# Validity of DIBELS Early Literacy Measures with Korean English Learners

**Michael L. Vanderwood · Jeanie E. Nam ·
Jennifer W. Sun**

**Abstract** The purpose of this study was to examine the validity of two Dynamic Indicators of Basic Early Literacy Skills (DIBELS) early literacy measures with first-grade Korean speaking English Learners (ELs). A total of 30 first-grade Korean ELs were screened three times during the year using early literacy measures from DIBELS. A sample of students was provided an empirically supported phonological awareness (PA) intervention to determine whether DIBELS measures were sensitive to growth caused by an intervention. Results suggest screening measures of phoneme segmentation fluency (phoneme segmentation fluency, PSF) were not predictive of reading performance at the middle and end of first grade. Measures of alphabetics (nonsense word fluency, NWF) were predictive of reading skills for this sample of first-grade Korean ELs. PSF and NWF were sensitive to change caused by an intervention, and appear to be able to be used to monitor the impact of a PA intervention. Implications and recommendations for the use of DIBELS with Korean speaking ELs are provided.

**Keywords** Early literacy · English learners · DIBELS · Phonological awareness

## Validity of DIBELS Early Literacy Measures with Korean English Language Learners

The population of English learners (ELs) has grown dramatically over the last decade, representing approximately 5.1 million students, or 10 % of the total public school enrollment (National Clearinghouse for English Language Acquisition 2007). ELs represent more than 350 native languages and come from diverse cultures and educational backgrounds (Hopstock and Stephenson 2003). According to data from 2000 to 2001, Kindler (2002) reported that Spanish and Asian languages were the most common languages spoken by ELs, with 79 % of students claiming Spanish as their native language, followed by Vietnamese (2.0 %), Hmong (1.6 %), Chinese (1.0 %), Korean (1.0 %), and other languages (15.4 %). Recent data from the National Center for Education Statistics (NCES 2009) suggests there are over 165,000 U.S. K-12 students who speak Korean at home. California currently has the largest population of Koreans in the U.S., with approximately one third of U.S. Koreans residing in this state (Korean American Coalition-Census Information Center 2003). Yet, there is very little research that provides guidance on the best way to prevent reading problems for this group. The purpose of this study is to evaluate early literacy assessment tools with a Korean speaking EL population to provide evidence regarding whether these tools should be used for the purposes of screening and progress monitoring.

A critical factor for academic success is a solid foundation in literacy skills in the early grades, but learning to read can be a challenging process for many students. Acquiring the skills and knowledge to read in a second language can be especially difficult for ELs (Snow et al. 1998). Past studies have found that many ELs are failing to meet state education standards (Kindler 2002), with 76 % of third-grade ELs performing below grade level in English reading (Zehler et al. 2003).

Given the importance of literacy skills, several authors suggest that school districts use a response to intervention (RtI) approach with ELs to more accurately identify and provide services to students in need of support (Klingner and Edwards 2006; Vanderwood and Nam 2008). Although there is a substantial amount of research supporting the use of an RtI approach when addressing reading difficulties with monolingual students (e.g., Vellutino et al. 1998; Wagner

M. L. Vanderwood (✉) · J. E. Nam · J. W. Sun
University of California-Riverside, Riverside, CA, USA
e-mail: mike.vanderwood@ucr.edu

et al. 1993), there is significantly less evidence regarding the effectiveness of this approach with ELs (Vanderwood and Nam 2008).

### Evaluating Screening and Progress Monitoring Measures

An important starting point in determining the appropriateness of using RtI with ELs is to ensure that the assessment tools used in the process are valid. As outlined in the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education [AERA, APA, and NCME] 1999), for constructs that could be influenced by language (e.g., cognitive ability and academic skills), it is not appropriate to use assessment tools with ELs based on studies that only included native English speakers. Using tests that have not been validated with ELs could lead to inaccurate or psychometrically biased conclusions. Therefore, it is necessary to provide separate validity evidence for groups in which certain identifying aspects (e.g., English language proficiency and home language) could possibly affect the reliability and validity of a measure. In addition, before assessments are used to make important decisions within an RtI approach, the tools should: (a) be psychometrically sound, (b) have the capacity to show growth over time, (c) be sensitive to change during the intervention, (d) be independent of specific instructional techniques, (e) have the capacity to inform teaching, and (f) be feasible to allow for frequent data collection on a large number of students (Fuchs and Fuchs 1999).

One major challenge with the RtI approach has been the development of screening decision rules for identifying students who are most in need of tier 2 services. In order to determine whether a student may be at-risk and in need of secondary-level reading intervention, educators using an RtI approach have typically relied on measuring skills related to literacy development and have applied a specific cutoff to classify these students. Recent reviews of EL literacy research have found that many of the same early literacy screening tools and predictors (e.g., phonological awareness, letter knowledge, alphabetic principle, and reading fluency) that are used with native English speakers can be used to screen for reading difficulties with ELs (Gersten et al. 2007). Unfortunately, the recommendations by Gersten and colleagues were mostly based on research that treated ELs as a homogenous group and did not examine whether there were differences in decision rules for different primary languages (e.g., Korean). Yet, there is substantial reason to believe the constructs to be assessed will be similar across primary languages for all students learning to read English (Vanderwood and Nam 2008).

