, 2004). Phonological awareness, the ability to hear and manipulate sounds in words, is measured by two tests. First, Initial Sounds Fluency (ISF) considers a student's ability to identify and produce the initial sound of a given word. The examiner presents and orally names four pictures (see Appendix A for samples of all the Grade 1 DIBELS measures). The child is then asked to point to the picture that begins with the sound produced by the examiner. In the second part of the ISF phonological awareness assessment, the examiner asks the child to produce the beginning sound for a word that is presented orally and matches one of the pictures. The length of time it takes the child to produce the initial sound for the twelve given words is recorded, and a calculation is made to convert the time into an Initial Sound Fluency score.

The second phonological awareness test is the Phoneme Segmentation Fluency (PSF) test. The PSF measure allows the examiner to assess a student's ability to fluently segment words into their individual phonemes. A phoneme is the spoken sound that makes up a word. For instance, the word "big" has three phonemes: "/b/ /i/ /g/". In the PSF test, the examiner orally presents words comprised of three to four phonemes, and the child is required to orally produce

the individual phonemes for each word. After a student response, more words are given. The correct number of phonemes identified in one minute is recorded.

Alphabetic principle includes alphabetic understanding, the notion that words are composed of letters that represents sounds, and includes phonological recoding which involves saying an unknown word by using the relationship between letters and their corresponding phonemes. DIBELS uses a Letter Naming Fluency (LNF) test which measures the number of letter names a child can correctly identify in one minute. The test consists of several rows of randomly arranged upper and lower cased letters. The student is given one minute to correctly identify as many letters as possible. Another measure of alphabetic principle is taken with the Nonsense Word Fluency (NWF). This test assesses students' knowledge of letter-sound correspondences as well their ability to blend letters together to form unfamiliar "nonsense" words (e.g., "tig") The child is given a random list of words with varying forms of vowel-consonant words (e.g., "ib") and consonant-vowel-consonant words (e.g., "neg"). The child can respond by saying the nonsense word as a whole, or by verbally saying each letter sound such as /n/e/g/. The final score is the number of correct letter sounds identified in one minute. Higher scores are given when the child can phonologically state the word instead of stating the letter sounds in isolation.

Fluency through connected text, the third area of early literacy, is measured through a test called the Oral Reading Fluency (ORF) test. With this measure, the child reads a selection of text at a designated level. The number of words the student reads correctly in one minute becomes the ORF score.

The various DIBELS measures are used at different times of the year: fall, winter, and spring. In this study, the spring measures will be used because such testing is typically done in

May and, therefore, close enough to mid June when teachers begin writing their final reports. The only exception is the Kindergarten ISF assessment. This study uses the winter assessment since there is no spring ISF test. The three Kindergarten DIBELS measures, therefore, are Initial Sound Fluency, Letter Naming Fluency, and Phoneme Segmentation Fluency. The three Grade 1 measures are Phoneme Segmentation Fluency, Nonsense Word Fluency, and Oral Reading Fluency.

Criterion Variable

School Based Performance Levels

Primary school teachers provide formal reports on student progress three times throughout the year. Two of those reports are given to parents during the school year, and the final June report reflects student progress and achievement to the end of the grade. The reports must comment on the “student’s school progress with reference to the expected development for students in a similar age range” (Ministry of Education, 2004).

During the year, teachers assign a numerical value to a student’s reading status on progress report cards. This mark corresponds to the comments necessary for reporting to parents the student’s development in relation to students of a similar range age. At the time the data for this study were collected, teachers could report student progress by using one of the four scores and accompanying comment for each learner outcome identified on the report card:

- 1 not yet meeting widely held expectations
- 2 minimally meeting widely held expectations
- 3 fully meeting widely held expectations
- 4 exceeding expectations

On each report card, separate performance levels are assigned for reading and writing. For instance, writing may receive a performance level of 2, while reading may receive a performance level of 3. At the end of the year, however, reading and writing are combined to create a single Language Arts performance level which is reported to the Ministry and identified in the student's Permanent Record Card as a verbal comment, not as a number. The Language Arts scores used in this study represent three possible categories or comments that were available to teachers for reporting student achievement on the Permanent Record Card: not yet meeting expectations, meeting expectations, or exceeding expectations.

Although the final Language Arts mark used in this study is a reflection of the combined reading, writing, listening, and speaking levels, informal discussions with primary teachers in the Prince George School District revealed that reading was identified as the determining skill when providing a final Language Arts performance level for students for the Permanent Record Card.

A study of concurrent validity requires that the criterion variable be accepted as valid. Teacher assigned grades that are reported to the Ministry via the Permanent Record Card are considered a valid measure in the British Columbia public school system. From Kindergarten to Grade 11 the Ministry uses teacher-assigned grades for students' Permanent Record Card, and even in cases where Provincial Final Exams occur, fifty percent of students' final grades are based on teacher-assigned grades. While teacher-assigned final grades are not based on standardized tests at the elementary level, these grades are still considered valid. First, the Ministry of Education collects and accepts teacher assigned grades, not standardized test scores, as valid indicators of student achievement; the Ministry does not require any justification for the teacher-assigned scores. Secondly, teacher grades become valid when considering the implications associated with them, especially at the kindergarten and Grade 1 level. Identifying a

student as “not yet meeting expectations” when reporting to parents often initiates a series of interventions including conferences, more in-depth assessments, school-based interventions, and possibly referrals to other agencies.

Other educators rely upon the assigned grades of previous teachers as well when making educational decisions. When developing classroom composition, for instance, or when anticipating learning assistance demands, school planners will use teacher reporting to guide their decisions. Similarly, when a student transfers to another school, the teacher assigned grades provide important information that assists the receiving school in deciding how to best accommodate that student’s needs.

Reading CBM Measures

According to Jenkins, Deno, and Mirkin (1979), a curriculum based measure (CBM) is described as an assessment that is directly related to a student’s curricula, is short in duration, is capable of having multiple forms, is low cost, and can indicate progress over time. Many CBM measures exist and are used to assess variety of skills associated with mathematics, writing, and reading. The CBM of interest in this study is the reading CBM. A reading CBM is a one minute oral reading of a passage from a classroom text that measures reading fluency. The number of words read correctly in a minute is used as the measure. It identifies the number of words read correctly. To avoid confusion over the type of CBM being discussed, and because the reading CBM measures words read correctly, this study uses the acronym WRC when referring to the reading CBM.

Although a rather simple assessment, the WRC measure has proven to be a useful and valuable assessment tool. Since the initial CBM validity study (Deno, Mirkin & Marston, 1982), an abundance of validity studies have been conducted (see Marston, 1989) and have shown that

the Reading CBM is a valid indicator of reading ability. Using such criterion measures as the Stanford Diagnostic Reading Test, and the Woodcock Reading Mastery Test along with a numerous other reading assessments, the researchers found correlation coefficients ranging from .73 to .91. Locally, School District No. 57 (Prince George) has recognized the validity of WRC assessments (Fewster, 2000). This assessment tool has been used to provide a “consistent standard for decision-making across schools” (p. 2).

Probes that measure WRC and are used by School District No. 57 were originally normed in 1995 and renormed again in 2003. The readings that received minor modifications in 2003 are comprised of short passages taken from reading anthologies used in the schools. Five levels of reading fluency are identified: well below average, below average, average, above average, and well above average. The cut score for these levels are simply the percentile scores corresponding to P₁₀, P₂₅, P₅₀, P₇₅ and P₉₀. Samples of the reading probes can be found in Appendix B.

Procedure

Data Collection

In order to facilitate data management, School District No. 57 provided schools with a data management tool called the Performance Standard Files (Prince George School District No. 57, 2000) that enables electronic storage and retrieval of school performance indicators. Included on these data bases are WRC scores, and DIBELS scores. June report card marks are stored on a different system called Turbo School (Wong, 1985). This data management system stores a variety of ministry required student information including core subject performance levels.

The administration of WRC and DIBELS tests varied between schools. In some cases, trained Learning Assistance / Support Teachers administered the tests, while in other cases, classroom teachers trained in the administration of DIBELS were involved.

In order to gather the necessary data from the 2003-2004 school year for this study, the School District was provided with a brief proposal outlining the purpose and method of the study. The school district granted approval, (see Appendix C) and the school principals were informed of the study by Ms. Bonnie Chappel, the Director of School Services. Principals from each participating school arranged for the necessary data to be forwarded. In most cases the data was forwarded electronically, but in some cases, hard copies of the data were forwarded and the data was changed into electronic form. A step-by-step procedure for data retrieval and instructions on forwarding the data were provided. Because some schools did not use DIBELS as their early literacy assessment in Kindergarten in 2004, the data from only those schools which used DIBELS are included in this study.

Data Cleaning and Screening

Sample Size

The data for the study were gathered, but required sorting and matching to the students' Personal Education Numbers (PENs) as the scores that were imported from the Performance Standard Files (Prince George School District No.57, 2000) could not be exported with the corresponding PENs. Even though PENs are a required piece of data that must be entered into that particular data base, a technical incompatibility made impossible an electronic transfer of that piece of data. Therefore, the students' names, PENs and matching final report marks were imported from each participating school's TURBO data base on a separate file. Some arrived as hard copies while others were electronic. Using student names and PENs, the scores from the

various measures were matched to allow for an analysis using the statistical program SPSS² Windows version 13.0. Student names were deleted once the matching was completed.

The sorting eliminated many cases, especially at the Grade 1 level. The Kindergarten sample size was large, approximately 500 for each measure, due mainly to a District requirement that schools collect DIBELS scores at the Kindergarten level for its early literacy monitoring purposes.

The Grade 1 sample size, in contrast, was smaller and numbers varied greatly between the different DIBELS measures. The sample size of Grade 1 students was potentially 991 based upon the available final Language Arts report marks, but were reduced (see Table 1).

Several factors combined to reduce the sample size. First, only those students who had at least one score on any of the assessments were included, thus reducing the samples size of students with Language Arts (LA) scores to 673. Further, not all schools used DIBELS and WRC measures at the Grade 1 level, and in one case a school had the data in a form that could not be used in this study. The largest sample for the Grade 1 criterion measures was the WRC assessment with 712 scores, while the number for the final LA scores was fewer. This discrepancy may be explained by students moving out of the district before the end of the school year resulting in having a WRC score but no LA final score.

Another factor that reduced the sample size was the manner in which the some students and their scores were identified. The Performance Standards File used by School District No. 57 schools requires both a student name and a PEN number before data can be saved. In numerous cases “fake” PENs were entered beside students. For instance, PENs such as “11111” or “3333” were used instead of the actual PEN. Fake PENs are used in instances where testing on a student is done before a PEN is available. For example, a student can often arrive to a school before his

² Statistical Program for the Social Sciences not to be confused with the School Plan for Student Success.

or her Permanent Record arrives from the previous school. Consequently, many LA scores could not be used because they could not be matched to the fake PEN numbers that accompanied the DIBELS and WRC scores. The final n for Grade 1 is found in Table 1.

Collapsing Language Arts Categories

Besides the sorting and matching task, the Language Arts scores were coded to assist with the analysis. The first task involved turning LA marks from verbal comments of either “not yet meeting expectations” or “meeting/exceeding expectations” into numbers. A system that reflects performance levels on the report cards was used. A “1” was assigned to the “not yet meeting expectations” category, and a “2” was assigned to the “meeting/exceeding expectations.”

However, three schools, differentiated the second category and identified students as either “meeting expectations” or “exceeding expectations.” Because this differentiation occurred in only three of the schools, students who were given “exceeding expectations” final grades were also assigned a “2”.

A final task in the data preparation involved assigning numerical rankings to the students on Individual Education Plans (IEP). Keeping in line with Ministry regulations, Language Arts grades are not given to students on IEPs. Those on IEP’s were given a “1” for the purpose of the study since students are provided with IEP’s in designated subject areas because they are unable to meet the intended learning outcomes for the grade, thus not yet meeting expectations.

Data Analysis

Descriptive statistics were first calculated for each of the variables at the Grade 1 and Kindergarten levels. Means, standard deviations, median, and skewness was calculated for the

dependent variable (LA final marks) and the covariates which included all DIBELS measures and WRC scores.

A measure of concurrent validity of DIBELS was created by calculating Pearson's product-moment correlation coefficients between the 2003 Kindergarten LA final marks and spring DIBELS subtest scores. A second test was done using 2003 Grade 1 LA final marks and spring reading WRC and DIBELS measures.

A second analysis was completed to address the use of a two-point scale in this study. Using a two-point system for Language Arts Scores created an artificial dichotomy. The two point scale that was used for Language Arts is not a true dichotomy (e.g., being either an elementary or high school student). Rather, the dichotomy created by making the Language Arts mark a two-point system is what Glass and Hopkins (1996) call an "artifact of a crude measurement" (p. 135). In other words, underlying the "meeting/not yet meeting" dichotomy is a normal distribution that has artificially been altered to create a dichotomous distribution. Using other measurement techniques, it might be possible to create a more normal distribution of final Language Arts marks.

