5-1-2001

Test-Retest Reliability of Phonemic Awareness Assessment Instruments of Kindergarten Students

Lorie Craycroft
Western Kentucky University

Follow this and additional works at: http://digitalcommons.wku.edu/theses
Part of the Education Commons, and the Psychology Commons

Recommended Citation
http://digitalcommons.wku.edu/theses/687

This Thesis is brought to you for free and open access by TopSCHOLAR®. It has been accepted for inclusion in Masters Theses & Specialist Projects by an authorized administrator of TopSCHOLAR®. For more information, please contact topscholar@wku.edu.
TEST-RETEST RELIABILITY OF PHONEMIC AWARENESS ASSESSMENT INSTRUMENTS ON KINDERGARTEN STUDENTS

A Thesis

Presented to

the Faculty of the Department of Psychology

Western Kentucky University

Bowling Green, Kentucky

In Partial Fulfillment

of the Requirements for the Degree

Specialist in Education

by

Lorie Ann Craycroft

May 2001
TEST-RETEST RELIABILITY OF PHONEMIC AWARENESS ASSESSMENT INSTRUMENTS ON KINDERGARTEN STUDENTS

Date Recommended 5-3-00

[Signature]
Director of Thesis

[Signature]
Dean, Graduate Studies and Research  Date
Acknowledgments

There are many important people to thank for the completion of this project. First, I would like to thank Kelli Bradbury and my other fellow graduate students for their constant support and empathy throughout the thesis process. I would also like to thank Jamie Byrne and Christen Gibbs for the many hours they contributed in collecting the data. I would like to thank Bowling Green Independent Schools for so graciously allowing their kindergarten students to participate in the study and WKU Graduate Studies and Research for providing funding for the project.

This project would never have gotten started nor completed without the tremendous amount of help from my thesis chair, Dr. Carl Myers. Without him I would have never been able to complete such a project of which I am so very proud. I would also like to thank my other committee members, Dr. Reagan Brown and Dr. Frank Kersting, who contributed many hours and provided exceptional guidance throughout this project.

I would also like to thank my parents, Gary and Jackie Dolwick. Their guidance and motivation throughout my life have enabled me to have the courage and determination to achieve my dreams.

Finally, I would like to thank my husband, David Craycroft, for his endless amount of support and understanding throughout my graduate training. Without him I would never have been able to endure the long hours and the tremendous amount of effort involved in receiving my graduate degree.
# Table of Contents

List of Tables ............................................................................................................. v  
List of Figures ............................................................................................................. vi  
Abstract ..................................................................................................................... vii  
Introduction ............................................................................................................... 01  

Literature Review .................................................................................................... 04  
  What is Phonemic Awareness?............................................................................... 04  
  The Importance of Phonemic Awareness............................................................ 06  
  Phonemic Awareness Tasks.................................................................................. 07  
  Standardized Phonemic Awareness Instruments.............................................. 09  
  Purpose of Present Study...................................................................................... 15  

Method ..................................................................................................................... 17  
  Participants ........................................................................................................... 17  
  Materials .............................................................................................................. 18  
  Procedure ............................................................................................................ 18  

Results .................................................................................................................... 20  
  Comprehensive Test of Phonological Processing............................................ 20  
  Test of Phonological Awareness ..................................................................... 22  
  Yopp-Singer Test of Phoneme Segmentation.................................................... 22  
  Dynamic Indicators of Basic Early Literacy Skills.......................................... 23  
  Change Scores in Standard Deviation Units...................................................... 24  

Discussion ............................................................................................................... 30  
  Limitations of the Present Study....................................................................... 33  
  Future Research .................................................................................................. 34  

References .............................................................................................................. 35  

Appendix A .............................................................................................................. 38  

Appendix B .............................................................................................................. 40  

Appendix C .............................................................................................................. 42
List of Tables

Table 1: Demographic Characteristics of Sample ........................................21

Table 2: Test-Retest Reliabilities of Phonemic Awareness Instruments ..........24
List of Figures

Figure 1: CTOPP Change Scores in Standard Deviation Units..........................26
Figure 2: TOPA Change Scores in Standard Deviation Units..........................26
Figure 3: Yopp-Singer Change Scores in Standard Deviation Units...............27
Figure 4: DIBELS-Onset Recognition Fluency Change Scores in Standard
Deviation Units.......................................................................................27
Figure 5: DIBELS-Phoneme Segmentation Fluency (1 min.) Change Scores in
Standard Deviation Units.........................................................................28
Figure 6: DIBELS-Phoneme Segmentation Fluency (total) Change Scores in
Standard Deviation Units.........................................................................28
Figure 7: DIBELS-Letter Naming Fluency Change Scored in Standard Deviation
Units........................................................................................................29
As a child develops, one of the most critical factors in future success in school is reading ability. The extent to which a child is phonemically aware can predict how well he or she will be able to read. With this in mind, many instruments exist that test for these phonemic awareness skills. The purpose of this research was to examine the test-retest reliability of four phonemic awareness instruments. Six elementary schools in the Bowling Green City School District participated in the study. A total of 152 students in the kindergarten grades of each school were administered the Comprehensive Test of Phonological Processing, the Test of Phonological Awareness, the Dynamic Indicators of Basic Early Literacy Skills, or the Yopp-Singer Test of Phoneme Segmentation (some students received two instruments) twice, with two weeks between the testings. Reliability coefficients for the two testings were calculated. Strong reliability coefficients were determined for each of the four instruments, ranging from .82 to .94.
Test-Retest Reliability of Phonemic Awareness Assessment
Instruments on Kindergarten Students

Introduction

The ability to read fluently is one of the most critical skills in a child’s development. As children progress through school, they develop this skill at different rates. Many children are often referred because of delays in reading achievement. They are falling behind compared to the progress of their peers, and often this delay affects other academic subjects as well. How is a child supposed to complete a word problem in math or understand a passage from a science book if he or she cannot read?