*Phonological Awareness (PA)* PA is a critical reading skill across all languages and is considered to be the understanding that spoken words and syllables can be divided into smaller components (e.g., Stanovich et al. 1984). Cho and McBride-Chang (2005) found Korean measures of PA were uniquely associated with word recognition for beginning readers of Korean. Similarly, studies found Spanish measures of phonological awareness were predictive of Spanish pseudoword and real word reading (Durgunoglu et al. 1993; Quiroga et al. 2002). There is also evidence that once PA is developed in the child's native language, these skills may be applied when learning to read in a second language. Durgunoglu et al. (1993) found first-grade Spanish-speaking ELs who displayed strong Spanish PA and word-decoding ability were better at decoding English words and pseudowords than students with poor native-language skills.

Results from intervention studies conducted with ELs suggest that many of the principles of instruction associated with improved outcomes for native English-speaking students can be applied to ELs. In a meta-analysis of PA instruction with English-speaking children, Ehri et al. (2001) found that PA interventions are most effective when conducted with small groups of two to seven students. They also found that effective interventions can be relatively short in duration, ranging from approximately 300 to 600 min.

*Letter Knowledge* Letter knowledge, specifically the ability to accurately and fluently name the letters of the alphabet, has been found to be predictive of reading outcomes (e.g., Kaminski and Good 1996; O'Connor and Jenkins 1999). In the study by O'Connor and Jenkins, measures of PA and rapid letter naming were the strongest predictors of reading difficulty in kindergarten and first grade. Unfortunately, one serious limitation of the rapid letter naming task is that ceiling effects have often been observed with this measure when administered to students at the end of kindergarten and first grade (Schatschneider et al. 2004).

*Alphabetic Principle* The alphabetic principle represents the knowledge of how letters correspond to sounds and the ability to blend these sounds to form words (National Reading Panel 2000). One measure of alphabetics, pseudoword reading, has been found to be strongly correlated with real-word reading for both monolinguals and ELs (Fien et al. 2008; Lesaux and Siegel 2003; Vanderwood et al. 2008). Vanderwood and colleagues examined the relationship between nonsense word fluency (NWF), a measure of pseudoword reading, and the reading performance of 134 first-grade ELs. First-grade NWF scores were significantly correlated with third-grade outcome measures, including oral reading fluency ($r=.65$), Maze ($r=.54$), and the California Achievement Test ($r=.39$). In addition, Lesaux and Siegel (2003) found that a measure of pseudoword reading in kindergarten was the best predictor of word reading and comprehension for second-grade ELs.

*Oral Reading Fluency* An assessment of oral reading fluency (ORF) can be used to make inferences about a student's decoding skills and can also serve as a reliable and valid index of general reading achievement, including reading comprehension (Shinn et al. 1992). A number of past studies have found moderate to strong correlations between ORF and measures of reading comprehension for both native English speakers and ELs (e.g., Baker and Good 1995; Good and Kaminski 2002; Shinn, et al. 1992). Baker and Good (1995) found ORF to be sensitive to growth over periods as short as 2 weeks and predictive of performance on second-grade statewide reading assessments for bilingual Latino students. Similarly, Wiley and Deno (2005) found ORF to be significantly correlated with reading performance on a statewide reading assessment for third ($r = .61$) and fifth-grade ELs ($r = .69$). Past studies have also demonstrated that ORF is superior to other, more "direct" assessments of comprehension, including oral and written retell of stories, cloze passages, Maze, and question answering (Fuchs and Fuchs 1992; Wiley and Deno 2005).

## Purpose

The purpose of this study is to provide validity data to guide the use of DIBELS with Korean English Learners. Using the Fuchs' (1999) guidelines about the evidence that should be provided for screening and progress monitoring tools as the structure for our study, we addressed four questions designed to address the question of whether DIBELS is appropriate for first-grade Korean ELs:

1. To what extent are DIBELS early literacy measures psychometrically sound for Korean ELs?
2. Are early literacy screening measures capable of showing growth for Korean ELs over the course of the year?
3. To what extent are early literacy progress monitoring measures sensitive to growth during a 6-week targeted PA intervention?
4. Do teachers value the data received from the DIBELS assessments for Korean ELs?

## Method

### Participants

A school with a large Korean EL population in Southern California volunteered to participate in the study. The school operated on a year-round schedule and served approximately 1,333 students in grades K–5 with approximately 84 % of the students receiving free or reduced-price lunch. The ethnic distribution of the school was 64 % Hispanic, 29 % Asian, 5 % Filipino, and 2 % African American. Seventy-five percent of the school population consisted of ELs, with 24 % listing Korean as their home language. Two out of the six first-grade classrooms provided dual immersion instruction in Korean and English and volunteered to participate in the study. These classrooms were primarily comprised of Korean ELs (36/41=88 %), but were open to all children, regardless of language proficiency or ethnicity.