Therefore, in order to get an approximation of the product-moment correlation between the hypothetically normally distributed Language Arts marks and the independent variables, a biserial correlation between the variables was calculated using the following formula (Glass & Hopkins, 1996, p. 137):

$$r_{\text{bis}} = r_{\text{pb}} \sqrt{\frac{n_1 n_0 (n_s - 1)}{u^2 n_s^3}}$$

Setting benchmarks for the Kindergarten and Grade 1 DIBELS involved using a contrasting groups method similar to that as identified by Nedelsky (in Crocker & Algina, 1986). This method requires that two groups with differing proficiency levels be identified and tested. The score distribution of each group is plotted on the same graph. The intersection point of the two curves becomes the cut-off point or benchmark. As will be seen in the next chapter, the high levels of skewness of most data in this study resulted in an alteration of the procedure.

The benchmarks resulted in two risk factor calculations. A *risk-factor* is a percentage that identifies the likelihood of an event occurring. While some students scored below the benchmarks on a particular DIBELS or WRC measures and were identified as not meeting Language Arts expectations, others were identified as meeting expectations. In other words, there were cases where a student scored below the benchmarks on a measure, but received a passing grade in Language Arts.

Therefore, a risk factor was generated. The first risk factor is the percentage of all students who scored below the benchmark on a test and who were also identified as not yet meeting expectations in Language Arts. For example, a benchmark for a measure with a risk factor of 25% indicates that a student who scores below the benchmark has a one in four chance of receiving a “not yet meeting expectations” in Language Arts. These results are identified in Table 5.

The creation of benchmarks in this study led to another analysis and a subsequent, second set of risk factors. As with the DIBELS and WRC assessments, some students who scored below the benchmarks on all the measures were still identified as meeting expectations. Upon closer examination of the data, it became evident that all the students who were identified as not yet meeting expectations scored below the benchmark on at least two of the measures. Yet, other

students who scored below the benchmarks on at least two of the assessments were identified as meeting expectations in Language Arts. Therefore, a risk factor was generated by calculating the percentage of all students who scored below the benchmark on a certain number of tests and also identified as not yet meeting expectations. These risk factors can be found in Table 6.

A chi-square test was then used to analyze the relationship between the student Language Arts scores and their scores on the DIBELS measure. Student DIBELS scores were transformed into nominal numbers representing scores that were above the School District's benchmark and below the benchmark. These scores were then correlated with the Language Arts scores. The resulting chi-square values are found in Table 7.

CHAPTER 4: RESULTS

Concurrent Validity

Descriptive Statistics

The descriptive statistics for both Kindergarten and Grade 1 revealed similar observations – the Language Arts marks for both groups are highly negatively skewed. As indicated in Table 1, for the Kindergarten Language Arts (LA) mark distribution, skewness was -3.66 while for the Grade 1 Language Arts mark the skewness was -3.06. A high negative skewness indicates an abnormally large number of cases distributed to the right of the mean. The high negative skew shows, therefore, that a large proportion of the Language Arts marks report students as meeting or exceeding expectations in Grade 1, and even more so in Kindergarten. The mean score ≥ 1.94 (out of 2) and a small SD ($< .27$) also show that the vast majority of students scored a 2, or “meeting expectations” in Language Arts.

On the other hand, the WRC and DIBELS measures were less skewed in their distributions. The most highly skewed distribution in the Kindergarten distributions were the Phoneme Segmentation Fluency (PSF) scores with a skew of 1.70 and Initial Sound Fluency (ISF) scores with a skew of 1.11. Not only was the skewness significantly less than the LA scores, but the skewness was positive rather than negative. The positive skewness of these measures indicates that while some students performed well on these measures, the scores tended towards the lower ranges. Such a result is opposite to the LA scores. Similar results were produced with the Grade 1 data. The scores with the highest skew were the reading fluency scores – Words Read Correctly (WRC) and Oral Reading Fluency (ORF). Again, as with the Kindergarten data, the skewness was not as significant as were the LA results, and the skewness was opposite to the LA results.

In Table 1, the frequency (*f*) refers to the percent of students “meeting expectations” in the case of the LA marks while the frequency for the DIBELS measures is based on the University of Oregon benchmarks. The frequency of students meeting the LA expectations was high at the Kindergarten level (94%) and at the Grade 1 level (92%). The disproportionate distribution of the LA scores resulted in the skew mentioned earlier in this chapter.

Table 1

Descriptive Statistics for Kindergarten and Grade 1

	<i>Test</i>	<i>n</i>	<i>f</i>	<i>Mean</i>	<i>Median</i>	<i>SD</i>	<i>Skew</i>
Kindergarten	LA	503	94%	1.94	2.00	0.24	-3.66
	PSF	451	55%	15.45	13.00	15.11	1.70
	NWF	462	63%	14.61	10.00	10.74	0.82
	ISF	513	60%	12.38	8.00	12.61	1.11
	LNF	463	55%	22.44	18.00	16.99	0.51
Grade 1	LA	673	92%	1.92	2.00	0.27	-3.06
	PSF	202	95%	36.96	39.00	17.93	-0.09
	NWF	168	83%	55.10	51.00	28.19	0.66
	ORF	178	71%	37.98	32.00	27.05	1.54
	WRC	712	85%	46.16	37.00	32.28	0.90

In the case of the DIBELS measures, the frequency of students scoring above the benchmarks is lower in Kindergarten than in Grade 1. At the Kindergarten level, between 55% and 65% of the students met or exceeded the benchmarks. By Grade 1, between 70% to 95% of the students scored at or above the benchmark on the different measures. At the Kindergarten level, the PSF had the smallest frequency of success as 55% of the scores were at or above the

benchmark. In contrast, the measure with the highest frequency of success at 95% was PSF at the Grade 1 Level.

Patterns in Convergent Validity

The correlation coefficients in Table 2 indicate no significant relationship between the DIBELS measures and final Kindergarten LA grades. The correlation coefficient values between the predictive and criterion variables suggest a trivial effect size between the DIBELS measures and the final Language Arts scores for Kindergarten. Effect size is a judgment of the strength of relationship between two variables. The magnitude of the effect size of Pearson's r is said to be trivial if $|r| < .1$, small if $|r| \geq .1$, medium if $|r| \geq .3$, and large if $|r| \geq .5$ (Cohen, 1988). All effect sizes between Kindergarten LA scores and the DIBELS measures were less than .1, and none of the correlation coefficient values were statistically significant at the $p < .05$ level or the $p < .01$.

The highest correlation coefficient for Kindergarten was $-.09$ between LA and ISF scores. However, not only is the highest correlation coefficient considered trivial using Cohen's effect size guidelines, it is negative, suggesting that a high ISF score has a trivial relationship with a low LA mark. The highest positive correlation at 0.02 was between the LA final mark and Letter Naming Fluency (LNF), but remains statistically insignificant. Such results are what might be expected with such high degree of negative skewness in the predictive variable and more normal, but still positive skewness distributions amongst the criterion variables. The n ranged from 513 to 451 for the DIBELS measures, and $n = 503$ for Language Arts.

Table 2

DIBELS Measures Correlations for Kindergarten and Grade 1

	Test	PSF	NWF	ISF	LNF
Kindergarten	LA	-.03	.01	-.09	.02
	PSF		.53**	-.02	.47**
	NWF			.11*	.38**
	ISF				.03
		PSF	NWF	ORF	WRC
Grade 1	LA	.27*	.31*	.31*	.30*
	PSF		.36*	.33*	.31*
	NWF			.79*	.72*
	ORF				.87

* Correlation is significant at $p < .05$ level (2-tailed).

** Correlation is significant at $p < .01$ level (2-tailed).

Some significant correlations were found internally, however, among the DIBELS measures. The strongest correlations exist between Nonsense Word Fluency (NWF) and the other DIBELS measures. They ranged from $r = .53$ (NWF and PSF) to $r = .11$ (NWF and ISF). The weakest correlations were between ISF and the other DIBELS measures. Interestingly, PSF produced two of the highest r values (.53 and .47) as well as the lowest value (-.024 with ISF). Moreover, the results show no significant relationship between PSF and ISF ($r = -.024$) despite both being indicators of phonemic awareness. That the ISF scores represented winter scores while the other DIBELS measures represented spring scores may have had an impact on these results.

The results for Grade 1 found in Table 2 produced higher correlations between the criterion and predictive variables. Despite the skewed LA results, a medium effect size was obtained between the criterion and predictive variables. A fairly consistent relationship between the LA and the criterion variables appears to exist as the r values between the measures were quite similar. The strongest relationship was between Language Arts and DIBELS with the NWF measure both ($r = .31$), while the weakest was with the PSF measure ($r = .27, p < .05$).

At the Grade 1 level, the relationship between the DIBELS measures and the WRC measure was high. As might be expected because of the similarity of the tests, a high correlation was found to exist between ORF and WRC scores ($r = .87$), and the correlation between these two measures and NWF were almost identical. The n was 673 for Language Arts and 712 for CBM while n ranged from 161 to 202 for the DIBELS measures.

In contrast to the Kindergarten results, the PSF measures produced the lowest r values between it and the other measures as the coefficients ranged from .31 to .36. However, similar to the Kindergarten results, the DIBELS measure that produced the highest correlations involved the NWF measures (.76, .36, and .79).

A biserial correlation was calculated because of the dichotomy created by reporting the LA scores on a two-point (meeting expectations - not yet meeting expectations) scale. Calculating a biserial correlation increased the magnitude of the r value for both Grade 1 and Kindergarten correlations (see Table 3). A biserial correlation analysis approximates a correlation by taking into account that the LA scores would typically resemble a normal distribution.

The biserial correlations suggest a significant relationship of a medium magnitude between the Grade 1 LA scores and the independent variables. Correlation coefficients ranged

from .48 to .56. On the other hand, while strengthened, the correlation between Kindergarten LA final marks and the independent variables remained non-significant and trivial. Biserial correlations for the Kindergarten results ranged from -.06 to .03.

Table 3

Biserial Correlations for Kindergarten and Grade 1 with Language Arts Marks

<i>Test</i>	<i>Kindergarten</i>	<i>Grade 1</i>
LNF	.03	-
ISF	-.02	-
PSF	-.06	.48
NWF	.02	.56
ORF	-	.56
WRC	-	.54

Benchmarks

Establishing benchmarks for Kindergarten and Grade 1 DIBELS was challenging. Following the procedures identified by Nedelsky (in Crocker & Algina, 1986), each of the frequency distributions of predictive variables was plotted. One frequency line represents students who were identified as not yet meeting expectations in LA. These students received a score of “1” for their final LA score. The second line represents students who received a “2” in LA, or those who were meeting expectations. The Nedelsky procedure would typically have two symmetrical curves, and the cut off point would be the intersection at which one curve begins rising and the other curve begins declining. Such normal curves were not produced with the data in this study. Because of the skewed nature of most of the distributions, establishing an

intersection of the two curves was problematic. Figure 1 exemplifies the nature of the difficulty. Note that the intersection of the curve occurs at the highest score on each curve.

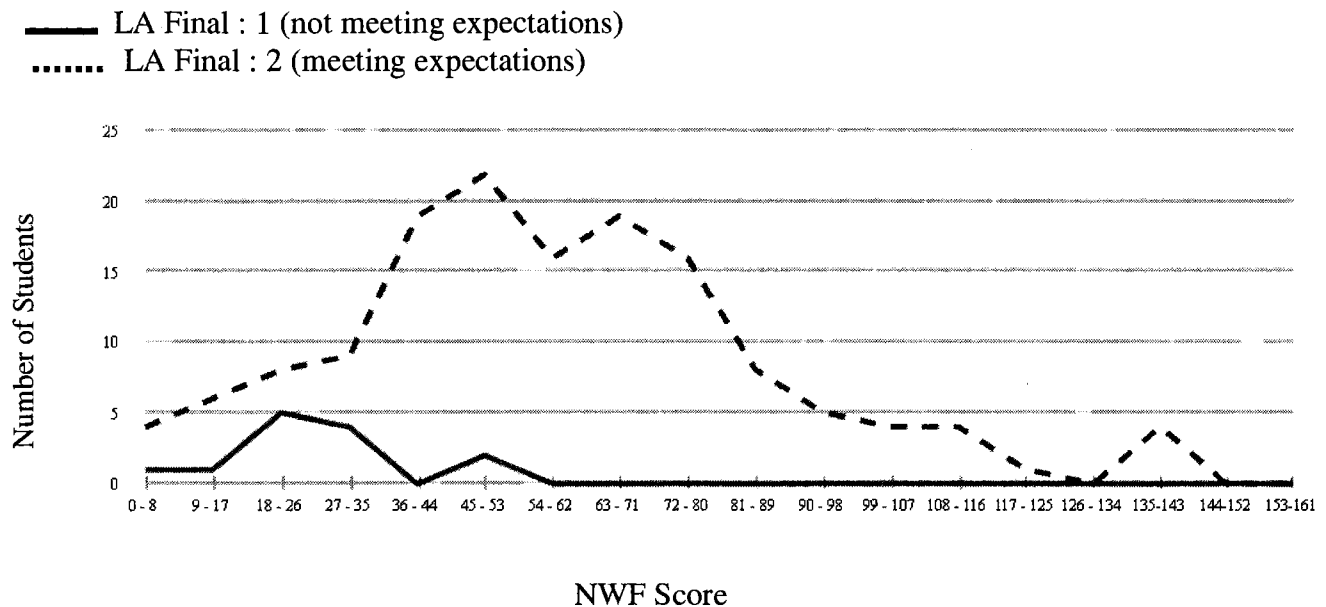


Figure 1

Frequency for Kindergarten NWF & Language Arts final using a score category of 9.