With this awareness in mind, it only makes sense to try and detect these potential delays and to intervene early. One way to detect potential reading difficulties is to assess a child’s phonemic awareness. Phonemic awareness has emerged as one of the most important indicators of a child’s potential reading ability. With its recent popularity, several phonemic awareness assessment instruments have been developed. As in the development of all new tests, one must be assured that the instrument being used can consistently measure the construct in question (i.e., the test should be reliable). Reliability is a test’s freedom from measurement errors. Methods for estimating reliability vary not only in the procedure but also in the information they offer. The test-
retest method gives information regarding the stability of test scores over time. The constructs being measured are assumed to be stable over time. Therefore, the repeated administration of a test measuring these constructs should yield similar results. Differences in test scores from one test to the next should be due solely to measurement error (Brown, 1976).

Low reading ability is often included in academic referrals to School Psychologists; the instruments involved in the assessment process need to yield reliable results. Phonemic awareness is a strong indicator of reading ability, and therefore instruments measuring it need to have appropriate test-retest reliability.

Many instruments that assess phonemic awareness are available to school personnel. One popular assessment of phonemic awareness is the Yopp-Singer Test of Phoneme Segmentation (Yopp, 1995). The Yopp-Singer measures a child's ability to articulate the separate sounds of a spoken word in order. It is easy to administer, score, and interpret. Two more recent assessments are the Comprehensive Test of Phonological Processing (Wagner et al., 1999) and the Test of Phonological Awareness (Torgesen & Bryant, 1994). A fourth assessment of phonemic awareness, the Dynamic Indicators of Basic Early Literacy Skills (Kaminski & Good, 1996), has also received considerable attention recently.

Due to the popularity and/or the recency of publication of these four phonemic awareness assessments, they were chosen to be studied in a test-retest reliability study. The researcher will evaluate the test-retest reliabilities of different phonemic awareness
instruments used with kindergarten students. Each instrument will be individually administered to a group of kindergarten students, and then re-administered approximately two weeks later. The reliability coefficients will then be determined, and then compared to the test-retest reliability found by the individual test developers, if one was determined.
Literature Review

What is Phonemic Awareness?

Phonemic awareness has recently emerged in the field of literacy as an important skill. Phonemes are the smallest units of speech that correspond to letters of an alphabetic writing system (Adams, Foorman, Lundberg, & Beeler, 1998). The English language is made up of 24 to 36 phonemes that can be combined to form every word (Snider, 1995). Phonemic awareness is the ability to hear and manipulate individual sounds, or phonemes, in a word. Griffith and Olson (1992) describe phonemic awareness as the understanding that language is made up of smaller units that can be examined independent of meaning.

Most youngsters enter kindergarten lacking phonemic awareness; however, many children gain this awareness and can manipulate phonemes in their speech by the end of first grade (Yopp, 1995). By first grade, phonemic awareness skills typically consist of splitting words up into sounds, splitting syllables, deleting sounds from words, substituting sounds, and reversing sounds (Smith, 1998). Poor readers who enter first grade phonemically unaware are likely to remain poor readers at the end of fourth grade, since their slow acquisition of word recognition skill is due in part to their lack of
phonemic awareness (Juel, 1986). Fortunately, phonemic awareness is a skill that can be taught and developed in children as early as their preschool years (Yopp, 1992).

Children can know the names of the letters and their corresponding sounds, but may not have phonemic awareness (Durica, 1998). The reason phonemic awareness can be so difficult is because people do not naturally attend to the sounds of phonemes as they listen to or produce speech. They process the phonemes automatically, attending to the meaning and the utterance as a whole (Adams et al., 1998). This conscious awareness that words are made up of sounds is not necessary to speak and understand speech, but it is necessary in order for children to be able to read and spell in the alphabetic language (Snider, 1995).

Another term that is closely linked with phonemic awareness is phonological awareness. Although some authors use the terms interchangeably, phonological awareness refers to the sound structure of language, examining words in a sentence, syllables in a word, and the ability to manipulate sound units smaller than words. It is the awareness of unconscious rules that govern speech-sound production (Adams et al., 1998). Phonemic awareness is more specific in that it deals strictly with the smallest sound units of the language (Kaminski & Good, 1998). Because phonemic awareness is included in the area of phonological awareness, both tests of phonemic awareness and phonological awareness were used as part of this study.
The Importance of Phonemic Awareness

The reason phonemic awareness is such a popular concept is due to the consistent finding that there is a strong relationship between phonemic awareness and reading ability (MacDonald & Cornwall, 1995; Snider, 1997; Stanovich, 1986). Ball and Blachman (1991) report the earliest studies regarding the relationship between phonemic awareness and reading ability came from two Russian psychologists, L.Y. Zhurova and D.B. Elkonin in 1963. Their work showed a relationship between phoneme segmentation abilities and subsequent growth in reading ability.

Snider (1997) examined the relationship between phonemic awareness and reading achievement in the primary grades. She individually administered the **Test of Phonemic Awareness** to 73 kindergartners in late April. Data were obtained from those same students in second grade using a standardized reading achievement test. Significant correlations of .34 for word analysis and .33 for reading comprehension were found between performance on phonemic awareness tasks and later reading achievement. A three-year follow-up study was done on 12 of the subjects who scored in the lowest quartile of the first kindergarten testing. These subjects were retested with the **Test of Phonemic Awareness** and given a reading passage. She found that only three of these students read at a fluent rate.