Regardless of ethnicity, all students from the two Korean dual-immersion classrooms participated in the initial screening. Participants who spoke a primary language other than Korean at home and participants who were classified as native or fluent English speakers were excluded from the analysis ($n=5$). Six children moved during the study, and their scores were also excluded from the analysis. Of the final screening data sample ($N=30$), 17 (57 %) of the participants were females and 13 (43 %) were males. All of the participants were between 6 to 7 years of age. As described more completely below, a second sample was produced by selecting the 10 lowest performing Korean ELs from the fall screening sample. These students were nominated to participate in the intervention that was designed to determine whether the progress monitoring tools were sensitive to change caused by an intervention. The teachers reviewed the data and agreed the scores represented current performance, and these students would benefit from a phonological awareness intervention. This sample of intervention students included six female and four male participants, and none of them were receiving special education services.

### Procedures

Participants were screened three times during the school (i.e., fall, winter, spring) year using measures from DIBELS 6.1 (Good and Kaminski 2002) and a reading comprehension measure (i.e., Maze; AIMSweb, Shinn and Shinn 2002). After the fall screening, a review of the data suggested a phonological awareness intervention was appropriate for the 10 students with the lowest performing scores. All 10 students were in the "at risk" level on the DIBELS measure of phonological awareness. Ten students were selected based on the amount of resources available for intervention, and due to the recommendation to provide intervention to no more than 25 % of students in a grade (Vanderwood and Nam 2008). (For the purposes of this study, the phonological awareness intervention was implemented to determine whether DIBELS 6.1 progress monitoring measures were sensitive to growth for this population and were not used to evaluate the intervention.)

The PA intervention (Vanderwood 2004), described below, was provided two times per week for 30 min per session for

6 weeks during the fall. Consistent with recommendations by Ehri et al. (2001), the intervention was provided in small groups, with five participants per group. Each session consisted of one lesson from the scripted intervention until all 12 lessons were completed. A baseline score was generated for each student using the initial screening data as well as two additional data points collected prior to the start of the intervention. Phoneme segmentation fluency (PSF) and nonsense word fluency (NWF) were used to monitor students' progress once per week using DIBELS Benchmarks as the goals for each measure. The progress monitoring sessions took approximately 5 min per student, and alternate forms of each measure were used each week. All students in the group received the same first-grade alternate form of PSF and NWF each week, and the probes were given in the order provided in the progress monitoring booklet for each measure.

Measures

*Letter Naming Fluency (LNF)* Letter Naming Fluency (LNF) is an individually administered test that assesses a child's ability to name upper and lower case letters of the alphabet. The 1-month, alternate-form reliability of LNF is .88 and the median criterion-related validity of LNF with the Woodcock-Johnson Psycho-Educational Battery-Revised Readiness Cluster is .70 (Good and Kaminski 2002). LNF was administered to all students in the fall of first grade.

*Phoneme Segmentation Fluency (PSF)* Phoneme segmentation fluency (PSF) is an individually administered test of PA. The alternate-form reliability of PSF is .88 for 2 weeks and .79 for 1 month (Kaminski and Good 1996). Good et al. (2002) report criterion validity of PSF with the Woodcock-Johnson Psycho-Educational Battery to be between .54 and .68. PSF was administered to all students in the fall, winter, and spring of first grade. The measure takes 1 min to administer.

*Nonsense Word Fluency (NWF)* Nonsense Word Fluency (NWF) is an individually administered test that assesses a child's ability to identify letter–sound correspondences and blend letters into pseudowords. The 1-month, alternate-form reliability of NWF in first grade is .83 (Good and Kaminski 2002). The predictive validity of NWF in January of first grade with ORF in May of first grade is .82 (Good and Kaminski). NWF was administered to all students in the fall, winter, and spring of first grade. The measure takes 1 min to administer.

*Oral Reading Fluency (ORF)* Oral Reading Fluency (ORF) is an individually administered measure of reading fluency and overall reading performance. Test–retest reliabilities for elementary students range from .92 to .97, and alternate-form reliabilities range from .89 to .94 (Good and Kaminski 2002). In the present study, ORF was administered to all students in

the winter and spring of first grade. For screening, three 1-min probes are administered and the median score is used.

*Maze* AIMSweb Maze is a multiple-choice cloze task that students complete while reading silently. This is a group or individually administered measure of reading comprehension and overall reading achievement (Shinn and Shinn 2002). Alternate-form reliability is reported to be .81, with 1- to 3-month intervals between testing (Shin et al. 2000). Criterion-related validity with published, norm-referenced tests of comprehension ranges from .77 to .85 (Fuchs and Fuchs 1992). Maze was administered to all students in a group format in the winter and spring of first grade and took approximately 5 min to administer. (Note: Maze was used to directly assess comprehension because DIBELS 6.1 did not include a comprehension measure.)

*Inter-observer Agreement* Inter-observer agreement was calculated for 20 % of the probes administered per screening period by having one person observe the examiner administering the measures. Total percentage agreement was calculated by dividing the number of agreements by the number of agreements plus disagreements, multiplied by 100. Inter-observer agreement for LNF ranged from 97 to 100 %, with an average combined agreement of 99 %. PSF ranged from 87 to 100 %, with an average combined agreement of 92 %. NWF ranged from 92 to 100 %, with an average combined agreement of 97 %. ORF ranged from 92 to 100 %, with an average combined agreement of 99 % and inter-observer agreement for Maze was 100 %.

