Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Volume 10 Number 18, November 2005 ISSN 1531-7714

Enhancing Validity in Phonological Awareness Assessment through Computer-Supported Testing

Jerrell C. Cassady, Lawrence L. Smith, & Linda K. Huber
Ball State University

Phonological awareness is an early indicator of emergent reading skill that is known to be reliably related to eventual reading performance. This established research based coupled with federal and state requirements to measure phonological awareness as an indicator of early reading program success has heightened the attention toward phonological assessment tools. The purpose of this paper is to identify two central threats to validity that are present in the standard assessment tools and provide a methodological solution to both threats using the Standardized Assessment of Phonological Awareness as an example.

The research on early literacy has provided several clear and articulate examinations of the developmental nature of young children's acquisition of phonological and phonemic awareness and the connection of those skills to reading proficiency (Ehri, Nunes, Willows, Schuster, Yaghoub-Zadeh, & Shanahan, 2001). In addition, federally-funded reading programs targeting emergent literacy development routinely require clear identification that children in the primary grades have demonstrated success on phonological awareness skills (Gordinier & Foster, 2004). As such, assessment tools targeting phonological and phonemic awareness abilities have become pervasive in educational assessment, evaluation, and program interventions (Lane, Pullen, Eisele, & Jordan, 2002).

The purpose of this paper is to identify what we believe to be two critical threats to validity in assessing phonological awareness that are present in most published phonological awareness measures. The first threat is the tendency in phonological awareness tools to assess broad domains of emergent skills rather than discrete abilities. The second threat is the individual variations in orally presented prompts that are unavoidable without a pre-recorded testing protocol. In addition, we demonstrate sufficient validity and reliability for an alternative method of assessing phonological awareness that eliminates these threats. We believe that the standard presentation of phonological material is a simple and reasonable fix that can be enacted with any existing phonological awareness assessment protocol. Such revisions to existing measures are expected to provide gains in the field by allowing researchers, evaluators, and educators to gain greater confidence in their assessment of children's phonological processing skills.

PHONOLOGICAL AWARENESS

Although there is widespread acceptance of the connection between phonological awareness and reading proficiency (e.g., Blachman, 2000: Ehri, Nunes, Willows, Schuster, Yaghoub-Zadeh, Shanahan, 2001), the corpus of literature in phonological and phonemic awareness has a consistent problem with operational definitions. For example, in our work we have found that the term "phonemic awareness" is often employed to describe skills and abilities that are beyond the purview of the phoneme, which fragments the literature base on the topic and impedes educators' understanding for research findings. . To establish clarity in our terminology, we offer the following operational definitions.

Phonological awareness is the awareness of constituent sounds of words and the ability to detect and eventually manipulate auditory units that do not necessarily hold syntactic meaning (Goswami, 2000; Snow, Burns, & Griffin, 1998; Sodoro, Allinder, & Rankin-Erickson, 2002; Harris & Hodges, 1995). Phonological awareness encompasses emergent readers' abilities to detect and manipulate progressively smaller units of sound within spoken words. Four established levels of phonological units are (a) syllables (/CAT/); (b) onset-rimes, which involves breaking the syllable into two parts with the split occurring directly before the vowel (/C/ /AT/); (c) body-coda, which involves breaking the syllable into two parts with the split occurring directly after the vowel (/CA/ /T/); and (d) phoneme, which is breaking the syllable into each distinct component piece (/C/ /A/ /T/; Cassady & Smith, 2004a; Goswami, 2000). Thus, the often-confused term phonemic awareness is a subset of the broader construct phonological awareness (Snow et al.) and involves conscious awareness of the smallest distinguishable auditory units in words (Harris & Hodges, 1995).

The abilities to detect and manipulate phonological units within words (i.e., syllable, onset-rime, body-coda, phoneme) are acquired in progressive fashion by emergent readers. The first step in gaining a phonological processing skill is to detect, or isolate, the component sound within a word. As the learner gains automaticity in these isolation and detection skills, they progress to the ability to manipulate the phonological units. Such tasks include the ability to blend two or more discrete sounds into a complete whole, segment or break apart whole words into component sounds, substitute alternate sounds for specific syllabic units, or report what would be left of a word when removing one identified phonological unit. To help elucidate the various phonological processing tasks, Table 1 presents a set of common phonological awareness tasks and example items. Phonological awareness mastery for a given phonological unit or task is considered mastered when the learner recognizes the alphabetic representations for auditory stimuli, also known as alphabetic insight (see Snow et al., 1998).