MacDonald and Cornwall (1995) did an eleven year follow-up study on 24 students who had participated in a study of phonological analysis and reading and spelling abilities eleven years earlier while in kindergarten. At the beginning of the study,
58 kindergarten students were randomly chosen from city schools in Nova Scotia for participation in the study. They were given a variety of measures, including the Auditory Analysis Test, the Peabody Picture Vocabulary Test, and the Reading and Spelling subtests of the Wide Range Achievement Test. In 1993, the researchers were able to contact 37 of the 58 participants and found 24 of the participants willing to participate in the follow-up study. They were administered the same measures as previously, and were also administered the Word Attack and Passage Comprehension subtests from the Woodcock Reading Mastery Tests-Revised. The results indicated that the phonological awareness assessed during kindergarten was a significant predictor of word identification and spelling skills eleven years later.

In a review of the research, Stanovich (1986) concluded that phonemic awareness is a more potent predictor of reading achievement than nonverbal intelligence, vocabulary, and listening comprehension, and that it often correlates more highly with reading acquisition than tests of general intelligence or reading readiness. "Most importantly, phonemic awareness tasks are the best predictors of the ease of early reading acquisition-better than anything else that we know of, including IQ" (Stanovich, 1994, p. 284).

**Phonemic Awareness Tasks**

Early researchers assessed phonemic awareness in children by having them tap out the sounds they heard in words (Liberman, Shankweiler, Fischer, & Carter, 1974).

Currently, evidence shows that phonemic awareness is not a unitary ability; instead it is a
general construct that consists of several dimensions. These dimensions can be assessed with a variety of tasks (Yopp, 1988). Some examples of phonological awareness tasks include: phoneme deletion, word to word matching, blending, sound isolation, phoneme segmentation, phoneme counting, deleted phoneme, odd word out, and sound to word matching (Stanovich, 1994). Adams (1990) arranged some of the phonemic awareness tasks, in order from easiest to hardest, as follows: rhyme (recognizing pairs of rhyming words or producing words that rhyme), sound oddity (identifying words that are the same or different in terms of beginning, middle, or ending sounds), blending (identifying a word when each syllable of a phoneme is pronounced separately), phoneme segmentation (pronouncing each separate phoneme in a one-syllable word), and phoneme manipulation (identifying the word left when phonemes are added, deleted, or moved). Most researchers discuss phoneme blending, substitution, and segmentation as being the most significant phonemic abilities in relation to future success in reading, and eventually spelling (Durica, 1998). Phoneme blending is the skill required to read multisyllabic words. Phoneme substitution is the ability to replace one phoneme for another, such as substituting the letters b, l, and t for the letter c in cook. Phoneme segmentation is the ability to isolate sequentially the sounds in a word and allows a reader to successfully decode words.

Ball and Blachman (1991) evaluated the effect training in phonemic awareness would have on kindergarten students’ early word recognition. Ninety kindergarten students were divided into three groups. The first group received training in phoneme
segmentation and correspondences between letter names and letter sounds. The second group received training in letter names and letter sounds only. The third group received no training. Results showed that the group with phonemic awareness instruction, combined with correspondence between letter names and letter sounds, significantly improved in early reading and spelling skills, more so than the other two groups.

Yopp (1988) examined the reliability and validity of several phonemic awareness tasks, including: a phoneme blending task, a phoneme counting task, two phoneme deletion tasks, a rhyming task, a sound isolation task, a word-to-word matching task, a phoneme reversal task, and two phoneme segmentation tasks. A learning test was also given in order to determine the predictive validity of each of the phonemic awareness measures, assessing each child's ability to use sound-symbol matches in order to decode printed artificial words. She calculated the highest Cronbach alpha reliabilities, over .90, for the phoneme blending task and one of the phoneme segmentation tasks. The predictive validity was determined by comparing the tasks to the criterion learning test, resulting in high predictive validity for four of the tasks: the modification of the sound isolation task, the two phoneme segmentation tasks, and the phoneme deletion task. These four tasks correlated the highest with the learning test.

**Standardized Phonemic Awareness Instruments**

Research on phonemic awareness tasks has led to the publication of numerous standardized phonemic awareness instruments. These instruments may be administered by a variety of school personnel, including: audiologists, speech and language
Four tests of phonemic awareness will be reviewed. The tests were selected due to their apparent popularity throughout the literature (i.e., Test of Phonological Awareness [Torgesen & Bryant, 1994]; Yopp-Singer Test of Phoneme Segmentation [Yopp, 1995]) or due to their recency of production (i.e., Comprehensive Test of Phonological Processing [Wagner, Torgesen, & Rashotte, 1999]; Dynamic Indicators of Basic Early Literacy Skills [Kaminski & Good, 1996]).

**Yopp-Singer Test of Phoneme Segmentation.** One test of phonemic awareness is the Yopp-Singer Test of Phoneme Segmentation, which measures a child's ability to articulate (segment) the individual sounds of a spoken word in order (Yopp, 1995). As an example, the child is given the word "old" and asked to separate the sounds, (not the letters), in the word. The answer would be /o/-/l/-/d/. The brief test, consisting of 22 items (words), is administered on an individual basis and requires approximately five to ten minutes per child. Students who segment all or nearly all of the words correctly are considered phonemically aware. Students who correctly segment some items are showing signs of emerging phonemic awareness. Students who are able to segment only a few items or none at all lack appropriate levels of phonemic awareness (Yopp, 1995). While an internal consistency reliability (Cronbach Alpha) for the Yopp-Singer was reported, no test-retest reliability information was reported.