Intervention Materials

This study utilized a PA intervention (Vanderwood 2004) that was previously validated with Spanish-speaking first-grade ELs (Healy et al. 2005). There were 12 scripted lessons based on a model-lead-test format: (a) the procedures were first modeled to the students, (b) the interventionist led the responses with the students, and then (c) students were asked to independently answer. Every 30-min lesson included a vocabulary section and five PA activities, including: phoneme production/replication, phoneme segmentation and counting, phoneme blending, phoneme isolation, and rhyming. Intervention fidelity checklists and observations were used to determine the degree to which the intervention was implemented in a manner consistent with the intervention protocol. The intervention was implemented with 100 % integrity during all 12 sessions.

Social Validity

A social validity survey was created to assess the teachers' views on the importance of the assessment tools and

intervention process and was given at the end of the intervention. The acceptability survey consisted of four Likert-scale questions regarding the intervention, with 1 being "strongly disagree" and 7 being "strongly agree".

## Results

The purpose of this study was to produce validity data that can be used to guide the use of DIBELS with first-grade Korean ELs. The first research question addressed the extent to which DIBELS is psychometrically sound when used with first-grade Korean ELs. Inter-rater reliability data varied slightly across measures with PSF having the lowest average reliability of .92. Although there is no agreement about what is considered adequate reliability, at least one set of authors suggest reliability should be above .90 when making decisions for individuals (Salvia et al. 2010). To provide evidence of validity of the measures, we conducted analyses to determine whether early literacy measures used for screening are related to future reading outcomes for this population of students. The strength of association between early literacy measures administered in the fall and reading achievement in the winter and spring are presented in Table 1. Fall LNF scores were significantly correlated with winter and spring ORF ($r$=.50 and .47, respectively; $p$<.01) and Maze ($r$=.45 and .56, respectively, $p$<.05). Significant correlations were not found between fall PSF and any of the other fall, winter, or spring measures. Fall NWF scores were significantly correlated with winter and spring ORF ($r$=.62 and .47, respectively, $p$<.01) and spring Maze ($r$=.44, $p$<.05). Spring NWF was also significantly correlated with spring ORF and Maze ($r$=.76

and .66, respectively, $p$s<.01). ORF in the winter was strongly correlated with ORF and Maze in the spring. Finally, a significant correlation was also found between spring ORF and Maze ($r$=.70, $p$<.01).

### Growth on Screening Measures

The second research question addressed the extent to which the screening data were sensitive to growth for all Korean ELs over the course of the year. Descriptive statistics for the screening measures for the full sample are summarized in Table 2. All data were normally distributed and were below 1.0 on skewness and kurtosis. Based on the DIBELS LNF benchmarks for the beginning of first grade (Good and Kaminski 2002), 25 out of the 30 total students scored in the low-risk category (LNF≥37) and five students scored in the some-risk category (25≤LNF<37) in the fall. Performance on fall PSF indicated that 10 students scored in the established category (PSF≥35), 15 students scored in the emerging category (10≤PSF<35), and five students scored in the deficit category (PSF <10). Results for fall NWF indicated that all but one student performed well above the at-risk category for this measure (NWF≥24).

Group averages for the fall, winter, and spring measures (see Table 2) suggested that the students made substantial growth between the screening periods. The mean ORF score indicated that the students were reading an average of 85.70 (SD= 27.71) words per minute in the winter and 109.23 (SD= 25.12) words per minute in the spring. Based on the DIBELS ORF 6.1 benchmarks for the middle and end of first grade (Good and Kaminski 2002), all of the students in this sample were in the low-risk range for this measure during the winter and spring screenings. Interestingly, although all students

**Table 1** Correlations between screening variables across screening periods

|    |          | 1     | 2    | 3     | 4     | 5     | 6     | 7    | 8    | 9     | 10    | 11 |
|----|----------|-------|------|-------|-------|-------|-------|------|------|-------|-------|----|
| 1  | LNF, T1  | –     |      |       |       |       |       |      |      |       |       |    |
| 2  | PSF, T1  | .29   | –    |       |       |       |       |      |      |       |       |    |
| 3  | NWF, T1  | .58** | −.02 | –     |       |       |       |      |      |       |       |    |
| 4  | PSF, T2  | −.22  | .22  | −.22  | –     |       |       |      |      |       |       |    |
| 5  | NWF, T2  | .34   | −.16 | .62** | −.15  | –     |       |      |      |       |       |    |
| 6  | ORF, T2  | .50** | −.15 | .61** | −.17  | .47** | –     |      |      |       |       |    |
| 7  | Maze, T2 | .45*  | .01  | .35   | −.42* | .30   | .33   | –    |      |       |       |    |
| 8  | PSF, T3  | −.19  | .35  | −.12  | .81** | −.06  | .07   | −.29 | –    |       |       |    |
| 9  | NWF, T3  | .24   | −.13 | .53** | −.04  | .33   | .72** | .26  | .21  | –     |       |    |
| 10 | ORF, T3  | .47** | .02  | .47** | −.05  | .22   | .80** | .25  | .21  | .76** | –     |    |
| 11 | Maze, T3 | .56** | .04  | .44*  | −.27  | .37*  | .68** | .58** | −.10 | .66** | .70** | – |