The skewness of the distribution and the resulting difficulty in finding an appropriate intersection point meant redefining the notion of a benchmark or cut-off score. Instead of being a line that establishes a definitive score at which a cutoff can be made to distinguish between *at risk* and *not at risk* readers, this study identifies a benchmark score that indicates students who have a certain degree risk of not meeting expectations. In other words, this study establishes a score or cut-off for which students who receive this score or less have a certain degree of risk at being identified as not yet meeting expectations in reading.

In order to identify such score, instead of choosing a specific cut-off point at the intersection of the two curves, a point was chosen that approximates an intersection point on the two curves (or a point as close as possible) where the distribution curve of students receiving a “1” (not yet meeting expectations) for an LA final mark began a noticeable decline and the

distribution curve of students receiving a “2” (meeting expectations) for an LA final mark began noticeable rise. A perpendicular line was drawn through these points and the score corresponding to that line became the benchmark. Figure 2 provides an example of the process. In this particular case, the cut-off score for Grade 1 NWF is 35.

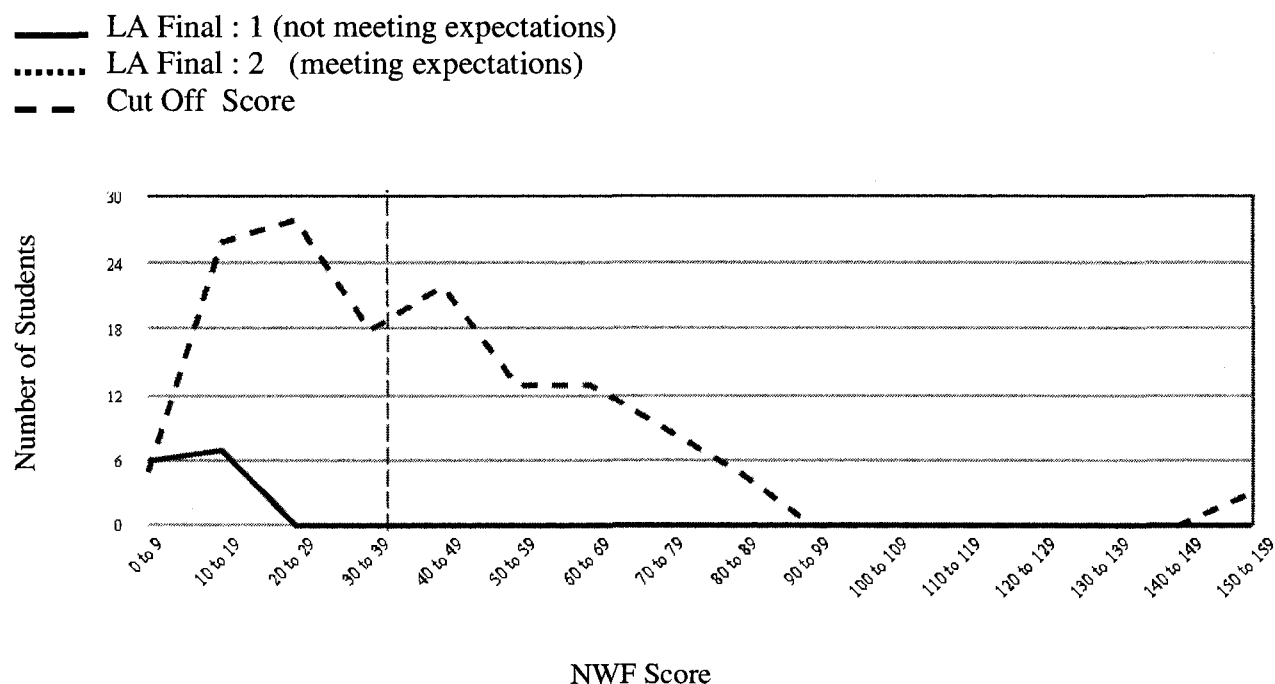


Figure 2

Sample benchmark identification process using grade 1 NWF using a score category of 10.

In order to smooth the curves so that those points could be identified as precisely as possible on the graphs, a variety of graphs were created using two different score categories. The first series of graphs had a score range of five (e.g., 0-4), and the second series of graphs had a score range of nine (e.g., 0-8).

Despite arranging the frequencies into various categories, the skewed nature of the data still made finding a benchmark problematic as the rise in the distribution of students who scored a 2 on their LA final did not necessarily correspond to the decline in the distribution of the students who received a 1 on their Language Arts score. Eventually, a series of graphs with a

score of 9 was used. The benchmarks are found in Table 4 and the graphs used for establishing benchmarks are found in Appendix D.

Identifying benchmark or cut-off points for Kindergarten was even more problematic. To establish the benchmark using a graph, a score category of 5 was used, (except in the ISF analysis) to provide greater detail on the shape of the distribution. Even so, trying to establish a cut-off for this test proved difficult.

Because the ISF scores have one decimal place in the scores, a smaller score category was established for this assessment. Doing so made the beginning of a decline in the number of students who received low ISF scores more apparent. Figure 3 the identifies this dip and the cut-off score of 7.4

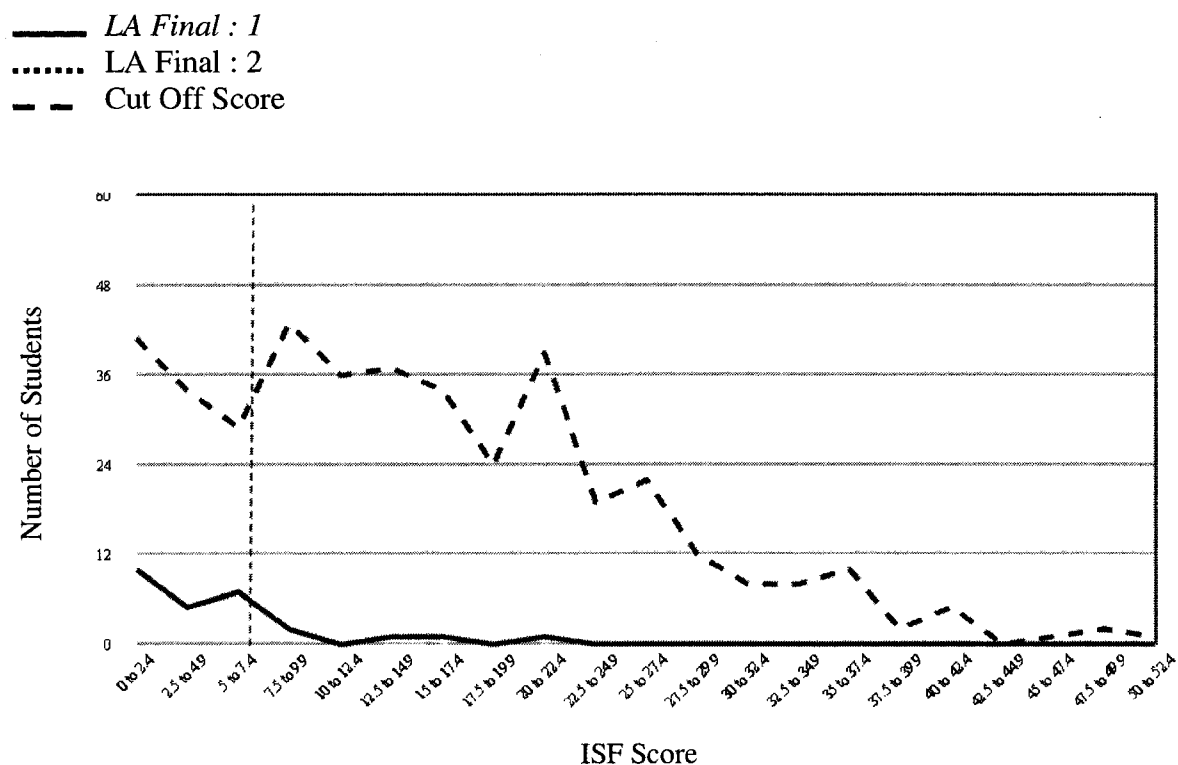


Figure 3

Frequency Distribution for Kindergarten ISF & Language Arts Final using a score category of 2.5

The benchmarks for Kindergarten DIBELS scores are also identified in Table 4. The graphs used to determine the benchmarks are found in Appendix D.

The benchmarks in this study can be used to determine the degree to which students are at risk for being identified as “not yet meeting expectations” for their grade level if they receive a score below the benchmark in any of the DIBELS or WRC measure.

Table 4

Benchmarks for Kindergarten and Grade 1

<i>Test</i>	<i>SD #57 Benchmark ("at risk")</i>	<i>University of Oregon Benchmarks ("at risk")</i>	<i>% of Students Scoring Below SD #57 Benchmark</i>	<i>Risk Factor</i>
Kindergarten				
ISF*	7.4	10	29%	17%
NWF	9.0	15	39%	10%
LNF	19.0	29	23%	22%
PSF	9.0	10	28%	12%
Grade 1				
NWF	35.0	30	25%	23%
ORF	19.0	20	29%	30%
PSF	29.0	10	26%	22%

* Middle of Year Scores

In Table 4, two benchmarks are identified. The first column of numbers are those established by this study and identified as School District No. 57 (Prince George) benchmarks. The second set, for comparison purposes, are those established by the University of Oregon

(2004). The School District No. 57 benchmarks at the Kindergarten level are lower than the University of Oregon benchmarks, except for the PSF measure which is fairly similar to that set by the University of Oregon. Lower benchmarks mean that students in School District No. 57 require a lower score on the DIBELS measures before they are considered at risk. In other words, fewer students would be considered at risk using the School District No. 57 benchmarks instead of using the University of Oregon benchmarks.

The reverse is true for the Grade 1 benchmarks. In this case, the School District No. 57 benchmarks are higher except for the ORF measure which is fairly similar to the University of Oregon benchmark. With higher benchmarks, students have to score higher on the DIBELS test before being considered at risk. This means that more students would be identified at risk using the School District No. 57 benchmarks rather than the University of Oregon benchmarks.

The final column in Table 4 is the “risk factor” and represents the percent of students who scored below the School District No. 57 benchmark and were identified as “not yet meeting expectations” at the end of the year. The Kindergarten measure with the highest risk factor was Letter Naming Fluency (LNF) at 22%. Such a risk factor means one in five kindergarten students who scored below the LNF benchmark were identified as “not yet meeting” expectations in Language Arts. In other words, a student who scores below the LNF benchmark has a one in five chance of being identified as “not yet meeting expectations” by the end of the year. The Kindergarten measure with the lowest risk factor is NWF with risk factor of 10%.

Scoring below the benchmark at the Grade 1 level carries a higher risk of being identified as not yet meeting expectations than at the Kindergarten level. The risk factor for scoring below the ORF benchmark is 30%, while scoring below the benchmark for NWF and PSF is just above 20%.

The data in Table 4 also indicate a large percentage of the sample scoring below the benchmark for a specific test yet still being identified as meeting expectations in LA. For instance, 28% of all Kindergarten students scored below the PSF benchmark, and 29% of Grade 1 students scored below the ORF benchmark. Yet Table 1 reveals that only 8% of Grade 1 students and 6% of the Kindergarten students received a 1 as a final LA mark, thus creating a significant discrepancy between students identified as “meeting expectations” in LA and their scores on the various DIBELS measures.

In order to investigate this discrepancy further, another analysis of the data was conducted in order to determine the number of measures in which students scored below the benchmark and their corresponding final LA mark. The results are shown in Table 5.

The analysis reveals that at the Kindergarten level, all students who were identified as not yet meeting expectations scored below benchmark levels on at least two or more DIBELS and WRC measures. No students with one or no scores below the benchmark were identified as “not yet meeting expectations” in Language Arts. However, 29% of the students who were identified as meeting expectations had DIBELS scores below the benchmark on two or more measures.

Table 5

*Frequency of Scores Below the Benchmark**

<i>Group</i>		<i>Test Scores Below the Benchmark</i>					<i>Total</i>
		<i>4</i>	<i>3</i>	<i>2</i>	<i>1</i>	<i>0</i>	
<i>Kindergarten Students Identified as Not Yet Meeting Expectations</i>	<i>%</i>	39	50	11	0	0	100
	<i>n</i>	7	9	2	0	0	18
<i>Kindergarten Students Identified as Meeting Expectations</i>	<i>%</i>	3	9	17	26	45	100
	<i>n</i>	9	27	51	79	136	302
<i>Grade 1 Students Identified as Not Yet Meeting Expectations</i>	<i>%</i>	100	0	0	0	0	100
	<i>n</i>	10	0	0	0	0	10
<i>Grade 1 Students Identified as Meeting Expectations</i>	<i>%</i>	20	9	16	25	30	100
	<i>n</i>	18	8	15	23	28	92

* Only those students who wrote all measure were included in this analysis.