**Comprehensive Test of Phonological Processing.** A recent test of phonological awareness is the Comprehensive Test of Phonological Processing (CTOPP) (Wagner et
This assessment instrument includes 13 subtests, yielding three composite areas: Phonological Awareness, Phonological Memory, and Rapid Naming. The testing time required to administer the entire CTOPP is approximately 30 minutes.

In order to determine the test-retest reliability of the CTOPP, 91 residents of Tallahassee, Florida, were administered the test. Out of this sample, only 32 were ages 5 through 7 in kindergarten and first grade at a local elementary school. The demographics of these participants were unavailable. The participants were tested twice, with a two week period between testings. The test-retest reliability coefficient for the children ages 5-7 for the Phonological Awareness composite scale was .79.

**Dynamic Indicators of Basic Early Literacy Skills.** An alternative measure to standard phonemic awareness tests is the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) (Kaminski & Good, 1998). It was developed by a team of researchers at the University of Oregon to make educational decisions regarding identification of children requiring early literacy skills intervention and to determine which interventions would be most effective. Because the DIBELS provides brief alternate forms, one is able to monitor the progress of a child’s literacy skills over time. The methods of assessment used in the DIBELS measures do not suggest the methods of instruction, but were developed to be indicators of key skills (Good & Kaminski, 1996). The measures are brief and each lasts approximately one minute. They are intended to provide a quick, reliable, and valid measure of key indicators of early literacy. Three DIBELS measures or
activities are available for use with kindergarten students: Phoneme Segmentation Fluency, Onset Recognition Fluency, and Letter Naming Fluency.

Phoneme Segmentation Fluency (PSF) is a measure of phonological awareness that is used to determine children's ability to segment orally presented words into phonemes. The authors recommend this task for children in the winter of kindergarten through the fall of first grade. Using a one-minute sample, a PSF probe consists of 24 words and reportedly takes approximately three minutes to administer and score (Good, Simmons, & Smith, 1998). While this task is very similar to the Yopp-Singer Test of Phoneme Segmentation, the scoring is much different. Using the Yopp-Singer, the child receives one point for correctly segmenting the entire word. With the Phoneme Segmentation Fluency measure, each correct segment of a word is worth one point. The PSF is individually administered by classroom teachers, teaching assistants/aids, school volunteers, school psychologists, and/or other related services personnel (Kaminski & Good, 1998).

Another DIBELS measure is the Onset Recognition Fluency task, intended to measure phonological awareness skills of children in the spring of preschool through winter of kindergarten. This task contains 16 items. The child is presented with four pages of pictures, each with four pictures of objects. On 12 of the items, the child is asked to point to the picture that begins with a target sound. The other 4 items require the child to give the initial sound of a picture. Twenty alternate forms are available.
A third DIBELS measure is Letter Naming Fluency. The authors of the DIBELS report that there has been a consistent finding of a strong positive relationship between accuracy of letter naming and later reading achievement. This one-minute timed task requires the child to name the randomly typed letters presented on a probe. This measure is intended for children in the fall of kindergarten through the fall of first grade.

Test-retest reliability for the DIBELS was reported, however, it was not measured in the traditional manner. In order to determine the test-retest reliability for the DIBELS, 37 kindergarten students and 41 first grade students from the Pacific Northwest were administered the test. Each cohort was divided randomly into two groups, a monitored and non-monitored group. The students in the monitored group were administered three DIBELS measures two times a week for a period of 9 weeks while the students in the non-monitored group were tested with the measures only at the beginning and at the end of the 9-week period. The three DIBELS measures used were: Phoneme Segmentation Fluency, Letter Naming Fluency, and Picture Naming Fluency. Only two of the three DIBELS tasks discussed in this review were included (Picture Naming Fluency is no longer included in the DIBELS.) No information for the Onset Recognition Fluency was provided. The reliability of the repeated measures was estimated using the average of the alternate forms coefficients for each measure. The average coefficient of stability for Letter Naming was .93. The average coefficient of stability for Phoneme Segmentation was .88 (Kaminski & Good, 1996).
**Test of Phonological Awareness.** The Test of Phonological Awareness (TOPA) (Torgesen & Bryant, 1994) was designed to identify the level of phonological awareness of children in kindergarten or early elementary grades. It measures a young child’s ability to isolate individual phonemes in spoken words, manipulating the initial sounds. In this manner, the instrument assesses initial phonemes. It can be group-administered, therefore being a quick and easy way to identify those children who are developmentally delayed in phonological awareness. Testing time is reported in the TOPA manual to usually require 15-20 minutes.

The TOPA consists of two forms: TOPA-Kindergarten and TOPA-Early Elementary. In order to assess the test-retest reliability of the TOPA-Kindergarten scale, the test was administered to 40 kindergarten students in Tallahassee, Florida. The test was re-administered 6 weeks later. The students ranged in age from 67 to 84 months. No other demographic information was provided. The correlation between the first and second testing was found to be .84.
Purpose

An evaluation of the literature revealed a need for more information on the test-retest reliability of phonemic awareness instruments. The tests reviewed either did not report test-retest reliabilities, used small, limited samples, or used nontraditional techniques to assess the test-retest reliability.