*Note*: *LNF* letter naming fluency, *NWF* nonsense word fluency, *ORF* oral reading fluency, *PSF* phoneme segmentation fluency, *T1* fall screening, *T2* winter screening, *T3* spring screening

*$p$<.05; **$p$<.01

**Table 2** Means and standard deviations for screening measures by screening period

| Variable | Fall screening | | Winter screening | | Spring screening | |
|---|---|---|---|---|---|---|
| | M | SD | M | SD | M | SD |
| LNF | 56.13 | 16.44 | | | | |
| PSF | 28.62 | 15.28 | 41.97 | 29.45 | 45.41 | 15.89 |
| NWF | 70.67 | 35.03 | 95.00 | 45.99 | 111.26 | 57.59 |
| ORF | | 85.70 | 27.71 | 109.23 | 25.12 | |
| MAZE | | | 8.30 | 4.60 | 16.50 | 5.91 |

*Note*: *LNF* letter naming fluency, *PSF* phoneme segmentation fluency, *NWF* nonsense word fluency, *ORF* oral reading fluency

achieved fluent word decoding skills in the winter and spring, as indicated by their ORF scores, 10 students continued to score in the emerging range for PSF in the winter, and seven students scored in the emerging range for PSF in the spring.

On the Maze measure, percentile ranks from AIMSweb norm-referenced tables (Edformation 2007) were used to compare the performance of these students to other first-grade students in the normative sample. In the winter of first grade, four students scored within the 25–50th percentile, nine students scored within the 50–75th percentile, and 17 students scored above the 75th percentile. By the spring of first grade, 29 students scored above the 75th percentile and one student scored within the 50–75th percentile. Average performance of the students in the winter (M=8.30, SD=4.60) and spring (M= 16.50, SD=5.91) of first grade indicated that the students made growth in reading comprehension.

Sensitivity to Intervention

Another way to determine whether the screening measures are sensitive to growth is to compare intervention with non-intervention students (Fuchs and Fuchs 1999). Students receiving intervention should close the gap on measures if the tools are sensitive to intervention effects. As presented in Table 3, the results indicated that the intervention students performed lower than the non-intervention students during the fall screening on all of the measures. By the winter screening, the intervention students outperformed the non-intervention students on both PSF and NWF, but not on ORF or Maze. The spring screening results indicated that the intervention and non-intervention students performed at a similar level on PSF, but the intervention group performed lower than the non-intervention group on NWF, ORF, and Maze. It is important to note that the average performance of both intervention and non-intervention students for each of the spring screening measures was in the established range.

**Table 3** Means and standard deviations for screening measures by group and screening period

| Variable | Intervention students (N=10) | | Non-intervention students (N=20) | |
|---|---|---|---|---|
| | M | SD | M | SD |
| Fall screening | | | | |
| LNF | 43.78 | 15.64 | 61.43 | 13.99 |
| PSF | 14.67 | 11.86 | 34.14 | 12.47 |
| NWF | 57.33 | 33.47 | 76.38 | 34.89 |
| Winter screening | | | | |
| PSF | 45.44 | 18.15 | 40.48 | 9.49 |
| NWF | 99.33 | 40.10 | 93.14 | 49.11 |
| ORF | 77.44 | 20.72 | 89.24 | 29.97 |
| Maze | 6.0 | 5.98 | 9.29 | 3.61 |
| Spring screening | | | | |
| PSF | 45.10 | 21.67 | 45.54 | 13.35 |
| NWF | 101.81 | 47.67 | 115.31 | 61.99 |
| ORF | 99.67 | 21.31 | 113.33 | 25.97 |
| Maze | 14.56 | 5.10 | 17.33 | 6.14 |

*Note*: *LNF* letter naming fluency, *NWF* nonsense word fluency, *ORF* oral reading fluency, *PSF* phoneme segmentation fluency

Growth on Progress Monitoring Measures

PSF and NWF progress monitoring data collected during the 6-week intervention for 10 students were used to examine the ability of the early literacy measures to monitor growth and goal attainment for Korean ELs. (Note: it was not the purpose of this study to validate the PA intervention. We selected this intervention because evidence indicated it worked with other ELs and it was based on sound literacy principles.) Initial PSF scores for the intervention students indicated that five students were performing in the deficit range and five were performing in the emerging range. By the end of the intervention, all of the students met their individual goals and performed within the DIBELS 6.1 established range for PSF. The average growth rate on PSF for individuals who received the intervention was 3.1 sounds per minute per week.

Initial NWF scores for the intervention students indicated that one student was performing in the at-risk range, one was in the some-risk range, and eight met benchmark criteria. In different circumstances, the students with established NWF scores would most likely not have been included in a PA intervention. In this case, we were attempting to validate the measures and the teachers indicated they believed that the students who were selected needed additional phonological awareness support. Progress monitoring data for NWF indicated that 9 out of 10 students were performing at benchmark by the end of the intervention. The average growth rate on NWF for students who received the intervention was 1.6

sounds per minute per week. The level of growth demonstrated on NWF and on PSF were slightly above the average growth demonstrated when this intervention was used with Spanish speaking ELs (Healy et al. 2005).