At the Grade 1 level, only students who scored below the benchmark on all four measures received a 1 (Not Yet Meeting Expectations) for LA at the end of the year. However, 20% of the students identified as “Meeting Expectations” also scored below the benchmark on all four DIBELS measures.

Because of these conflicting results, a risk factor table was also completed for the number of measures with scores below the benchmark. Table 6 indicates that the risk of being identified as “Not Yet Meeting Expectations” in Language Arts increases with the number of scores below the benchmark.

Table 6

Risk Factor for the Number of Tests Scored Below the Benchmark

<i>Grade</i>	<i>Number of Scores Below the Benchmark</i>				
	<i>4</i>	<i>3</i>	<i>2</i>	<i>1</i>	<i>0</i>
<i>Kindergarten</i>	44	25	4	0	0
<i>Grade 1</i>	36	0	0	0	0

Table 6 shows that in Kindergarten no student who scored below the benchmark on one measure was identified as Not Yet Meeting Expectations, but 25% of the students who scored below the benchmark on three of the measures and 44% of the students who scored below the benchmark on all four of the measures were identified as Not Yet Meeting Expectations in LA. In other words, a Kindergarten student who scores below the benchmark on three DIBELS measures has a one in four chance of being identified as not yet meeting expectation in LA. A Kindergarten student who scores below the benchmark on all four measures has close to a one in two chance of being identified as Not Yet Meeting Expectations in LA.

At the Grade 1 level, students who receive a score below the benchmark on all four measures have a one in three chance of receiving Not Yet Meeting Expectations in LA. Stated alternately, students at the Grade 1 level have no risk of being identified as Not Yet Meeting Expectations in Language Arts unless they score below the benchmark on all four measures. Even then, there is only a 1 in 3 chance of being identified as Not Yet Meeting expectations in LA.

To test for a goodness of fit between student final LA scores and their DIBELS scores, a chi square test was performed. The LA scores were already nominal numbers representing

Meeting Expectations and Not Yet Meeting Expectations. Using the SPSS (version 15), the student scores on the DIBELS measures were transformed into nominal numbers representing scores above and scores below the School District No. 57 benchmarks.

The results of the chi square test found in Table 7 show that there is a fit between students scoring below the School District No. 57 benchmark and their LA score at the Grade 1 level, but not at the Kindergarten level as the χ^2 results for that population are all well below four while the values for Grade 1 are all well above the critical value of 3.84. The chi square results are similar to the biserial correlations - a significant relationship exists between the LA marks and DIBELS scores in Grade 1, but not in Kindergarten. The highest value attained at the Kindergarten level was with ISF $\chi^2(4, n = 463) = 1.82, p < .05$. On the other hand, the Grade 1 χ^2 values were statistically significant with NWF producing the highest value $\chi^2(4, n = 161) = 28.03, p < .005$. Such results mean that at the Grade 1 level, there is a likelihood that being identified as not yet meeting expectations in LA is dependent upon a student's performance on the DIBELS measures. On the contrary, the chi square results suggest that final LA marks are not related to the DIBELS scores at the Kindergarten level.

Table 7

Pearson Chi-Square Values Between Students Meeting Expectations in Language Arts and Students Meeting Expectations on DIBELS Measures

<i>Test</i>	<i>n</i>	<i>x</i> ²	<i>r</i> _{bis}
Kindergarten			
ISF	463	1.82	-.02
NWF	425	.20	.02
LNF	417	.08	.03
PSF	402	.00	-.06
Grade 1			
NWF	161	28.03***	.56
ORF	163	27.86***	.56
PSF	186	13.67***	.48

df = 1; *88p<.005 $.05x^2_{1\text{ critical}} = 3.84$

CHAPTER FIVE: DISCUSSION

Concurrent Validity

In 2002, Prince George School District No. 57 selected the University of Oregon's Dynamic Indicators of Basic Early Literacy Skills (DIBELS) as its performance indicators for the Kindergarten and Grade 1 levels. One of the research questions posed in this study is "What is the concurrent validity of the sub-tests of DIBELS?"

Testing for concurrent validity using teacher assigned final Language Arts grades was problematic and was limited by the high proportion of students identified at meeting expectations. With only 6% of Kindergarten and 8% of Grade 1 students in the study being identified as not yet meeting expectations in Language Arts (Table 1), the Language Arts data appears to be skewed as a result of a Type II error. A Type II error is a false negative. In relation to this study, a Type II error in the case of the Language Arts data means that a number of students who were identified as meeting expectations in Language Arts were in fact not meeting expectations. In other words, there is likely a greater number of students not meeting expectations than were actually identified by teachers.

Teachers at the Kindergarten and Grade 1 levels appear to have an unwillingness or reluctance to identify students as Not Yet Meeting Expectations. Other data suggests that a proportion greater than 6% or 4% of students typically do not meet expectations in reading in School District No. 57 (Prince George). Foundation Skill Assessment (FSA), for example, are standardized assessments administered by the Ministry of Education in British Columbia to students in Grade 4 and Grade 7 and include a reading comprehension component. Results for School District No. 57 (Prince George) show that in 2002-2003, the same year in which the Language Arts marks for this study were collected, 28% of Grade 4 students and 30% of Grade 7

students were not meeting expectations in reading comprehension (Ministry of Education, 2007a). Similarly, in the spring of 2006, 25% of Grade 4 students were not meeting expectations on the Reading FSA. That particular Grade 4 cohort represents the Grade 1 sample used in this study. While the sample population would have changed somewhat in the three years between Grade 1 and Grade Four, the data supports the likelihood that a Type II error is occurring at the Kindergarten and Grade 1 levels. Referring to Table 5, the number of students in Grade 1 who scored below school district benchmarks total 28 which is near 28% of the students. Had all those students who scored below the benchmark been identified as Not Yet Meeting Expectations in Language are, the percent not meeting expectations would have been similar to the 2006 Reading FSA results.

Despite the possibility of a Type II error affecting the LA data, a Pearson's product-moment correlation (r) was still calculated between the DIBELS measures and the teacher-assigned Language Arts mark, and between the DIBELS measures and WRC with the Grade 1 data. The r values generated for Kindergarten and provided in Table 2 ranged between .008 and -.086. Such low r values suggest no significant relationship between the two variables. Even when calculating the biserial correlations, the r values remained non-significant as they ranged from 0.03 to -0.06. A chi-square test also found no significant relationship between students achieving above School District No. 57 benchmarks and students identified as meeting expectations in Language Arts.

These Kindergarten results are not surprising considering the skewness and dichotomous nature of the LA scores. The Language Arts scores had only two categories, and almost all the scores were found in one category as 94% of the Kindergarten students and 92% of the Grade 1 students were identified as meeting expectations. The independent variables, on the other hand,

were not dichotomous and, therefore, less susceptible to skewness (see Table 1). Consequently, no relationship between the two variables was found to exist at the Kindergarten level which is contrary to other validity claims in the literature (for example, see Kaminsky & Good, 1996; Barger, 2003; Buck & Torgesen, 2003; Hinze, Ryan, & Stoner, 2003).

Despite not identifying concurrent validity for the separate measures using Pearson's product-moment correlations, establishing a benchmark provided an indication of some positive relationship between DIBELS assessments and the final Language Arts scores.

The benchmarks for Kindergarten established by this study (Table 4) suggest that the battery of DIBELS tests appear to be valid indicators of reading success. In other words, while the measures individually may not predict student reading success, the results of all four measures in combination have strength in predicting early reading success.

The data in Table 5 shows that 89% of Kindergarten students who were identified as Not Yet Meeting Expectations in Language Arts generated a score below the benchmark in at least three of the DIBELS measures. Further, 100% of Kindergarten students who were identified as Not Yet Meeting Expectations in Language Arts scored below the benchmarks in at least two of the measures. Using the battery DIBELS scores in conjunction with the benchmarks generated by this study, therefore, appears to be highly predictive in identifying students who are identified as Not Yet Meeting Expectations in Language Arts in Kindergarten.

However, not all students who did poorly on the battery of DIBELS assessments were identified by teachers as Not Yet Meeting Expectations in Language Arts. While 100% of Kindergarten students who were identified as Not Yet Meeting Expectations in Language Arts and scored below the benchmarks in at least two of the measures, close to one-third (29%) of those who were identified as meeting expectations in Language Arts also scored below the

benchmark on at least two of the measures. Such discrepancy suggests that while scoring below the benchmark does not necessarily mean a student will be identified as Not Yet Meeting Expectations in Language Arts, it does suggest a certain degree of risk for such identification.

In order to clarify the magnitude of being at risk when students scored below the benchmark on several measures, the benchmark data were translated into risk factors. These risk factors (Table 6) support the conclusion that the larger the number of measures with scores below the benchmark, the greater the risk factor. For instance, Kindergarten students who score below the benchmark in all four measures have a 44% chance that they will be identified as Not Yet Meeting Expectations at the end of the year.

As previously stated, the Grade 1 Language Arts data also appear to be limited by a Type II error. Despite that, evidence of concurrent validity of DIBELS and Reading WRC is stronger at the Grade 1 level. The Pearson product-moment correlations (Table 2) ranged from 0.27 to 0.31. The strongest correlations were with the NWF ($r = .31$) and ORF ($r = .31$) measures. These values signify a low-moderate strength in the relationship.

As might be expected because of the similarities in the type of measure, the correlations between WRC Reading and Language Arts scores ($r = .295$), and ORF and Language Arts scores ($r = .311$) were also similar. Such similarity is also evident in the high correlation between the two measures ($r = .870$).

Because the Language Arts final marks for Grade 1 were also an artificial dichotomy, biserial correlations were calculated and increased the r value range from 0.48 to 0.56 (Table 3). These are considered moderate positive relationships and help to confirm concurrent validity of DIBELS as indicators of Grade 1 student reading success. These results are consistent with other validity studies (Hinze, Ryan, & Stoner, 2003), but were not as large as the results produced by

Kaminsky and Good (1996). The strength of the relationship between the ORF and NWF measures with the LA measure were largest ($r = .56$), and once again WRC Reading and ORF had similar correlation values with LA. A chi-square test also found significant relationships between the DIBELS measures and the Language Arts scores as the χ^2 values in Table 7 ranged from 28.03 to 13.67.

As with the Kindergarten scores, the predictive strength of DIBELS at the Grade 1 level becomes evident as well by comparing the number of DIBELS measures in which students scored below the benchmarks to their final Language Arts score. All students who received a 1 (not yet meeting expectations) as a Language Arts final mark scored below the benchmarks in all four of the indicators (Table 5). In addition, and similar to the Kindergarten case, a large number of Grade 1 students were identified as meeting expectations despite scoring below the benchmark on all four measures.

In order to address this discrepancy once again, a risk factor was calculated for Grade 1. Table 8 shows that the risk of being identified as not yet meeting expectations in Language Arts increases with the number of DIBELS scores below the benchmarks. Students who score below the benchmarks on all four measures have a 36% chance of being identified as not yet meeting expectations in Language Arts. The risk decreases as the number of measures with scores above the benchmark increases.

Benchmarks

In order for the DIBELS and Reading WRC to be valid indicators of student reading performance, benchmark or cut-off scores that identify students who are expected to perform below grade expectations is important. The impetus for assessing students is to establish whether

or not they are within expectations in order to provide appropriate interventions to those who are not.

The University of Oregon has established its own benchmarks. Good, Simmons, Kame'enui, Kaminski and Wallin (2002) identified benchmarks for the various DIBELS assessments and these are found in Table 4. Because the benchmarks are based on a population different than the Prince George district, benchmarks based on a sample from the Prince George district would be more valuable to identify students requiring intervention.

The Kindergarten benchmarks established in this study are lower than the University of Oregon benchmarks. This result suggests that students in the School District No. 57 have to score lower than the samples used in the University of Oregon research in order to be identified as not yet meeting expectations. Needing to score lower is consistent with the Type II error in the Language Arts score as it means that many students are identified as meeting expectations in Language Arts despite having a low DIBELS score.

At the Grade 1 level, benchmarks for both NWF and PSF are higher than the University of Oregon benchmarks, while the ORF benchmark is slightly lower. The higher benchmarks at the Grade 1 level means that School District No. 57 students need to score higher on these measures than the population used for the University of Oregon studies before they are considered not yet meeting expectations. The benchmarks established in this study have a degree of latitude because the skewed Language Arts data made establishing definite benchmarks difficult. However, they are useful in creating the risk factors in Table 4.

As separate assessments at the Kindergarten level, ISF and LNF appear to be the stronger indicators of reading success as 1 of 5 students who score below the benchmark on these measures are identified as not yet meeting expectations in Language Arts. This strength is

supported with the chi-square test as the χ^2 value between these measures was also the highest, albeit non-significant, at 1.815 (Table 7). Students who score below the benchmark in PSF and NWF have about a 1 in 10 chance of being identified as not yet meeting expectations.

Similar risk factors can be made at the Grade 1 level. As a single assessment, NWF and ORF appear to provide the largest risk factor as 1 in 3 students who scored below the benchmark were identified as not yet meeting expectations in Language Arts at the end of the year. The χ^2 values between Language Arts scores and NWF ($\chi^2 = 28.03$) and ORF ($\chi^2 = 27.86$) also support this strong risk factor. The PSF subtest provides a 1 in 5 chance of identifying an at risk-reader.