The purpose of this study was to evaluate the test-retest reliability of four phonemic awareness and phonological awareness measures with a sample of kindergarten students. The four assessment instruments were chosen because of current popularity and/or recency of production. For the two tests that did report test-retest reliabilities, it is questioned whether a different sample of kindergartners will yield the same results. Neither the Test of Phonological Awareness nor the Comprehensive Test of Phonological Processing provided demographic information of the subjects, leaving one to question if the same results would occur if repeated on a different sample. It is unclear as to whether a test-retest analysis was completed on the Yopp-Singer Test of Phoneme Segmentation. For purposes of this study, the DIBELS-PSF task was administered using a one-minute timed score and a total score. The reason for following this procedure was to insure that the child was given ample time and items to attempt segmentation.
It is hypothesized that the TOPA will have higher correlations than the test developers’ study due the TOPA being given individually instead of group-administered. Also, it would be reasonable to predict that a higher correlation will occur with more children being used in the sample than the original study and a shorter time interval between testings. Concerning the CTOPP, it is predicted that a similar correlation to the test developers’ study will be found because the test developers and the current study both use the same 2-week interval between testings. Even though there is no retest data available for the Yopp-Singer, it is predicted that a strong correlation will exist between the two testings due to the skill being assessed by the measure. Regarding the Phoneme Segmentation Fluency of the DIBELS measure, the same hypothesis exists as the Yopp-Singer, due to the skill being assessed. With correlations being determined for the task administered for one minute and for the entire task, it would be expected that a higher correlation would be found for the entire task. One would assume a high correlation for the Onset Recognition Fluency task of the DIBELS due to its being a fairly simple task. The Letter Naming Fluency task of the DIBELS measure is predicted to have a slightly lower correlation between the testings. Although the task involves the child naming the letters for only one minute, when a time element is added, correlations will often decrease.
Method

Participants

Phonemic awareness measures are usually intended for use with children who are not yet reading. Therefore, kindergarten students were used as participants in the study. Approval for the study was obtained from the Western Kentucky University Human Subjects Review Board (see Appendix A) and from the Bowling Green (Kentucky) City School District (see Appendix B). A letter of consent (see Appendix C) was sent home to the parent/guardian of every kindergarten student in the district, which consisted of 261 students. Due to the young age of the participants, no assent form was signed by them. Students who returned the signed consent form, regardless of whether their parents allowed them to participate in the study, received a pencil as a reward. Consent to participate in the study was obtained for 165 students (63%) from six elementary schools in the district. Over the two-week period, between the first and second testing sessions, 13 students were dropped from the study due to illnesses or families moving, leaving 152 students (58%) participating in the study. The children ranged in age from 5 years, 3 months to 7 years, 0 months. Of the 152 participants, 83 (55%) were male and 69 (45%) were female. One hundred eleven children were Caucasian (73%), 26 were African-American (17%), 6 were Hispanic (4%), 5 were Asian (3%), and 4 were placed in the
Other category (3%), which included recent immigrants from European and Middle Eastern countries.

Materials

The four tests administered included the Test of Phonological Awareness-Kindergarten Version (Torgesen & Bryant, 1994), the Yopp-Singer Test of Phoneme Segmentation (Yopp, 1995), the Comprehensive Test of Phonological Processing (Phonological Awareness Composite)-Five- and Six-Year-Olds version (Wagner et al., 1999), and the Dynamic Indicators of Basic Early Literacy Skills (Onset Recognition Fluency, Phoneme Segmentation Fluency, and Letter Naming Fluency subtests). Because the DIBELS instrument has 20 alternate forms for each subtest, one alternate form was randomly chosen. Mini-tape recorders were used to record a random number of administrations in order to test for inter-rater reliability.

Procedure

The participants were randomly divided into three groups. Each participant in Group 1 was individually administered the CTOPP. Each participant in Group 2 was individually administered the TOPA. The participants in Group 3 were each individually administered the Yopp-Singer and the DIBELS. The total testing time took approximately 8-12 minutes per child. The testing was done in libraries and workrooms of the schools. The instruments involved the participants responding orally to items or marking a box corresponding to their answer. To minimize inconsistencies due to differences in the examiners, retest administrations were conducted on the same
participants by the same examiners approximately two weeks later. The two-week time interval was considered long enough to lessen the chances of a practice effect occurring and short enough to minimize possible learning that may take place during the interval. Four examiners took part in the study, including two undergraduate students who were trained in administering these measures. Once consent was given, a master list was created which included all of the participants’ names and a corresponding code number. Only the code numbers were written on the instrument protocols to ensure participant confidentiality.
Results

Percentages of inter-rater reliability were calculated for all three of the DIBELS measures and for the Yopp-Singer Test of Phoneme Segmentation. Some of the administrations were tape-recorded and a sample of those were selected to be re-scored by the author for comparison. As a result, 20.6% of all DIBELS and Yopp-Singer test administrations were reviewed. The DIBELS-Onset Fluency task had an inter-rater agreement of 97%. The DIBELS-Phoneme Segmentation Fluency task had an inter-rater agreement of 69%. A 100% agreement was found for the DIBELS-Letter Naming Fluency task and a 90% agreement was found for the Yopp-Singer Test of Phoneme Segmentation. According to Alessi and Kaye (1983), an inter-rater reliability coefficient greater than .80 is acceptable. Each of these measures met this criteria except for the DIBELS-Phoneme Segmentation Fluency task. The low inter-rater reliability coefficient for this measure is addressed later as a limitation of the study.