## Social Validity

The results of the social validity survey indicated that the two teachers found the intervention and the assessment tools to be highly acceptable. The teachers strongly agreed that the assessment focused on important reading skills, did not take too much of the student's time away from class, and could be used to improve a student's overall performance. All ratings ranged from 6 to 7 on a 7-point scale.

## Discussion

The purpose of this study was to gather evidence about the validity of DIBELS with a Korean speaking English Learner population. In some districts throughout the USA, Korean speaking students represent a substantial proportion of the overall EL population (Korean American Coalition-Census Information Center 2003). To appropriately identify who needs additional reading skills support, many school districts are employing tools like DIBELS with all their students, including ELs. Yet, at this point, there is only a small amount of evidence that these tools can work with students who have different language backgrounds than those who were part of the tests' development.

## Psychometric Properties

Using inter-rater reliability as an index, all the DIBELS 6.1 measures meet guidelines for the level of reliability to feel confident about making individual decisions (Salvia et al. 2010). PSF was the least reliable early literacy measure, although an inter-rater reliability of .92 should be sufficient for most type of decisions. These results are consistent with other studies that examined the use of DIBELS with Spanish speaking ELs (e.g., Vanderwood et al. 2008).

To provide initial evidence of the tools' validity with Korean ELs, the measures' ability to predict future performance was obtained by correlating the early literacy, fluency and comprehension measures across the three time periods. Consistent with past studies examining native English speakers and Spanish-speaking students, LNF and NWF in the fall were found to be significantly correlated with end-of-the-year reading achievement, as measured by ORF and Maze. In recent years, there has been increasing support for the use of NWF with both ELs and native English speakers to determine risk status. For example, Vanderwood et al. (2008)

examined the relationship between first-grade NWF and third-grade reading outcomes with a sample of ELs primarily composed of Latinos and Asians. Significant correlations were found between first-grade NWF scores and third-grade ORF ($r=.65$), Maze ($r=.54$), and the California Achievement Test Reading Composite score ($r=.39$). Fien et al. (2008) also found NWF in kindergarten to be a strong predictor of first- and second-grade reading outcomes, including ORF and SAT-10 ($r=.51–.65$). It is important to note the relationship of NWF to ORF grew over time as would be expected in normal reading development. The results of the present study with Korean ELs are consistent with current research, providing further support for the use of NWF in first grade to predict future reading outcomes.

In contrast, fall PSF was not significantly correlated with any of the fall, winter, or spring measures. While there is consensus that PA is a prerequisite to reading acquisition (e.g., Cunningham and Stanovich 1997; Stanovich 1986; Wagner and Torgesen 1987), relatively weak correlations between PSF and later reading outcome measures have been found in previous studies. For example, a study by Kaminski and Good (1996), conducted with primarily native English speakers, did not find PSF to be significantly correlated with the Stanford Diagnostic Reading Test or the Metropolitan Readiness Test.

It is also important to note the winter measures of ORF and Maze were strongly correlated with the spring measures of the same test. In addition, ORF and Maze were also strongly correlated within and across time periods. This result is significant because it suggests both ORF and Maze are sensitive enough to be used with Korean ELs as early as the winter of first grade. It is not surprising that a strong relationship was found between spring ORF and Maze. ORF can be used to make inferences about a student's decoding skills and also serve as a reliable and valid index of general reading achievement, including reading comprehension (Shinn et al. 1992). The results of this study are consistent with Perfetti's Verbal Efficiency Theory (1985) which states that fluent readers are able to read text quickly and accurately, allowing them to expend less cognitive resources on decoding and have a higher capacity for comprehension.

## Ability to Show Growth

Performance during the fall, winter, and spring of first grade suggested that DIBELS 6.1 measures and AIMSweb Maze were sensitive to growth for this sample of first-grade Korean ELs. Students exhibited growth in PA and alphabetic knowledge, as reflected by the increase in their PSF and NWF scores between the fall and spring. Average ORF and Maze scores in the spring indicated that the students achieved high levels of English word recognition and comprehension skills by the end of first grade.

Despite their EL status and limited English vocabulary, the students in our sample showed growth in reading fluency and comprehension throughout the year that was comparable to results with native English speakers (Fien et al. 2008). By the spring of first grade, all of the students in our sample met the DIBELS 6.1 benchmark goal for ORF and were performing above the AIMSweb 50th percentile for Maze.

One possible reason why students in this study reached high levels of reading fluency and comprehension despite low scores on phoneme segmentation is that the development of English PA may have been affected by native language skills. Although several of the phonemes represented in Korean are similar to those in English (e.g., /m/ and /n/), there are many phonemes in English that are not represented in Korean. For example, Korean lacks the sounds /f/, /v/, initial /l/, and the /th/ sounds as in *think* and *this* (Taylor and Taylor 1995). As a result, the students in the present study may have encountered difficulties in distinguishing the sounds during the PA task, resulting in lower performance.