Detecting discrete abilities. The various actions of detecting, recognizing, manipulating, and substituting the sounds that make up words dictate specific and isolated phonological processing skills. Although it is clear that there are various degrees of complexity in processing sub-syllabic utterances, there is steady debate regarding the level of specificity necessary in assessment tools designed to measure these abilities. In the 1980's, two research teams explored phonological awareness with sets of items tapping 10 phonological tasks (Stanovich, Cunningham, & Cramer, 1984; Yopp, 1988). While both groups found validity in those distinct tasks, they reduced the number of levels of phonological awareness through factor analytic procedures (see Table 2 for illustration). This approach to synthesizing a complex body of data surely facilitates the establishment of a more simplified theoretical model, but the simplification in theory simultaneously poses a threat to diagnostic or prescriptive testing intended to highlight areas of deficit or excellence. For instance, combining the beginning and ending sound isolation tasks into one factor (Stanovich et al.) has since been shown to provide imprecise measurement of phonological awareness because children acquire the ability to isolate sounds in words in a progressive fashion; first they master beginning sounds, then the end, and finally the middle (Cassady & Smith, 2003). Also outlined in Table 2 is the representation of a more recent attempt to build an assessment model using IRT designs which provided yet another conceptualization for the steps of phonological awareness development and a new set of subskills to focus on in the assessment process (Schatschneider, Francis, Foorman, Fletcher, & Mehta, 1999).

Table 1. *Phonological Awareness Assessment Task Examples*

Phonological Awareness Task	Basic Instructions	Sample Item(s)
Rhyme recognition	Rhymes are words that sound the same at the end... Tell me if these words rhyme.	**ape-knee; dip-hip**
Rhyme application	Tell me a word that rhymes with:	**cap**
Oddity tasks: Beginning sounds	Listen to the names of these pictures. Tell me which one has a different **beginning** sound.	**nest, soap, nails**
Oddity tasks: Ending sounds	Listen to the names of these pictures. Tell me which one has a different **ending** sound.	**bell, web, crib**
Oddity tasks: Middle sounds	Listen to the names of these pictures. Tell me which one has a different **middle** sound.	**beak, cone, heel**
Blending body-codas	I will say two parts of a word separately. You tell me the word.	**/co/ /p/**
Blending onset-rimes	I will say the first sound of a word and then the rest of the word separately. Tell me the whole word	**/c/ /op/**
Blending phonemes	I'm going to say each sound of a word slowly, then you tell me the word.	**/s/ /a/ /ve/ -- "what is the word put together?"**
Segmenting onset-rimes	Split the word by saying the first sound and then the rest of the word:	"Split the word **coat** by saying just the first sound and then the rest of the word.
Segmenting phonemes	Say each sound you hear in the word	**job**
Phoneme deletion	Listen to the word ____. Take away the **first** sound, what is left?	Listen to the word **book**. Take away the **/b/** sound, what is left?
Phoneme Substitution: Beginning sounds	If I change the first sound in the word **man** to **/p/**, the new word is **pan**.	Change the first sound in **cat** to **/h/**. What is the new word?
Phoneme Substitution: Ending sounds	If I say the word **rat** and change the last sound to **/g/**, the new word is **rag**.	Change the last sound in **cat** to **/p/**. What is the new word?
Phoneme Substitution: Middle sounds	If I say the word **pan**, change the middle sound to **/i/**, the new word is **pin**.	Change the middle sound in the word **cat** to **/o/**, what's the new word?

Note: The phonological awareness task examples are based on the structure of the SAPA. There are variations across measures on the instructions, types of items, and number of tasks assessed.

However, the most influential model of phonological awareness to date has been Adams' (1990) five-stage developmental approach. The first level is described as having "an ear for the sounds of words" (Adams, p. 80), which is primarily measured through children's knowledge of nursery rhymes or ability to remember rhyming words more easily than non-rhyming words. The second level is the ability to successfully master oddity tasks, where the child can compare and contrast words on the dimensions of rhyme and alliteration. Third in this model is the ability to blend syllables or phonemes, as well as recognize that syllables can be split. The fourth level is characterized by the actual ability to split words into phonemes on demand. Finally, the fifth level is phoneme manipulation, in which the reader can add or delete specified phonemes from target words and produce the new word (or non-word). Despite the popularity in the field held for this model, there are assessment barriers presented through this perspective. For instance, the third level "syllable and phoneme blending and awareness of the ability to segment syllables" examines the full acquisition of an overall blending skill. However, controlled empirical investigations have repeatedly demonstrated that children are able to segment and blend specific sub-syllabic units more readily than others (Cassady & Smith, 2004a; Treiman, 1985; Treiman & Zukowski, 1996).