These benchmarks increase the likelihood of identifying at-risk readers. Referring to Table 1, the probability of identifying a student as Not Yet Meeting Expectations is 8% at Grade 1 and 6% at Kindergarten. Use of the benchmarks could increase the likelihood of identifying at-risk students as the Kindergarten benchmarks capture 10% to 20% of the individuals while the Grade 1 benchmarks capture 20% to 30% of the individuals.

Limitations

Based on my experience as an elementary school administrator, and in my interactions with primary school educators, early literacy skills related to reading are the primary concern and the primary focus when considering the Language Arts placement of students of primary school age. Nonetheless, one of the limitations in this study is the final report for Language Arts performance levels. Because the Language Arts score is based on a combination of reading, writing, speaking, and listening, there is no control over the weight each component received in determining a level or the assessments used in determining a level. There is no systematic procedure that teachers across the district use when determining the Language Arts performance level.

Another limitation may be the pressure placed upon teachers to rate students as Meeting Expectations at the early stages of literacy. Considering a variety of factors including developmental factors, social and emotional factors as well as school based support, teachers appear to be reluctant to rate students as Not Yet Meeting Expectations in Language Arts in the early stages of literacy development.

Implications

Implications for Education

While the nature of the data, specifically the Kindergarten and Grade 1 Language Arts scores, contributed to making the results for a DIBELS validity test less conclusive than I would have preferred, some other issues that have educational implications arose from the results.

First, the descriptors or score categories used by teachers for identifying student performance levels in Language Arts and for reporting to the Ministry at the end of the year may be too limiting, thus providing an unclear, if not distorted report on student achievement. As mentioned in the discussion, the large proportion of students identified as Meeting Expectations in Language Arts at both the Kindergarten and Grade 1 levels is inconsistent with other data such as the FSAs. The Ministry is receiving what appears to be incomplete, distorted, or misleading data.

The skewed Language Arts marks appear to be the result of teachers having, at the end of the year, basically two options for reporting student ability in Language Arts for year-end reporting on the Permanent Record Cards. While teachers did have three possible choices - exceeding, meeting or not yet meeting expectations - the data that was gathered for this study identifies only a small number of instances where some schools used the “exceeding expectations” option. What appears to be the case, therefore, is that the number of choices available to teachers

to identify student performance in LA was, for practical purposes, only two - either a student was meeting expectations, or a student was not meeting expectations. There was no option to signify that a student was “borderline.”

However, such was not the case for reporting during the school year. At the time this data was collected, teachers in School District No. 57 had four options for reporting student progress on report cards. One score category for the report cards, “minimally meeting widely held expectations,” was available to teachers to indicate that a student was approaching grade standards and indicated that the student’s performance was borderline. Such a category allows teachers to communicate that there are some concerns about student performance. Had interim marks been used instead of final marks, stronger correlations between the teacher scores and the DIBELS scores may have been found.

At the end of the year, however, when reporting to the Ministry, teachers were faced with the decision of rating the “borderline” student as either “meeting” or “not yet meeting” expectations. Considering the social, emotional, and resource cost (e.g., learning assistance time) of such identification, there appears to be reluctance in identifying “borderline” students as not yet meeting expectations. This may account for the skewed results and the large number of students doing poorly on the measures, yet still being identified as meeting expectations.

By referring to the frequencies found in Table 5 in chapter four, clearly many students who scored below the benchmark scores on the various measures were still identified as “meeting expectations.” For instance, 20% of the students who were identified as “meeting expectations” in Grade 1 LA scored below the benchmarks in all four measures. These students may have been reported as “minimally meeting expectations” on the report cards, but the only options available to the teachers when reporting to the Ministry was “meeting” or “not yet

meeting” expectations. Quite possibly, therefore, a rating scale with more categories may assist teachers in identifying more at-risk readers when submitting Permanent Record Card data to the Ministry.

Being restricted to what is essentially a meeting/not yet meeting reporting system for Permanent Record Card reporting likely has little direct impact on identifying students at risk in the school. Having the four options available provides teacher with a score that indicates to parents that the child is possibly at risk - a warning flag of sorts. After such identification, teachers, parents, and other support staff can monitor such student more carefully and provide interventions to help the child move closer to the fully meeting expectations category.

Despite not affecting in-school identification of at risk readers, schools may be impacted in an indirect manner. When communicating to the public through such reports as the *Foundation Skills Assessment 2001/2002 - 2005/2006: School District 057 Prince George* (Ministry of Education, 2007a) the Ministry separates student performance into categories similar to those that primary teachers must use when submitting final marks: not meeting expectations, meeting expectations, and exceeding expectations. However, the individual student FSA reports that are sent home to parents contain a fourth category - minimally meeting expectations. This fourth category appears to be combined with the meeting expectations category in school reports, district reports, and indeed Ministry provincial level reports.

The first implication that is drawn from the study, therefore, is that while educators and parents (through progress reports) likely have a fairly accurate view of individual student achievement, the Ministry of Education in British Columbia and school districts, are likely receiving a less clear picture of student achievement in the sense that there are likely many students who are struggling but are not being identified by their teachers as Not Yet Meeting

Expectations in Language Arts. The Ministry and public is not getting information about those borderline students. With an increased emphasis on data-driven decision-making, the Ministry is receiving data that may not reflect the true picture of early literacy ability levels in the province.

A second implication emerges from this study. While the benchmarks are not definitive cut-off points and therefore to be used with caution, they do provide a starting point for which district benchmarks can be made. Certainly, clarity of district priorities is required before benchmarks can be established. Raising the benchmarks would increase the likelihood of false positives. That is, a higher benchmark would identify more potential at-risk readers, but, at the same time, identify more students who are not necessarily at-risk. Referencing Table 5, for instance, a lower benchmark would increase the number of Grade 1 students who score below the benchmark on four of the measures. Potentially, the 9% of students who scored below the benchmarks in three of the four measures and were identified as meeting expectations in Language Arts could end up scoring below the benchmark on all four measures. Since 36% of those who score below the benchmark on all four measures are identified as not yet meeting expectations, one would expect a substantial increase in the number of students requiring intervention.

Therefore, raising the benchmark would increase the number of students receiving intervention and reduce the risk of not identifying at-risk readers or letting students “fall between the cracks,” but it would also negatively impact financial and resource costs.

Conversely, lowering the benchmark scores for DIBELS would increase the likelihood of false negatives. In other words, doing so would decrease the number of students identified as requiring intervention. Such reduction would have a positive impact on the resource demands at

the school and district. However, a lowering of the benchmark would also increase the likelihood of at-risk readers continuing their education without appropriate interventions.

This study also has implications at the classroom level. This study demonstrates that when using DIBELS as an early literacy assessment, one should pay close attention to students who score below the benchmarks on three and especially all four DIBELS measures as there is a strong likelihood that they will be identified as “not yet meeting” expectations. DIBELS assessments occur three times a year, and students who score low in the fall should be monitored closely as the year continues.

Finally, this study raises the question as to whether or not Reading WRC and ORF are both needed at the Grade 1 level. The data throughout this study suggests that both tests have similar strengths at identifying at-risk readers. While ORF appears to have slightly higher validity, either, but not both, should be sufficient to measure reading fluency.

Implications for Further Research

A key question that emerges from this research is “Why is there such a large number of students identified as meeting expectations in Language Arts when they perform just as poorly as those identified as not yet meeting expectations on the DIBELS measures?”

One possible area of research may be to investigate what other aspects of reading require valid assessment strategies, or what other measures teachers are using to assess student ability in Language Arts. Many teachers use other classroom assessments such as Running Records and assessments that measure comprehension, and finding out the degree to which they rely on these other measures when reporting student reading may address the key issue.

A second and similar area of investigation would be to examine the degree to which teachers use DIBELS and WRC results to guide instruction and reporting. In 2003-2004 when these data were collected, school planning and assessment based instruction was in its infancy. In years prior to this data collection, teachers depended upon their own assessment practices which they trusted and with which they had familiarity. Assessment for learning strategies are a new development in education, and teachers may not have sufficient training in using assessment to guide instruction. Are DIBELS and WRC assessments shared and examined by teachers and support staff, or is the data collected and stored for the purposes of school planning?

The third area of research is connected to the second. Teachers are with their students daily, interacting with them, hearing them read, watching them read, and helping them choose appropriate reading material. Experienced teachers have worked with hundreds of young readers and have established practices for instruction, assessment, and intervention. Is there a difference in the reporting of experienced teachers versus less experienced teachers? In other words, is the number of students who do poorly on DIBELS and WRC measures and identified as meeting expectations greater from a class with an inexperienced teacher?

A final area of research that emerges from this study involves using progress report data instead of Permanent Report Card scores. The progress reports use a four scale scoring system and, unlike the Permanent Report Card scores, include a score category to identify students who are not fully meeting expectations, but are not considered to be not yet meeting expectations. Using such scores may provide more accurate corrections and more sensitive benchmark data. Similarly, using a population where rating students as Not Yet Meeting Expectations occurs more frequently would provide greater clarity about the relationship between DIBELS and early literacy development.

References

- Adams, M. J. (1990). *Beginning to read: Thinking & learning about print*. Cambridge: MA MIT Press.
- Barger, J. (2003). *Comparing the DIBELS Oral Reading Fluency indicator and the North Carolina end of grade reading assessment* (Technical Report). Ashville, NC: North Carolina Teaching Academy.
- Big Ideas in Beginning Reading (2000 - 2004) Retrieved November 6, 2004, from http://reading.uoregon.edu/site_map.php
- Blachman, B. A. (1984). Relationship of rapid naming ability and language analysis skills to kindergarten and first-grade reading achievement. *Journal of Educational Psychology*, 76, 610-622.
- Blachman, B.A. (1989). Phonological awareness and word recognition: Assessment and intervention. In A. B. Kamhi & H. W. Catts (Eds.), *Reading disabilities: A developmental language perspective* (pp. 133-158). Boston: College-Hill.
- Black, P., and Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education*, 5 (1): 7-74.
- Black, P., & Wiliam, D. (1998b). Inside the black box. *Phi Delta Kappan*; 80, (2) Issue 2, 139-147.
- Buck, J. & Torgesen, J. (2003). The relationship between performance on a measure of oral reading fluency and performance on the Florida Comprehensive Assessment Test, *Florida Center for Reading Research*.
- Byrne, B. (1991). Experimental analysis of the child's discovery of the alphabetic principle. In Laurence Rieben and Charles A. Perfetti (Eds.), *Learning to Read: Basic research and its implications* (75-84). Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Byrne, B. (1992). Studies in the acquisition procedure for reading: Rationale, hypotheses and data. In Philip B. Gough, Linnea C. Ehri & Rebecca Teiman (Eds.), *Reading Acquisition* (1-34). Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Carnine, D. W., Silbert, J., & Kameenui, E. J. (1997). *Direct instruction reading* (3rd ed.). Upper Saddle River, NJ: Merrill/Prentice-Hall.
- Chard, D. J. & Dickson, S. V. (1999) Phonological awareness: Instructional and assessment. *Intervention in School and Clinic*, 34(5), 261-270.
- Clark, P., Owens, T., & Sutton, R. (2006) *Creating independent student learners, school leaders: A practical guide to assessment for learning*. Winnipeg: Portage and Main Press.