## Limitations

The most significant limitation of this study was the small number of participants. As a result, it is difficult to generalize findings and eliminate all alternative hypotheses about the relationships among the variables. However, given the scarcity of research with Korean speaking ELs, these results can be used to support future research and practice in providing assessment-driven services. In fact, it is clear from our study there is a need to further understand the validity of using PSF with first-grade Korean ELs.

The purpose of the intervention in this study was to determine whether DIBELS 6.1 early literacy tools were sensitive to change caused by an intervention focused on phonological awareness. The data were not used to modify the intervention nor determine the intervention's overall effectiveness. An empirically supported intervention previously used with Spanish speaking ELs was implemented with the expectation it would have a positive impact on the phonological awareness skills of Korean ELs. The intervention did have a positive effect, but given the lack of use of a design that can appropriately measure intervention effects (e.g., randomized control trial, regression discontinuity), conclusions about intervention effectiveness based on this study should not be made.

## Implications

Although past studies have reported that many of the same tools and strategies that have been used to screen native English speakers can be used to improve outcomes for ELs (e.g., Baker and Good 1995; Haager and Windmueller 2001), there is clearly a need for further research and continued development of assessment tools and strategies that are

sensitive to the language proficiency and unique cultural background of ELs. The results of this study and others (e.g., Riedel 2007) suggest educators should use caution when interpreting the results of PSF with first-grade ELs. It is quite possible a student's PSF score will not be predictive of future reading outcomes. Yet, this study and others (e.g., Vanderwood, et al. 2008) do support the use of NWF and ORF with ELs. Besides evidence of positive psychometric validity for all measures except PSF for screening, the teachers who completed the social validity measures agreed the tests were appropriate for Korean ELs. It is important to note these conclusions are only focused on first-grade students and are similar to the results obtained with native English speakers (Riedel 2007). In fact, in the most recent version of DIBELS (DIBELS Next), the use of PSF was scaled down significantly in first grade and is only used in the fall. It appears for all students, including ELs, a student's skill in the alphabetic principle during first grade is more predictive of future reading skills than performance on phonological awareness tasks.

## References

APA, AERA, & NCME. (1999). Testing individuals of diverse linguistic backgrounds. In AERA/APA/NCME (Ed.), *Standards for educational and psychological testing* (pp. 91–97). Washington, D.C.: American Educational Research Association.

Baker, S. K., & Good, R. H. (1995). Curriculum-based measurement of English reading with bilingual Hispanic students: a validation study with second-grade students. *School Psychology Review, 24*, 561–578.

Cho, J. R., & McBride-Chang, C. (2005). Correlates of Korean Hangul acquisition among kindergartners and second graders. *Scientific Studies of Reading, 9*, 3–16.

Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology, 33*, 934–945.

Durgunoglu, A. Y., Nagy, W. E., & Hancin-Bhatt, B. J. (1993). Cross-language transfer of phonological awareness. *Journal of Educational Psychology, 85*, 453–465.

Edformation. (2007). *AIMSweb growth tables.* Retrieved July 13, 2007, from http://www.aimsweb.com/

Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly, 36*, 250–287.

Fien, H., Baker, S. K., Smolkowski, K., Smith, J. L., Kame'enui, E. J., & Beck, C. T. (2008). Using nonsense word fluency to predict reading proficiency in kindergarten through second grade for English learners and native English speakers. *School Psychology Review, 37*, 391–408.

Fuchs, L. S., & Fuchs, D. (1992). Identifying a measure for monitoring student reading progress. *School Psychology Review, 21*, 45–58.

Fuchs, L. S., & Fuchs, D. (1999). Monitoring student progress toward the development of reading competence: a review of three forms of classroom-based assessment. *School Psychology Review, 28*, 659–671.

Gersten, R., Baker, S. K., Shanahan, T., Linan-Thompson, S., Collins, P., & Scarcella, R. (2007). *Effective literacy and English language instruction for English learners in the elementary grades: a practice guide.* Washington, D.C.: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Good, R. H., & Kaminski, M. A. (Eds.). (2002). *Dynamic indicators of basic early literacy skills* (6th ed.). Eugene, OR: Institute for the Development of Education Achievement.

Haager, D. S., & Windmueller, M. P. (2001). Early reading intervention for English language learners at-risk for learning disabilities: student and teacher outcomes in an urban school. *Learning Disabilities Quarterly, 24,* 235–250.

Healy, K., Vanderwood, M., & Edelston, D. (2005). Early literary interventions for English language learners: support for an RTI model. *The California School Psychologist, 10,* 55–63.

Hopstock, P. J., & Stephenson, T. G. (2003). *Special Topic Report #1: Native languages of LEP students.* Retrieved February 9, 2007, from http://www.ncela.gwu.edu/resabout/ research/descriptivestudyfiles/native_languages1.pdf

Kaminski, R. A., & Good, R. H. (1996). Toward a technology for assessing basic early literacy skills. *School Psychology Review, 25,* 215–227.

Kindler, A. L. (2002). *Survey of the states' limited English proficient students and available educational programs and services: 2000-2001 summary report.* Retrieved from National Clearinghouse for English Language Acquisition website: http://www.ncela.gwu.edu/files/rcd/BE021853/Survey_of_the_States.pdf

Klingner, J. K., & Edwards, P. (2006). Cultural considerations with response to intervention models. *Reading Research Quarterly, 41,* 108–117.