Contesting the validity of well-established theoretical models of phonological awareness development is beyond the scope of this study. However, a fundamental point on the assessment of phonological awareness makes these points relevant. Specifically, there has been a disquieting trend in the past 15 years to build simplified or broad assessment tools for phonological awareness that are based primarily on these theoretical models. As such, assessment instruments have been losing the specificity needed to get more prescriptive and diagnostic information regarding the development of these discrete phonological awareness tasks. Using a more specific assessment tool, our research team has been able to find developmental trends within the broad theoretical stages offered in the literature. In addition, the subscales that are quite specific to phonological processing skills can be combined to generate combined ability subscores that mimic the more broad assessment tools and can be used to test the theoretical models offered by the various research teams.

Reliability across testing conditions. The second critical error that is common to most measures of phonological awareness is based on the typical mode of assessment. The traditional method of assessing phonological awareness involves individual or group administered tasks in which the test administrator reads an auditory prompt to the child, requesting that the child either identify or manipulate a specified phonological unit in the word (see Table 1 for examples). For items that require comparison of multiple words (e.g., rhyme awareness, oddity tasks) the administration typically involves presenting associated images to limit the burden of working memory during specific phonological processing tasks (Gibbs, 2004; Sodoro, Allinder, & Rankin-Eriskson, 2002).

We argue that this methodology presents a second threat to validity and reliability in assessing phonological awareness skills. Specifically, given that the task is one in which the student is required to identify, manipulate, or substitute meaningful information about auditory units, there is an inherent validity risk posed when multiple test administrators are involved in reading the auditory content to the learner. Variations in administrators' dialects, speech rate, enunciation, diction, or accent can make each presenter provide a different test stimulus than her or his colleagues. This reality was highlighted in a recent study demonstrating that African American first graders with normal reading skills received disproportionately lower and negatively skewed scores on a popular phonological awareness test (Thomas-Tate, Washington, & Edwards, 2004). The authors concluded that dialectical differences were interfering with the students' performance levels on the orally presented test and called into question the use of existing tests of phonological awareness given the variations in dialect observed in diverse settings. Given the current availability of technical delivery devices, there is no longer a reasonable rationale for continued presentation of these phonological awareness prompts in a non-standardized fashion.

It is important to document that there are tests and subtests that involve non-auditory assessment of phonological awareness (e.g., asking children to tap out the number of syllables in response to picture cards; Adams, Foorman, Lundberg, & Beeler, 1998). These alternative strategies overcome the concern we have regarding the individual variation in language production driven by multiple test administrators. However, tasks that require participants to tap out the syllables in response to a picture card may require more than phonological processing. Specifically, the individual is being asked to access orthographic information, perceive visual content whilst attempting to develop an auditory representation for the object in working memory, and respond non-verbally. Such a task is complex at best and may not provide a realistic or specific test of phonological detection or manipulation. Even more difficult (and not a test of just phonological awareness in our estimation) are those tasks that require clear alphabetic insight, where the participant is asked to blend sounds represented on a set of letter cards placed in a row.

COMPUTER-SUPPORTED PHONOLOGICAL AWARENESS ASSESSMENT

As mentioned earlier, we propose that there are two critical threats to valid assessment of phonological awareness in most existing measures. First, broad measures produce imprecise information that hampers the ability to clearly identify tasks within the emergent readers' skill sets. Second, when various test administrators are involved in the data collection process, there is an uncontrollable level of inconsistency in the presentation stimuli. In our own work, we have observed variations in pace, spacing, inflection, presence of schwa, and simple pronunciation in popular tests of phonemic awareness and manipulation. Whether the goal of administering the test is to provide diagnostic information relative to a norm group, establish program efficacy in a school-based literacy initiative, or draw upon experts' suggested curricular materials using a standard assessment protocol linked to instructional content, non-standard assessment procedures call the conclusions into question. In an era of educational research where there is increasing attention to demonstrate with "sound research practices" the impact or efficacy of programs and activities, researchers and practicing educators need to be sensitive to validity threats to make research and programmatic results meet the level of scrutiny exacted upon educational and reading research in particular.