Comprehensive Test of Phonological Processing

Fifty-two participants made up Group 1, receiving the Comprehensive Test of Phonological Processing (CTOPP). The demographics of the participants in this group are listed in Table 1. The ages of the participants in this group ranged from 63 months to 84 months, with an average age of 69.65 months. The participants were individually
administered the CTOPP during one session, then approximately two weeks later were administered the CTOPP again. The average number of days between testing and retesting was 15.6.

The three subtest raw scores of the CTOPP were added together to make a total raw score for each participant. The raw scores from the first test were correlated with the raw scores of the second testing. The coefficient of stability for the CTOPP-Phonological Awareness Composite for the Five- and Six-Year-Olds Version was .90 (see Table 2).

Table 1

Demographic Characteristics of Sample

<table>
<thead>
<tr>
<th></th>
<th>CTOPP</th>
<th>TOPA</th>
<th>Yopp-Singer/DIBELS</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Males</td>
<td>30 (57.7)</td>
<td>32 (60.4)</td>
<td>21 (44.7)</td>
<td>83 (54.6)</td>
</tr>
<tr>
<td>Females</td>
<td>22 (42.3)</td>
<td>21 (39.6)</td>
<td>26 (55.3)</td>
<td>69 (45.4)</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>40 (76.9)</td>
<td>34 (64.2)</td>
<td>37 (78.7)</td>
<td>111 (73.0)</td>
</tr>
<tr>
<td>Black</td>
<td>4 (7.7)</td>
<td>15 (28.3)</td>
<td>7 (14.9)</td>
<td>26 (17.1)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>2 (3.8)</td>
<td>2 (3.8)</td>
<td>2 (4.3)</td>
<td>6 (3.9)</td>
</tr>
<tr>
<td>Asian</td>
<td>4 (7.7)</td>
<td>1 (1.9)</td>
<td>0 (0.0)</td>
<td>5 (3.3)</td>
</tr>
<tr>
<td>Other</td>
<td>2 (3.8)</td>
<td>1 (1.9)</td>
<td>1 (2.1)</td>
<td>4 (2.6)</td>
</tr>
</tbody>
</table>
Test of Phonological Awareness

The participants in Group 2 were each administered the Test of Phonological Awareness (TOPA). The group consisted of 53 participants. The demographics of the participants in this group are listed in Table 1. The ages of the participants ranged from 63 months to 76 months, with 69.23 months being the average age. Each participant individually received the TOPA, then approximately two weeks later received the same measure. The average number of days between the two testings was 15.7.

The raw scores for the first test were correlated with the second test raw scores. The coefficient of stability for the TOPA-Kindergarten Version was .88, significant at the 0.01 level (see Table 2).

Yopp-Singer Test of Phoneme Segmentation

Group 3 consisted of 47 participants receiving Yopp-Singer Test of Phoneme Segmentation (Yopp-Singer). The demographics of the participants in this group are listed in Table 1. The ages of the participants in this group ranged from 64 months to 76 months, with an average age of 69.28 months. The Yopp-Singer was individually administered to each participant. It was re-administered approximately two weeks later. The average number of days between testings was 15.3.

The raw scores for the first session were correlated with the raw scores for the second session. The coefficient of stability for the Yopp-Singer was .94, significant at the 0.01 level (see Table 2).
Dynamic Indicators of Basic Early Literacy Skills

Along with the Yopp-Singer, the participants in Group 3 were also administered the Dynamic Indicators of Basic Early Literacy Skills (DIBELS). The demographics of the participants of this group are listed in Table 1. The ages of the participants in this group ranged from 64 months to 76 months, with an average age of 69.28 months. Each participant was individually administered the three DIBELS measures, which were Onset Recognition Fluency, Phoneme Segmentation Fluency, and Letter Naming Fluency. These measures were re-administered approximately two weeks later, with an average of 15.3 days between testings.

The scores for each measure for the first session were correlated with the scores of the second session, giving a correlation coefficient for each measure. One of the measures, Phoneme Segmentation Fluency, was correlated using the total raw score and a raw score for only a one-minute administration. The coefficient of stability for the Phoneme Segmentation Fluency measure for a one-minute administration was .82, significant at the 0.01 level (see Table 2). The coefficient of stability for the Phoneme Segmentation Fluency measure for the entire administration was .92, significant at the 0.01 level (see Table 2). The coefficient of stability for the Onset Recognition Fluency measure was .85, significant at the 0.01 level (see Table 2). The coefficient of stability for the Letter Naming Fluency measure was .88, significant at the 0.01 level (see Table 2).
Table 2

**Test-Retest Reliabilities of Phonemic Awareness Instruments**

<table>
<thead>
<tr>
<th></th>
<th>Pearson r</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTOPP</td>
<td>.90*</td>
</tr>
<tr>
<td>TOPA</td>
<td>.88*</td>
</tr>
<tr>
<td>DIBELS</td>
<td></td>
</tr>
<tr>
<td>ORF</td>
<td>.85*</td>
</tr>
<tr>
<td>PSF- 1 Min.</td>
<td>.82*</td>
</tr>
<tr>
<td>PSF-Total</td>
<td>.92*</td>
</tr>
<tr>
<td>LNF</td>
<td>.88*</td>
</tr>
<tr>
<td>Yopp-Singer</td>
<td>.94*</td>
</tr>
</tbody>
</table>

*Note. ORF = Onset Recognition Fluency, PSF = Phoneme Segmentation Fluency, LNF = Letter Naming Fluency.