Korean American Coalition-Census Information Center. (2003). *Korean population in the United States, 2000.* Retrieved from California State University, Center for Korean American and Korean Studies website: http://www.calstatela.edu/centers/ ckaks/census/ 7203_tables.pdf

Lesaux, N. K., & Siegel, L. S. (2003). The development of reading in children who speak English as a second language. *Developmental Psychology, 39,* 1005–1019.

National Center for Education Statistics. (2009). *Reading 2009. National assessment of educational progress at grades 4 and 8.* Retrieved from National Assessment for Educational Progress website: http://nationsreportcard.gov/reading_ 2009 /nat_g4.asp.

National Clearinghouse for English Language Acquisition. (2007). *The growing numbers of LEP students: 2005-2006 poster.* Retrieved April 9, 2008, from http://www.ncela.gwu edu/stats/2_nation.htm.

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction.* Retrieved June 10, 2007, from http://www.nationalreadingpanel.org/Publications.

O'Connor, R. E., & Jenkins, J. R. (1999). The prediction of reading disabilities in kindergarten and first grade. *Scientific Studies of Reading, 3,* 159–197.

Perfetti, C. A. (1985). *Reading ability.* New York: Oxford University Press.

Quiroga, T., Lemos-Britton, Z., Mostafapour, E., Abbott, R. D., & Berninger, V. W. (2002). Phonological awareness and beginning reading in Spanish-speaking ESL first graders: Research into practice. *Journal of School Psychology, 40,* 85–111.

Riedel, B. W. (2007). The relation between DIBELS, reading comprehension, and vocabulary in urban first grade students. *Reading Research Quarterly, 42,* 546–567.

Salvia, J., Ysseldyke, J. E., & Bolt, S. (2010). *Assessment in special and inclusive education* (11th ed.). Belmont, CA: Wadsworth.

Schatschneider, C., Fletcher, J. M., Francis, D. J., Carlson, C., & Foorman, B. R. (2004). Kindergarten prediction of reading skills:

a longitudinal comparative analysis. *Journal of Educational Psychology, 96,* 265–282.

Shin, J., Deno, S. L., & Espin, C. A. (2000). Technical adequacy of the maze task for curriculum-based measurement of reading growth. *Journal of Special Education, 34,* 164–172.

Shinn, M. R., & Shinn, M. M. (2002). *AIMSweb training workbook: administration and scoring of reading maze for use in general outcome measurement.* Retrieved February 10, 2007 from http://www.aimsweb.com

Shinn, M. R., Good, R. H., Knutson, N., Tilly, W. D., & Collins, V. L. (1992). Curriculum-based measurement reading fluency: a confirmatory analysis of its relation to reading. *School Psychology Review, 21,* 459–479.

Snow, C. E., Burns, S. M., & Griffin, P. (1998). *Preventing reading difficulties in young children.* Washington, D.C: National Academy Press.

Stanovich, K. E. (1986). Matthew effects in reading: some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21,* 360–407.

Stanovich, K. E., Cunningham, A. E., & Cramer, B. B. (1984). Assessing phonological awareness in kindergarten children. Issues of task comparability. *Journal of Experimental Child Psychology, 38,* 175–190.

Taylor, I., & Taylor, M. M. (1995). *Writing and literacy in Chinese, Korean, and Japanese.* Philadelphia: John Benjamins.

Vanderwood, M. L. (2004). *Phonemic Awareness and Vocabulary Intensive Intervention (PAVII) Version 1.0.* Unpublished manual, Graduate School of Education, University of California-Riverside, Riverside, CA.

Vanderwood, M. L., & Nam, J. (2008). Best practices in using a response to intervention model with English language learners. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology V.* Bethesda, MD: National Association of School Psychologists.

Vanderwood, M. L., Linklater, D., & Healy, K. (2008). Predictive accuracy of nonsense word fluency for English language learners. *School Psychology Review, 37,* 5–17.

Vellutino, F. R., Scanlon, D. M., & Tanzman, M. S. (1998). The case for early intervention in diagnosing specific reading disability. *Journal of School Psychology, 36,* 367–397.

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin, 101,* 192–212.

Wagner, R. K., Torgesen, J. K., Laughon, P., Simmons, K., & Rashotte, C. A. (1993). The development of young readers' phonological processing abilities. *Journal of Educational Psychology, 85,* 1–20.

Wiley, H. I., & Deno, S. L. (2005). Oral reading and maze measures as predictors of success for English learners on a state standards assessment. *Remedial and Special Education, 26,* 207–214.

Zehler, A., Fleischman, H., Hopstock, P., Stephenson, T., Pendzick, M., & Sapru, S. (2003). *Policy report: summary of findings related to LEP and SPED-LEP students.* Retrieved from http://onlineresources.wnylc.net/pb/orcdocs/LARC_.

**Michael L. Vanderwood** is an Associate Professor and Coordinator of the School Psychology program at UC-Riverside.

**Jeanie E. Nam** is a graduate of the UCR School Psychology program and is now practicing school psychologist.

**Jennifer W. Sun** is a graduate of the UCR School Psychology program and is now practicing school psychologist.