In response to a perceived need to provide more specific and comprehensive reports on the phonological awareness skills mastered by students in a research initiative targeting emergent readers' development, our research team developed a measure previously referred to as the Phonological Awareness Test (PAT; Cassady, Smith, Bauserman, Jordan, Walker, & Popplewell, 2002; Cassady & Smith, 2004a). Early use of the tool demonstrated it helped to overcome the first critical threat to validity discussed earlier by providing valid and reliable assessment data on 14 distinct dimensions of emergent phonological awareness skills, sensitive to both phoneme position (beginning, middle, end sounds in words) and linguistic complexity (structural components within a syllable; phoneme, onset-rime, body-coda; see Stahl & Murray, 1994). However, we continued to see the potential threat to validity in non-standardized presentation of the test stimuli. To overcome this glaring potential problem in the assessment systems used in so many educational initiatives, we modified our own phonological measure and created the Standardized Assessment of Phonological Awareness (SAPA). Specifically, we now deliver the SAPA to students using a computer that plays pre-recorded digital audio and video test stimuli. Given the age of the standard subject screened for emergent literacy skills and our use of the tool as a measure of program success in the beginning phases of literacy development (i.e., ages 4-7), we have maintained a standard practice of individual administration. However, as some early phonological screening tools are administered to groups of young children (Lane et al., 2002), it is feasible that this process could be used in the field to increase the number of children who can be tested at any one time.

Other than the SAPA, we are aware of only two computerized phonological awareness assessment tools that have been validated in the research community. The first program, Cognitive Profiling System (CoPS), is a broad developmental assessment program that has 27 tasks addressing cognitive functioning in children (including phonological awareness activities; Singleton, Thomas, & Horne, 2000). The primary phonological processing activities included in the CoPS suite used by Singleton et al. included detecting or identifiying appropriate examples of rhyme and alliteration awareness as well as simple auditory discrimination by choosing the correct pronunciation of a pictured word (e.g., rock) from a set of auditory presentations (e.g., "wock" and "rock"). The assessment activities in CoPS are delivered through an engaging game format.

The second program, Heps-Kups Land, is a Finnish language program that assesses word-level segment identification, syllable-level segment identification, phonological unit synthesis (blending), and continuation of phonological units where the subject provides the ending sound to a word unit presented in conjuction with a photo depicting the target word (Puolakanaho, Poikkeus, Ahonen, Tolvanen, & Lyytinen, 2003). This program is also an engaging animated environment and is designed for children under the age of 4.

The SAPA differs from the pre-existing programs in 3 primary ways. First, the phonological awareness processing tasks represented by the SAPA's 14 subscales are far more distinct than the broad scales offered by CoPS or Heps-Kups Land. This is a known unique feature of the SAPA, and was intentionally created to provide more specific and discreet information on isolated phonological awareness tasks. Second, the SAPA is not embedded in a game format. The SAPA is more aligned with traditional tests of phonological processing (e.g., Stanovich et al., 1984; Yopp & Yopp, 2000) or broad emergent reading ability (e.g., DIBELS; Good & Kaminski, 2002). We consider this difference to be important, as the tasks in the SAPA are specific, clear, and unencumbered by the context of characters, animation, or plot as in the animated programs. Third, the SAPA provides complete standardization in the presentation of phonological units. That is, the stimuli on the SAPA are all read by one professional male voice with no obvious dialect and all phonological units are manipulated digitally to ensure that all breaks between phonological units are at 1-sec intervals. This standardization of timing is particularly important in the blending tasks where the children are asked to bring together distinct auditory units to make a coherent word (e.g., blending the three phonemes for /c/ /a/ /t/). Differential pauses between the three sounds can make the task dramatically easier or more difficult for individual test takers.

To illustrate the possibilities of improved psychometric attributes for the assessment of phonological awareness through computer-supported presentation, we present a summary of the validity, reliability, and procedural benefits observed in using the SAPA. While the SAPA is the only measure available to us for this presentation, we affirm that it is the method—not the measure—that provides the proposed advantages. That is, any phonological assessment tool that provides specific assessment of discrete phonological awareness skills through pre-recorded stimuli that have controlled for dialect and timing in the presentation of materials would be expected to overcome the threats we identify.

Validity

Validation of the SAPA was undertaken in waves of analyses, as iterations of the final scale were developed. Driven initially by Adams' (1990) conception of phonemic awareness development, with influences from other established theories of phonological processing, the initial 13-subscale and current 14-subscale versions of the instrument were tested with emergent readers across four academic years.