*p < .01

**Change Scores in Standard Deviation Units**

The number of points each participant’s raw score increased or decreased from test one to test two was also calculated and reported in standard deviation units. The differences were charted as frequencies. The frequencies are indicated for each instrument used in the study, found in Figures 1 through 7. It appears that the TOPA, Yopp-Singer, and DIBELS-ORF remained fairly consistent from one test to the next.
However, the DIBELS-PSF and the DIBELS-LNF had significant point differences between the two testings.
Figure 1. CTOPP change scores in standard deviation units.

Figure 2. TOPA change scores in standard deviation units.
Figure 3. Yopp-Singer change scores in standard deviation units.

Figure 4. DIBELS-Onset Recognition Fluency change scores in standard deviation units.
Figure 5. DIBELS-Phoneme Segmentation Fluency (1 min.) change scores in standard deviation units.

Figure 6. DIBELS-Phoneme Segmentation Fluency (total) change scores in standard deviation units.
Figure 7. DIBELS-Letter Naming Fluency change scores in standard deviation units.
Discussion

The results of this study indicated favorable test-retest reliability correlations for all the phonemic awareness instruments under consideration. For the two instruments (TOPA and CTOPP) that had reported test-retest reliabilities based on small samples of students from Florida, the current study revealed even higher correlation coefficients than reported in the test manuals. A higher correlation for the TOPA was expected due to the current researcher’s use of a larger sample and a shorter time interval between testings. Also, a higher correlation was expected because in the current study, the TOPA was individually administered, whereas the test developers used group administration. A similar test-retest correlation as the test developers’ finding for the CTOPP was expected due to the same two-week interval being used. While the CTOPP used kindergartners and first-graders in the standardization sample, the current study found a higher correlation (.90 versus .79) using only kindergarten students in the sample. Thus, the publishers of the TOPA and the CTOPP would view the current results very favorably. The test-retest correlation coefficients were also high for all of the subtests of the DIBELS and for the Yopp-Singer, two instruments for which no test-retest analysis was found. Due to the skills being assessed with these instruments (phoneme segmentation, recognition of initial sounds) high correlations were expected and found. These are skills
that kindergartners seem to either have or not have; therefore, tests given two weeks apart would result in high correlations. A slightly lower correlation was expected for the DIBELS-Letter Naming Fluency due to the task being a 1-minute timed sample; however, a strong correlation was found. The test-retest reliability correlation for the 1-minute Phoneme Segmentation Fluency task was found to be slightly lower than for the entire Phoneme Segmentation Fluency task, which was expected. The current results are very favorable for these two previously unexamined phonemic awareness instruments.

To further evaluate test-retest characteristics of the test instruments, the raw point differences between testings were also examined. The frequency of each point difference for each instrument was graphed in standard deviation units in order to provide an illustration of the consistency between the two testings. The TOPA and the Yopp-Singer appeared to have the highest percentage of students with no change between the two testings. An explanation for this outcome may be due to the limited sample of skills assessed. The Yopp-Singer assesses only phoneme segmentation, which is a difficult skill for kindergartners. Many students received scores of zero on both test administrations. Thus, although the Yopp-Singer may have a high test-retest reliability, it does not appear to be a useful measure at the kindergarten level. The TOPA only evaluates whether the student can match initial sound phonemes. Whether the measurement of a single skill (i.e., initial sound matching) is predictive of reading ability remains to be seen.
For the CTOPP, more than two-thirds of the students had an increased score on the second testing. The CTOPP, while a standardized instrument, is comprised of three different tasks. Changing the tasks throughout the administration may have affected the students' performance. Since many students had increased scores, the exposure to the instrument once may have affected the child's performance the second time. The instructions and the tasks were novel and, perhaps, unclear the first time, but exposure to these types of tasks may have given the participants a better understanding of what they were supposed to do.

Unlike the Yopp-Singer phoneme segmentation test, the DIBELS-Phoneme Segmentation task did show differences from test 1 to test 2. The largest percentage of children on this task had a score increase of 1.01 to 1.50 standard deviations between testings, even though it was essentially the same task as the Yopp-Singer Test of Phoneme Segmentation. The difference is likely due to scoring procedures of the two tests. The DIBELS-Phoneme Segmentation task is scored by giving credit for each phoneme correctly segmented, whereas the Yopp-Singer is scored by giving credit only if the entire word is segmented correctly. Thus, it seems the DIBELS’s scoring method for the phoneme segmentation task would provide a more sensitive measure of a child’s skills. Unfortunately, as noted by a low inter-rater scoring agreement (.69), scoring each individual phoneme is not an easy task.

The DIBELS-Letter Naming task also had many students with an increased score. Because this task involved simply naming letters that were randomly presented on a
probe, perhaps an increase in points was due to learning of new letters between the two testings.

The change scores in standard deviation units for the DIBELS-Onset Fluency showed most children receiving the same or similar scores for both testings. Again, this outcome may be due to the consistency of the task at hand. The children appeared to have an understanding of what they were required to do, and assessing whether or not they knew the answer was clear. Therefore, the same or similar scores were yielded for each testing. Similar scores may also be due to the simplicity of the task. Many students received high scores on both test administrations. Thus, this task may not be a valid measure for mid-year kindergartners.

The current study provides support for the test-retest reliability of the instruments being used to assess phonological and phonemic awareness. Because these pre-reading skills are so critical to a child’s emerging literacy, it is important that these skills are assessed accurately. It is the author’s conclusion that these instruments will provide stable estimates of a child’s pre-reading skills. However, the issue of which instrument is the best predictor of a child’s reading skills remains to be evaluated.