- Clay, Marie. (1985). *The early detection of reading difficulties*. Heinemann: Hong Kong.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.
- Collins V. L., Good, R. H., Knutson N., Shinn, M. R. and Tilly, W. D. (1992). Curriculum-based measurement of oral reading fluency: A confirmatory analysis of its relation to reading. *School Psychology Review*, 21(3), 469-479.
- Crocker, L. & Algina, J (1986) *Introduction to classical and modern test theory*. New York: Holt, Rinehart and Winston..
- Felton, R. H., & Wood, F.B. (1989). Cognitive deficits in reading disability and attention deficit disorder. *Journal of Learning Disabilities*, 22, 3-13.
- D'Angiulli, A. & Siegel, L. S. (2003) Cognitive Functioning as Measured by the WISC-R. *Journal of Learning Disabilities*, 36(1), 48-58.
- Deno, S.L., Mirkin, P.K., & Chiang, B. (1982). Identifying valid measures of reading. *Exceptional Children*, 49, 36-45.
- Deno, S.L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children*, 52(3), 219-232.
- Dynamic Indicators of Basic Early Literacy Skills. (2000-2004). Retrieved November 6, 2004, from <http://dibels.uoregon.edu/measure.php>.
- Elliot J., Lee S. W., & Tollefson, N. (2001). A reliability and validity study of the Dynamic Indicators of Basic Early Literacy Skills – Modified. *School Psychology Review*, 30(1).
- Felton, R. H., & Wood, F.B. (1989). Cognitive deficits in reading disability and attention deficit disorder. *Journal of Learning Disabilities*, 22, 3-13.
- Fewster, S.(2000). *School Based Evidence for the Validity of Curriculum -Based Measurement Norms in School District #57*. Unpublished master's thesis, University of Northern British Columbia, Prince George, British Columbia.
- Foorman, B. R., Fletcher, J.M., Schatschneider, C. & Mehta, P. (1998). The role of instruction in learning to read: Preventing reading failure in at-risk children. *Journal of Educational Psychology*, 90(1), 37-55
- Fowlert, A. E. (2004). Relationship of naming skills to reading, memory, and receptive vocabulary: Evidence for imprecise phonological representation of words by poor readers. *Annals of Dyslexia*, 54 (2), 247 - 280

- Fullan, M., Bertani, A., & Quinn, J. (2004). New lessons for district wide reform. *Educational Leadership*: 61(7), 42-48.
- Gray, A. & McCutchen. (2006). Young readers' use of phonological information: phonological awareness, memory, and comprehension. *Journal of Learning Disabilities*, 39 (4), 325-333.
- Glass, G. & Hopkins, K (1996). Statistical methods in education and psychology (3rd ed.). Boston: Allyn and Bacon.
- Good, R. H., Simmons, D. C., & Smith, S. B. (1998).). Effective academic interventions in the United States: Evaluating and enhancing the acquisition of early reading skills. *School Psychology Review*, 29(1), 45-56.
- Good, R.H., & Jefferson, G. (1998). Contemporary perspectives on curriculum based measurement validity. In M.R. Shinn (Ed.), *Advanced applications of curriculum-based measurements* (pp. 61-88). New York: Guilford Press.
- Good, R.H., Wallin, J., Simmons, D.C., Kame'enui, E.J., & Kaminsky, R. A. (2002). System-wide percentile ranks for DIBELS Benchmark assessment (Technical Report 9). Eugene, OR: University of Oregon.
- Good, R. H., Simmons, D., Kame'enui, E., Kaminski, R. A., & Wallin, J. (2002). *Summary of decision rules for intensive, strategic, and benchmark instructional recommendations in kindergarten through third grade* (Technical Report No. 11). Eugene, OR: University of Oregon.
- Goswami, U. & Bryant, B. (1992). Rhyme, analogy, and children's reading. In Philip B. Gough, Linnea C. Ehri & Rebecca Teiman (Eds.), *Reading Acquisition* (49-64). Hillsdale, New Jersey: Lawrence Erlbaum Associates
- Gough, P. (1992). Reading, spelling, and the orthographic cipher. In Philip B. Gough, Linnea C. Ehri & Rebecca Teiman (Eds.), *Reading Acquisition* (35-48). Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Hinze, J. M., Ryan, A. L. & Stoner, G. (2003). Concurrent validity and diagnostic accuracy of the Dynamic Indicators of Basic Early Literacy Skills and the Comprehensive Test of Phonological Processing. *School Psychology Review*, 32(4).
- Kaminski, R. A., & Good, R. H. (1998). Assessing early literacy skills in a problem-solving model: Dynamic indicators of basic early literacy skills. In M.R. Shinn (Ed.), *Advanced applications of curriculum-based measurements* (113- 142). New York: Guilford Press.

- Kaminski, R. A., & Good, R. H. (1996). Toward a technology for assessing basic early literacy skills. *School Psychology Review*, 25 (2) , 215-227.
- Jenkins, J.R., Deno, S.L., & Mirkin, P.K. (1979). Measuring pupil progress toward the least restrictive environment. *Learning Disability Quarterly*, 2, 81-92.
- Liberman, Y. & Shankweiler, D. (1987). Phonology and the problems of learning to read and write. *Advances in Learning and Behavioral Disabilities*, Supplement 2 , (pp. 203 - 234).
- Liberman, I. Y., Shankweiler, D. & Liberman, A.M. (1989). The alphabetic principle and learning to read. In Shankweiler, D & Liberman, I.Y. (Eds.) *Phonology and Reading Disability: Solving the reading puzzle*. (pp. 1 - 33). Ann Arbor: University of Michigan Press.
- Mann, V.A. (1993) Phoneme awareness and future reading ability. *Journal of Learning Disabilities*, 26(4), 259-69.
- Marston, D.B. (1989). A curriculum based measurement approach to assessing academic performance: what it is and why do it. In M.R. Shinn (Ed.), *Advanced applications of curriculum-based measurements* (18- 78). New York: Guilford Press.
- Ministry of Education. (2004a). *Student progress report order191/94*. Retrieved November 6, 2004, from http://www.bced.gov.bc.ca/policy/policies/student_reporting.htm
- Ministry of Education. (2004b). *Student progress report order190/91*. Retrieved November 6, 2004, from http://www.bced.gov.bc.ca/policy/policies/student_reporting.htm
- Ministry of Education. (2005a). *Foundation skills assessment*. Retrieved January 20, 2005 from <http://www.bced.gov.bc.ca/assessment/fsa/>
- Ministry of Education. (2005b). *Provincial exam reports & information*. Retrieved January 20, 2005, from <http://www.bced.gov.bc.ca/reporting/exams/>
- Ministry of Education. (2005c). *Satisfaction survey*. Retrieved January 20, 2005, from http://www.bced.gov.bc.ca/sat_survey/
- Ministry of Education. (2007a). *Foundation Skills Assessment 2001/2002 - 2005/2006: School District 057 Prince George*. Retrieved August 26, 2007, from <http://www.bced.gov.bc.ca/reports/pdfs/fsa/057.pdf>
- Ministry of Education. (2007b). *District Data Summary 2001/02 - 2005/06: 057 Prince George*. Retrieved August 25 from http://www.bced.gov.bc.ca/reports/pdfs/sd_perf/057.pdf

- Moats, L. C. (1999). Teaching reading is rocket science: What expert teachers of reading should know and be able to do. Washington, D. C.: American Federation of Teachers.
- Nelson, J.R., Benner, G. J., & Gonzalez, J. (2003) Learner characteristics that influence the treatment effectiveness of literacy interventions: A meta-analytic review. *Learning Disabilities Research and Practice*, 18(4), 255-267.
- Perfetti, C. A. (1999). Cognitive research and the misconceptions of reading education. In Oakhill, J. & Beard, R. (Eds.) *Reading development and the teaching of reading*, (pp. 42-58). Malden, Massachusetts : Blackwell.
- Pikulski, J. J. & Chard, D. J. (2005). Fluency: Bridge between decoding and reading comprehension. *The Reading Teacher*, 58 (6), 510-518.
- Phillips, B.M. & Torgesen, J.K.(2006). Phonemic awareness and reading: Beyond the growth of initial reading accuracy. In Neuman, S. B. & Dickinson, D.K. (Eds.)*Handbook of Early Literacy Research*, (pp. 101-112), New York: Guilford Press.
- Popham, J.W. (2004). Why assessment illiteracy is professional suicide. *Educational Leadership*, 62 (1), 82-83.
- Ross, J. A. The effects of running records assessment on early literacy assessment. *Journal of Educational Research*, 97(4), 186-195.
- Prince George School District No.57 (1996). *Guidebook for the use of curriculum based measurement in School District #57*. Prince George, British Columbia: School District #57.
- Prince George School District No.57 (2000). Performance Standards File (Version 1.6) [Computer software]. Prince George, British Columbia.
- Shaw, R & Shaw, D. (2002). DIBELS oral reading fluency-based indicators of third grade reading skills for Colorado State Assessment Program. (Technical Report). Eugene, OR: University of Oregon.
- Shinn, M.R. (1989). Identifying and defining academic problems: WRC screening and eligibility procedures. In Shinn (Ed.), *Curriculum-based measurement: Assessing special children*, (pp. 90-129). New York: Guiliford Press.
- Shinn, M.R., Good, R.H., Knutson, N., Tilly, WD., & Collins, V. L. (1992). Curriculum-based measurement of oral reading fluency: A confirmatory analysis of its relation to reading. *School Psychology Review*, 21(3), 459-479.
- Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
- Stahl, S. A., & Murray, B. A. (1994). Defining phonological awareness and its relationship to early reading. *Journal of Educational Psychology*, 86, 221-234.

- Stanovich, K.E. (1991). Changing models of reading and reading acquisition in . In L. Rieben & C.A. (Eds.) *Learning to Read* (pp. 33 –46). Hillsdale, New Jersey : Lawrence Erlbaum Associates.
- Stanovich, K.E (1994) Romance and reality. *The Reading Teacher*, 47 (4), 280-291.
- Treiman, R. (1992). The role of intrasyllabic unity in learning to read and spell. In Gaough, P.B., Ehri, L. C., and Treiman, R. (Eds.) *Reading Acquisition* (65-106). Hillsdale, New Jersey : Lawrence Erlbaum Associates.
- Torgesen, J. K. (2000) Individual Differences in Response to Early Interventions in Reading. *Learning Disabilities Research & Practice*, 15(1), 55-64.
- Torgesen, J. K. & Wagner, R. K. (1994) Longitudinal studies of phonological processing and reading. *Journal of Learning Disabilities*, 27(5), 276-286.
- Uhry,J. (2002) Finger point reading in kindergarten: The role of phonemic awareness, one-to-one correspondence, and rapid serial naming. *Scientific Studies of Reading*, 6(4), 319-342.
- University of Oregon Center on Teaching and Learning (2004) Benchmark Goals and Indicators of Risk Three Assessment Periods Per Year. Retrieved November 10, 2004, from <http://dibels.uoregon.edu/benchmarkgoals.pdf>.
- Vandre Meer, C. D., Lentz, F. E. and Stollar, S. (2005). The relationship between oral reading fluency and Ohio proficiency testing in reading (Technical Report). Eugene, Oregon: University of Oregon.
- Wilson, J. (2005). The relationship of Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Oral Reading Fluency to performance on Arizona Instruments to Measure Standards. Tempe, Arizona: Tempe School District No. 3.
- Wong, J. (1985) Turbo School [Computer Software]. Vancouver, British Columbia: TCS Developments Inc.
- Ysseldyke, J. E. & Christensen, S.L. (1988). Linking assessment to intervention. In J. L. Graden, J. E. Zins, & M. J. Curtis (eds.). *Alternative educational delivery systems: Enhancing instructional options for all students* (pp. 91-110). Washington, DC: National Association of School Psychologists.

APPENDIX A

Sample DIBELS Measures : Grade 1

Phoneme Segmentation Fluency

Short Form Directions

Refer to Administration and Scoring Guide p. 16

Phoneme Segmentation Fluency

I am going to say a word. After I say it, you tell me all the sounds in the word. No, if I say, "sum," you would say /s/ /u/ /m/. Let's try one. (one second pause) Tell me the sounds in "mop".

CORRECT RESPONSE: If student says, /m/ /u/ /p/, you say,	INCORRECT RESPONSE: If student gives any other response, you say,
Very good. The sounds in "mop" are /m/ /u/ /p/.	The sounds in "mop" are /m/ /u/ /p/. Your turn. Tell me the sounds in "mop".

OK. Here is your first word.

Correct Segmentation: a correct sound segment is any different, correct, part of the word represented by sounds that correspond to the word part. For example, the sound /t/ is a correct sound segment of "trick", as are /tr/ and /trɪ/.

Benchmark 3

Phoneme Segmentation Fluency*

star	/s/ /t/ /ər/	give	/g/ /ɪ/ /v/	/6
yet	/y/ /e/ /t/	bid	/b/ /ɪ/ /d/	/6
steal	/s/ /t/ /i/ /əl/	walk	/w/ /ɔ/ /k/	/7
you	/y/ /oo/	brook	/b/ /roo/ /k/	/5
she	/sh/ /ee/	swing	/s/ /w/ /ɪ/ /ŋ/	/6
coal	/k/ /oa/ /l/	oak	/oa/ /k/	/5
safe	/s/ /a/ /f/	bones	/b/ /oa/ /n/ /z/	/7
moves	/m/ /oo/ /v/ /z/	rooms	/r/ /oo/ /m/ /z/	/8
guess	/g/ /e/ /s/	stole	/s/ /t/ /oa/ /l/	/7
light	/l/ /ee/ /t/	she	/sh/ /ea/	/5
crow	/k/ /oa/ /r/	crow	/k/ /r/ /oa/	/6
mad	/m/ /a/ /d/	face	/f/ /ai/ /s/	/6

Total

Form P-Item

*See p. 2 this booklet for directions

* This research was supported, in part, by the Early Childhood Research Institute on Monitoring Growth and Development (1180241006) funded by the U. S. Department of Education, Special Education Programs
Good, R. H., & Kaminski, R. A. (Eds.). (2001). *Dynamic Indicators of Basic Early Literacy Skills* (5th ed.). Eugene, OR: Institute for the Development of Educational Achievement. Available: <http://dibels.iuegon.edu/>

Sample Letter Naming Fluency (Fall Only) - Grade One

Benchmark 1 Letter Naming Fluency

Short Form Directions

Here are some letters (point to the student probe). Tell me the names of as many letters as you can. When I say "begin", start here (point to first letter), and go across the page (point). Point to each letter and tell me the name of that letter. If you come to a letter you don't know I'll tell it to you. Put your finger on the first letter. Ready, begin.