Limitations of the Study

While the demographic sample of the study represented a fair amount of diversity, only one school district in the state of Kentucky was used. Perhaps using a school district in a different geographic region would provide different results. Another limitation of the current study is the relatively low inter-rater reliability for the DIBELS-Phoneme
Segmentation task. Additional training over administration of this task could have increased the agreement at which the examiners scored the students’ responses.

Future Research

Research conducted in the future could examine the same or additional instruments in different geographic regions and with more diverse ethnic and SES samples. Different samples may produce somewhat different reliability coefficients. As noted, some instruments measure only single skills while others are perceived as too hard or too easy for mid-year kindergartners. These observations, however, may be irrelevant to a test’s predictive validity. Future research needs to examine which task or combination of tasks best predicts a child’s later reading abilities. Future research could also focus more on effective methods of teaching phonemic awareness and those skills that should be taught in kindergarten classrooms. Results of this area of research could aid teachers in knowing the necessary pre-reading skills children should have and how teachers can implement those skills in their classrooms.
References


APPENDIX A

Letter of Approval from Human Subjects Review Board

at Western Kentucky University
Dear Ms. Craycroft:

1. Your research project "Test-Retest Reliability of Phonemic Awareness Assessment Instruments on Kindergarten Students," has undergone review by the Western Kentucky University IRB for human subjects of research and it has been determined that risks to subjects are: (1) minimized and reasonable; and that (2) research procedures are consistent with a sound research design and do not expose the subjects to unnecessary risk. Reviewers determined that: (1) benefits to subjects are considered along with the importance of the topic and that outcomes are reasonable; (2) selection of subjects is equitable; and (3) the purposes of the research and the research setting is amenable to subjects' welfare and producing desired outcomes; that indications of coercion or prejudice are absent, and that participation is clearly voluntary.

2. In addition, the IRB found that: (1) informed consent will be sought and documented from each prospective subject. (2) Provision is made for collecting, using and storing data in a manner that protects the safety and privacy of the subjects and the confidentiality of the data. (3) Appropriate safeguards are included to protect the rights and welfare of the subjects. Please store all data securely at an on campus location for a minimum of three years after the project is completed.

3. Your research therefore meets the criteria of Full Board Review and is approved. Please note that the institution is not responsible for any actions regarding this protocol before approval. Copies of your request for human subjects review, your application, and this approval, are maintained in the Office Sponsored Programs at the above address. Please report any changes to this approved protocol to this office. A Continuing Review protocol will be sent to you in the future to determine the status of the project.

Kindest regards.

Sincerely,

Phillip E. Myers, Ph.D.
Director, Office of Sponsored Programs and
Human Subjects Coordinator

c: Human Subjects File0020

HSApprovalCraycroft0020Rev
APPENDIX B

Letter of Approval from

Bowling Green Independent School District
November 17, 1999

Lorie Craycroft
1957 Stonehenge Avenue
Apt. D
Bowling Green, KY 42101

Dear Ms. Craycroft:

The members of the Bowling Green Board of Education approved your request to utilize kindergarten students in a phonemic awareness test-retest reliability study at the regular November Board meeting.

If I may be of further assistance, please do not hesitate to contact me.

Sincerely,

Dr. John C. Settle
Superintendent

JC:rc
APPENDIX C

Letter and Consent Form Sent

to Parents
Dear Parents:

Your child is being asked to participate in a study about children's reading abilities. This study is being conducted by Lorie Craycroft and Dr. Carl Myers of Western Kentucky University. The aim of our study is to get a better understanding of how to test children's early reading abilities. The study will be conducted in two short sessions (at your child's school) in cooperation with your child's teacher so that your child does not miss important learning activities.

Each session will take approximately 20 minutes. Your child will individually be given two tasks that deal with separating and combining letters and sounds. Your child's responses may be tape recorded but will be kept confidential. Your child is free to discuss this activity with you. The second session, taking place approximately two weeks later, will be conducted the exact same way.

We emphasize that your child's participation in this project is entirely voluntary. If you or your child decide not to participate, it will have no negative outcome for you or your child in any way. Your child may refuse to respond to any of the items and may withdraw from the study at any time. All information collected in this study will be kept strictly confidential and is accessible only to the project staff and your child's school. Data will be identified with a code number, not your child's name.

The procedures in this study have been reviewed and approved by the Western Kentucky University Committee for the Protection of Human Research Participants. Any questions about this study may be directed to Lorie Craycroft at 745-2695 or Dr. Carl Myers at 745-4410. We urge you to call us if you have any questions.

We hope that you will allow your child to take part in our study. We promise to make it a pleasant experience for your child and to schedule our sessions in cooperation with your child's teacher. Please fill in your child's name, your child's date of birth, and your child's teacher's name on the attached form. To indicate your consent, check the "yes" box, sign your name, and fill in the date. When your child returns this letter to the teacher, whether you check yes or no, your child will receive a small reward.

Thank you for your help.

Carl Myers, Ph.D.
Assistant Professor of Psychology

Lorie A. Craycroft
School Psychology Graduate Student
WESTERN KENTUCKY UNIVERSITY

PARTICIPANT CONSENT FORM

Child’s name: ___________________________ Date of birth: __________________

Teacher’s name: ____________________________

____ No, I do not give my consent for my child to participate in this study.

____ Yes. I have read the information provided about this study, and give my consent for my child to participate in the study conducted by Lorie Craycroft and Dr. Carl Myers of Western Kentucky University. I understand that I may withdraw my child from the study at any time without penalty.

Parent/Guardian signature: ___________________________ Date: ______________

* Please return this form by Thursday, December 16.

When this form is returned, whether it is checked yes or no, your child will receive a small reward.