K	q	T	n	s	S	g	G	D	W
g	U	g	S	s	t	B	P	A	j
Z	V	l	r	D	G	Q	H	I	N
C	C	Z	l	a	f	p	e	r	h
z	s	M	p	T	P	m	v	X	w
K	E	l	y	d	r	n	z	n	i
N	E	X	F	a	H	R	Y	o	w
J	c	W	m	d	R	O	i	y	c
b	J	l	e	h	L	O	A	B	U
u	Q	V	y	t	L	F	k	x	Y
M	u	k	b	K	q	T	n	x	S

Total:

Sample Nonsense Word Fluency - Grade 1

Nonsense Word Fluency Short Form Directions

Each at this word point to the first word on the practice sheet. It's a make-believe word. Watch me read the word: /m/ /h/ /m/ "sim" (point to each letter then run your finger last beneath the whole word). *I can say the sounds of the letters, /s/ /m/ /m/ (quietly each letter), or I can read the whole word "sim" (run your finger last beneath the whole word).*

You turn to read a make-believe word. Read this word the first you can (get into the word "m"). Make sure you say every sound you hear.

FOR THE RESPONSE:	INTERPRET OR NO RESPONSE
If the child responds "no" or with something other than the word.	If the child does not respond within 1 second, or responds incorrectly, say:
That's right. The sounds are /h/ /m/ /m/ or "sim"	Remember, you can say the sounds or you can say the whole word. Watch me: the sounds are /h/ /m/ /m/ (quietly each letter) or "sim" (run your finger last through the whole word). Let's try again. Read this word the best you can (point to the word "m").

Place the student copy of the probe in front of the child.

Here are some make-believe words (point to the student probe). Start here (point to the first word) and go across the page (point across the page). When I say "begin", read the words the best you can. Put your finger on the first word. Ready, begin.
Prompt: If a child seems unsure of what to do, say "Remember you can point to each letter and tell me the sound or read the whole word".

Benchmark 3 Nonsense Word Fluency

uk	sot	pez	hus	sav	/14
vol	wuf	bav	fik	mak	/15
rav	nid	rev	zcd	vaf	/15
nop	jip	kep	dif	sus	/15
mip	jus	lot	og	ef	/13
hic	fum	rij	niz	je	/14
dij	zuf	ruc	ig	hur	/14
miz	ju	lac	kog	wog	/14
ov	dos	zec	ap	lak	/13
pej	joz	wij	lim	bif	/15
Total:					

Error Pattern:

See p. 4 this booklet for directions.

Sample Oral Reading Fluency - Grade 1

Benchmark 2

DIBELS Oral Reading Fluency Passage 2

Do not administer if the student scored below 10 on passage 1.

When I say begin, start reading aloud at the top of the page. Read across the page. Try to read each word. If you come to a word you don't know, I'll tell it to you. Be sure to do your best reading. The name of this story is "My Soccer Team". Ready, Begin. At the end of 1 minute, place a bracket () after the last word and say "Stop."

My Soccer Team

I am so happy! I just found out I can be on the soccer	14
team. We have our first practice on Saturday. We practice at	25
my school right after lunch.	30
Our team is called the Blue Bombers. Our colors are blue	41
and white so I get to wear blue shorts and a blue and white	55
shirt. The number on my shirt is seven. I'm seven years old,	67
too. I think seven must be my lucky number.	76
We play our first game next week on Saturday. I can't	87
wait to play. My dad said if I practice a lot I will do well at	103
the games. My dad is going to practice with me tonight.	114
Right after dinner my dad is going to take me to the store	127
to buy some soccer shoes and a soccer ball. Then we will play	140
on the grass by my school. My dad will help me to kick the	154
ball and to run fast and kick the ball at the same time.	167
I am so excited I don't think I will get to sleep tonight.	180
But I better sleep so that I can be rested and strong for my	194
soccer practice.	196

Total: _____

Kind Parent:

APPENDIX B

Reading WRC – Grade 1

Reading WRC – Grade 1

Probe 1
G.OTGO.771

A long, long time ago, some men
were sitting around a fire.
They were looking at smoke going up
into the air.
This made them think.
Could smoke make a bag go up too?
The men wanted to see.

They took a bag and held it over the
smoke. The warm air made the bag
go up into the sky.
Then the men made a big bag.
Under this first balloon they put a basket.
In the basket they put a pan.
In the pan was a fire.
The fire made the air warm.
The warm air made the big balloon go up too!
The men made one more balloon.
It, too, had a basket and a pan.
There was a fire in the pan.
But this time there was something new.
In the basket were a duck, a rooster,
and a sheep.
When the balloon went up into the sky,
the animals went too.
When the balloon came down,
the animals were safe.

APPENDIX C

School District No. 57 Approval Letter



SCHOOL DISTRICT NO. 57 (PRINCE GEORGE)

2100 Ferry Avenue, Prince George, B.C. V2L 4R5

Phone: (250) 561-6800 Fax: (250) 561-6801
www.schdist57.bc.ca

October 27, 2004

Ted Zarowny
P.O. Box 357
McBride, BC
V0J 2E0

Dear Ted:

This letter is to confirm our email and phone discussions regarding your request to obtain access to schools in the Prince George School District for the purpose of educational research. As we discussed, the school district recognizes the integral part that research plays in education. We support the research sponsored by our local tertiary institutes as a priority. Your project will provide valuable information for our schools and for our district.

This letter's purpose is to indicate that you have district approval to proceed with your project. ***"District approval" allows the researcher to approach principals, and where appropriate, teachers to request permission for access to the assessment data.*** Researchers must understand that circumstances may be difficult and school administrators have the final decision. Your next step will be to contact the principals of the schools in order to request their assistance in extracting the assessment data.

If you have any questions, please do not hesitate to call me. Good luck with your project. I look forward to receiving a copy of the final report.

Sincerely,

Bonnie Chappell
Director, School Services

CC: Elementary Principals

BC/hg

APPENDIX D

Benchmark Graphs

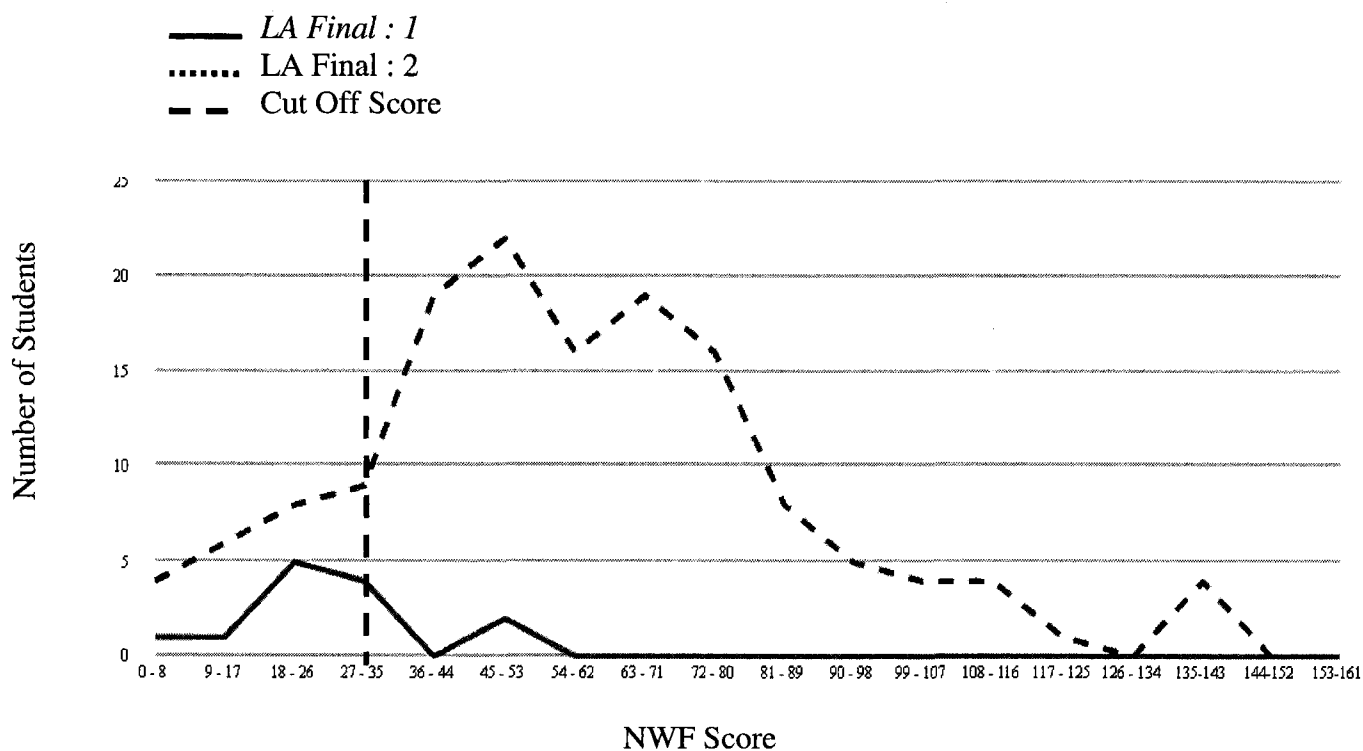


Figure 1

Benchmark for grade 1 NWF (Score Category = 9).

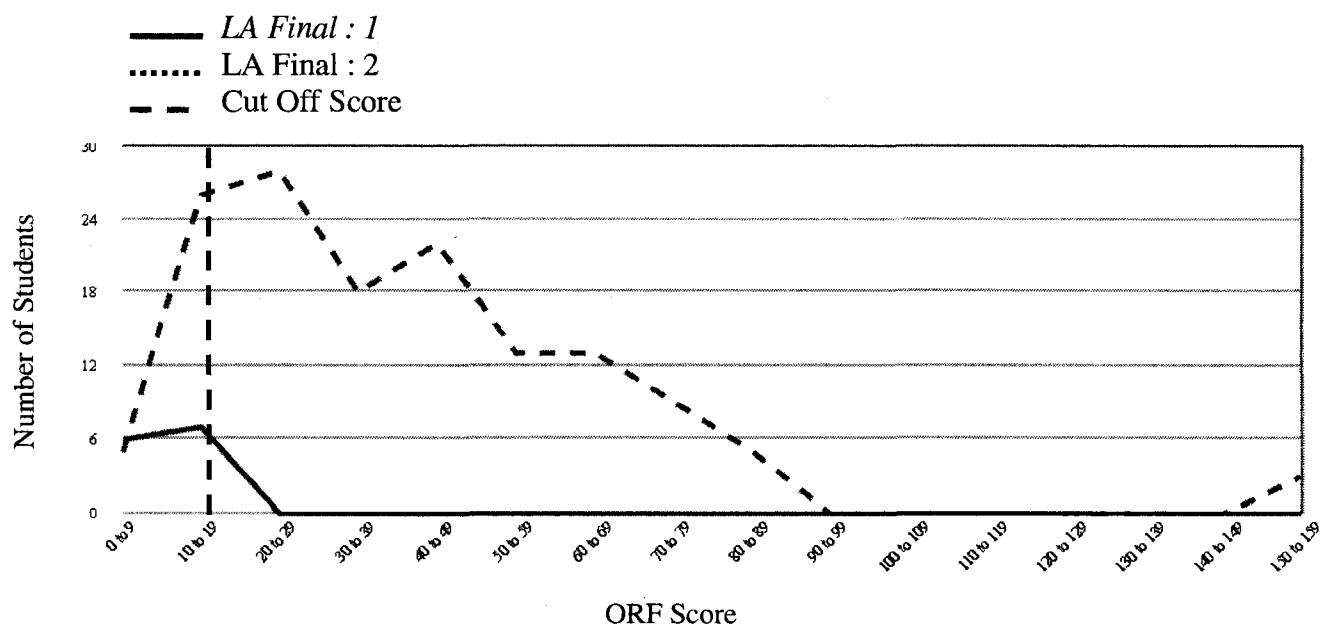


Figure 2

Benchmark for grade 1 ORF (Score Category = 10).

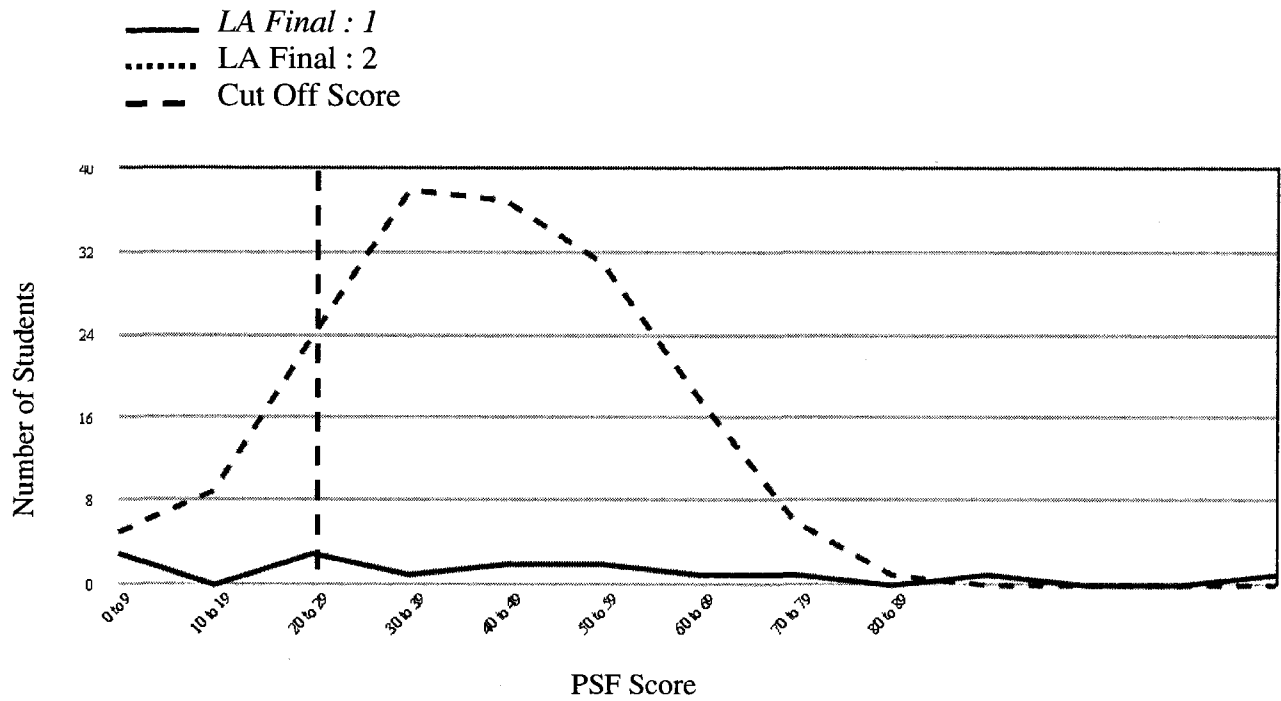
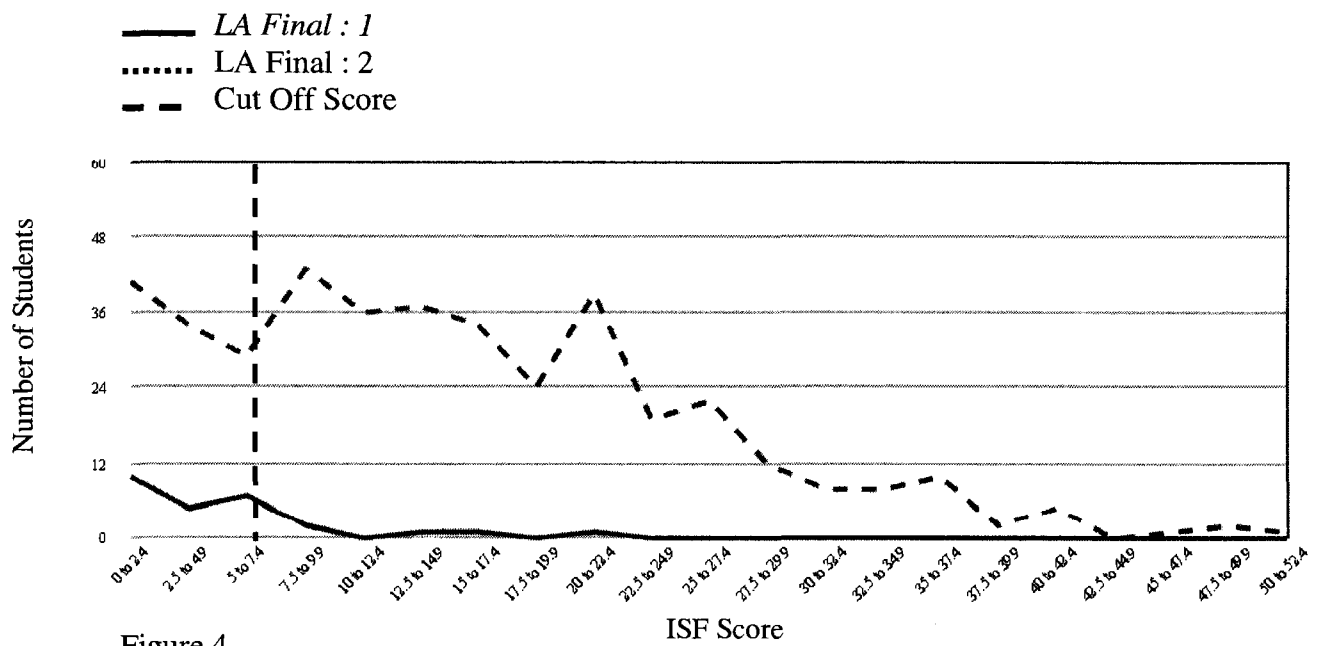


Figure 3

Benchmark for grade 1 PSF (Score Category = 10).



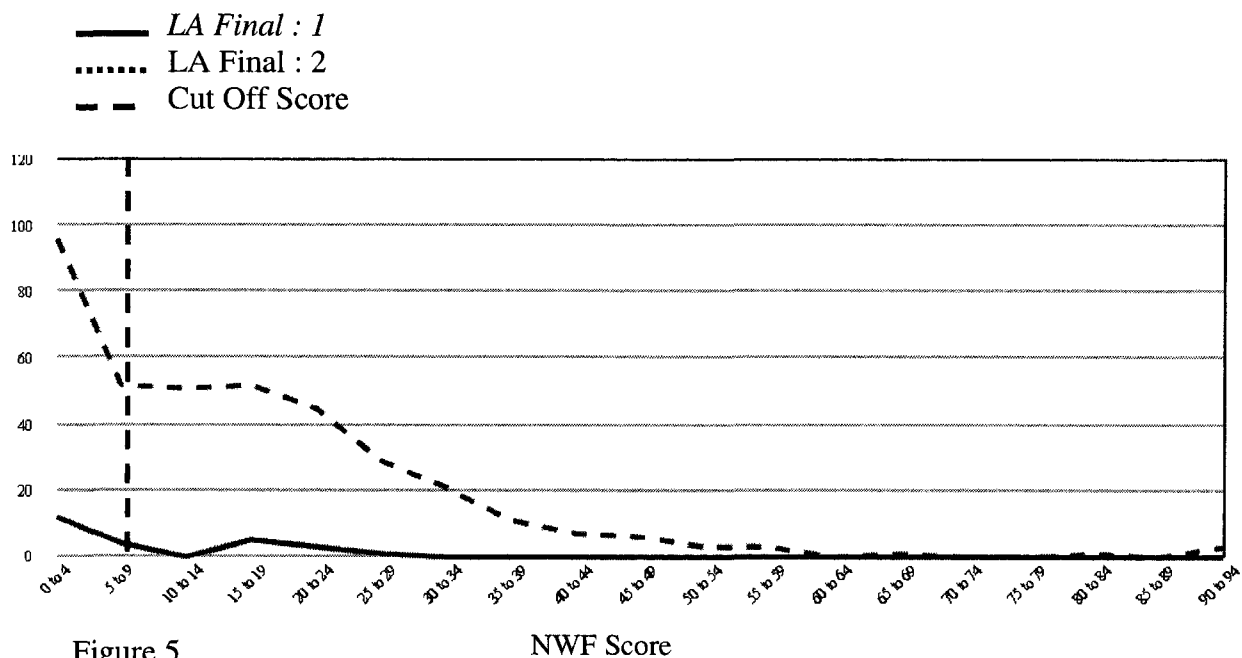


Figure 5

Benchmark for kindergarten NWF (Score Category = 5).

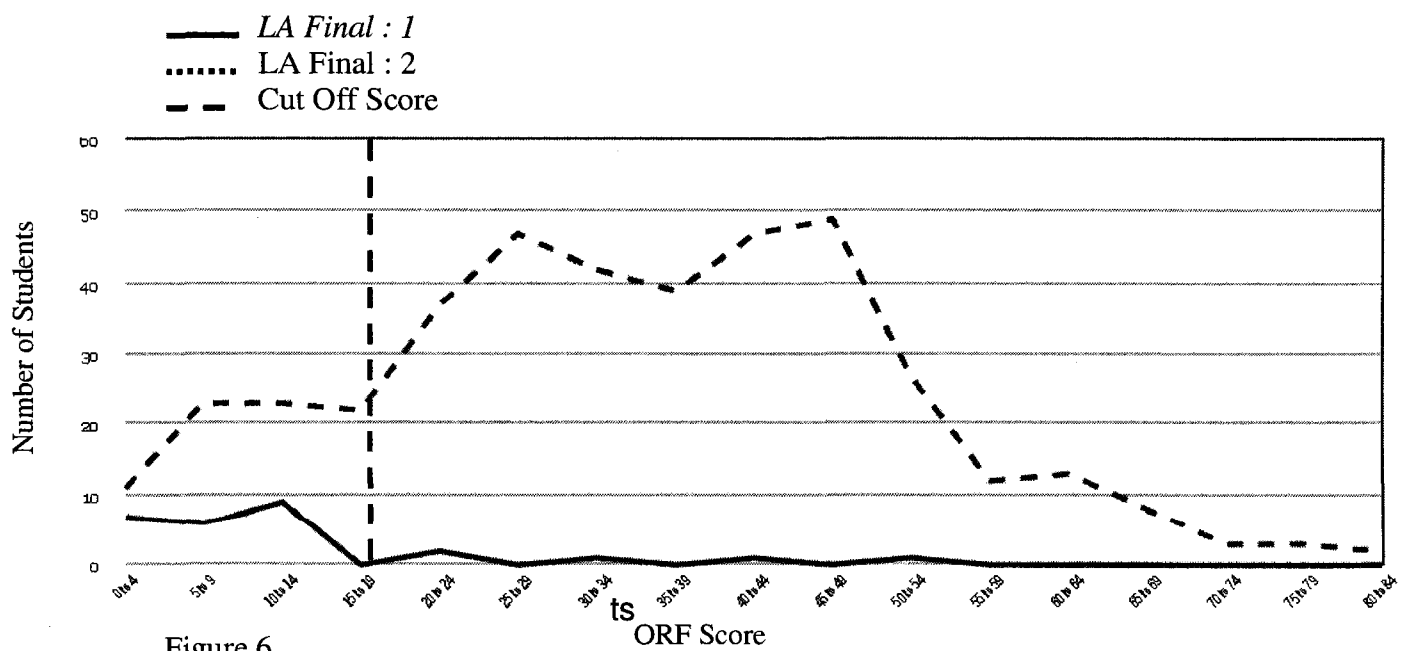


Figure 6

Benchmark for kindergarten LNF (Score Category = 5).

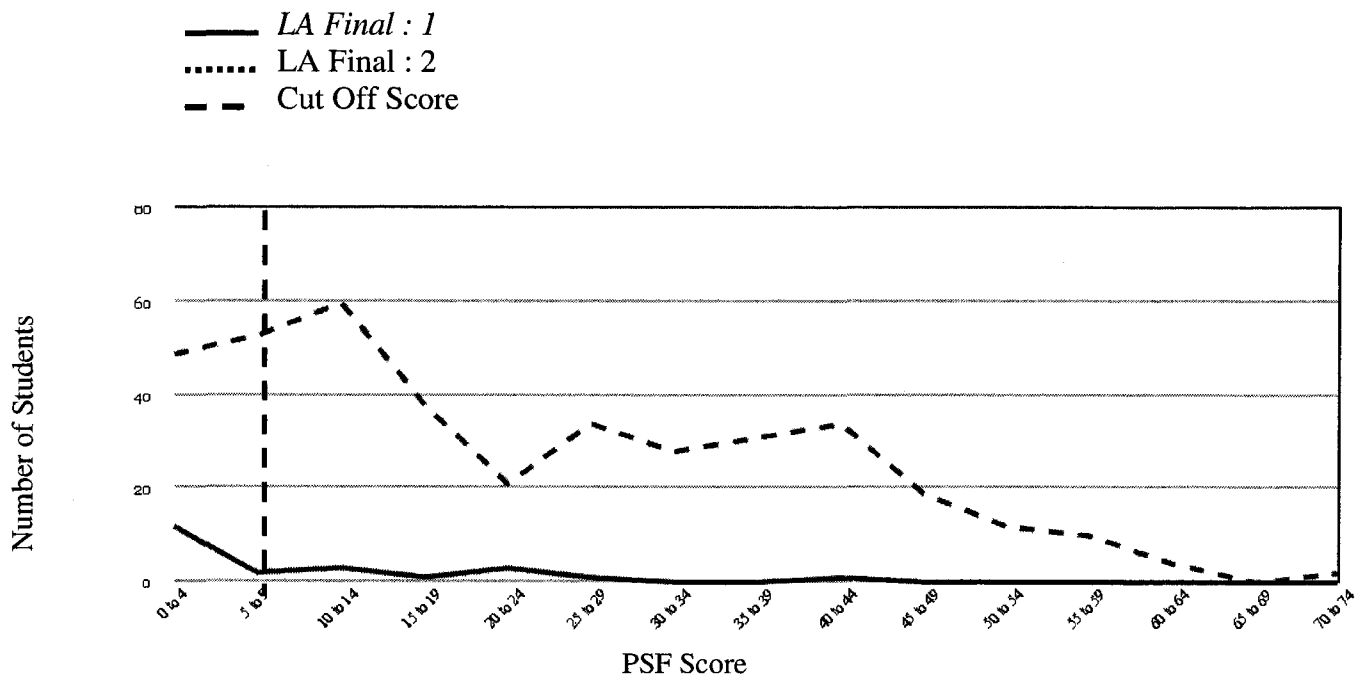


Figure 7

Benchmark for kindergarten PSF (Score Category = 5).