Content validity. A non-empirical validation approach is available through content, or face, validity estimation. The subtests in the SAPA were developed to follow the leads of several existing phonological and phonemic awareness tests. Table 2 displays the theoretical orientation of the SAPA with existing proposed models' explanations for the progressive development of phonological awareness skills. The overlap of the subtests on the SAPA across the existing theoretical and empirical models supports our assertion that the discrete tasks measured by the SAPA are consistent with the models of emergent literacy development that have guided the field for the past 2 decades. In particular, this measure provides discrete assessment of tasks with sensitivity to phonemic position (beginning, middle, ending sounds),

Table 2. *SAPA Subtest Alignment with Established Models.*

SAPA Subscales	Adams (1990)	Yopp (1988)	Stanovich et al. (1984)	Schatschneider et al. (1999)[a]	Stahl & Murray (1994)[b]
Rhyme recognition	1	1	1		1
Rhyme application		1	1		
Oddity: Beginning		3	2	1	
Oddity: Ending	2	3	2		
Oddity: Middle		3			
Blend body-codas					
Blend onset-rimes	3			2	2
Segment onset-rimes	4				2
Blend phonemes	3	2		3	3
Segment phonemes	4	4		5	3
Phoneme deletion	5	5	3	4	
Phoneme substitute: Beginning Sound	5				
Phoneme substitute: Ending Sound	5				
Phoneme substitute: Middle Sound	5				

Note. This comparison is an illustration of our conceptualization for how the cited models best fit into the 14 subscales of the SAPA and are not endorsed by the theorists cited.

[a] Schatschneider et al. (1999) include a sixth level that involves blending phonemes into non-words, which is not represented in the SAPA.

[b] Stahl & Murray's (1994) model regarding linguistic complexity also includes Level 4-Manipulate Cluster Onsets and Level 5-Manipulate Cluster Codas.

phonological awareness tasks (rhyme, detect oddity, blend, segment, substitute), and linguistic unit (syllable, body-coda, onset-rime, and phoneme).

Concurrent validation with teacher ratings. A second test of construct validity was a comparison of SAPA scores with the simultaneously provided teacher ratings of reading ability for students completing the assessment in the spring of their kindergarten year (see Cassady et al., 2002). Teachers were asked to rate students on a 5-point classification scheme judging their reading skills as compared to "grade level" expectations (well-above grade level to well-below grade level). The teachers making these ratings were well trained in state academic standards that placed premium focus on the acquisition of phonological awareness skills during the kindergarten year. There was a meaningful positive correlation between teacher ratings of student ability and SAPA total score, $r = .67$, $p < .001$, $n = 121$. Significant, positive correlations were repeated for each of the 13 subscales as well.

An alternative method of viewing this relationship between teacher ratings and SAPA performance is possible through analyses of differences among teacher-defined groups. A multivariate analysis of variance (MANOVA) was conducted to verify overall differences among the groups on the 13 original subscales. The MANOVA's omnibus effect was significant, $F(52, 428) = 2.70$, $p < .001$, $\eta^2 = .25$, with statistically significant differences among the 5 groups for 12 of the 13 original subscales. Only the Oddity Tasks-Middle Sounds subtest demonstrated no meaningful differences, apparently due to a combination of two factors. First, there was a small range in the scores among the 5 groups on this subscale, with means progressively growing from 1.67 for the well-below average group to 2.60 for the well-above average group. It is possible that the subscale is merely not sensitive enough to detect the minor differences between the groups on this domain. This low-level of sensitivity appears to be caused by the overall difficulty of the items on this subtest. Students at the end of kindergarten may have not yet mastered detection of phonemic differences for middle units, which is consistent with the state curriculum standards that do not call for mastery of this skill until the end of first grade (see Cassady & Smith, 2004a for related discussion). Second, lack of power provided by small sample sizes in the group cells inhibits detection of weak to moderate effects in these analyses.

Examination of the total score was conducted for a simplified validation of demonstrating separation among identified ability groups. Scheffe's post-hoc pairwise comparisons demonstrated that the readers identified by their teachers as well-below and below average performed significantly worse on the SAPA than the other three groups (p's $< .001$), but did not differ significantly from one another. While the average and above average groups did not differ significantly from one another, the well-above average reading group did outperform these two groups (p's $< .001$).

Concurrent validation with standardized reading achievement tests. The availability of standardized test performances for a subset of the sample allowed further validation of the SAPA. For 90 of the 135 spring kindergarten participants discussed in the teacher rating analyses, first-grade CTBS Terra Nova reading, language arts, and total scores were available. Correlational analyses revealed strong, positive correlations among the SAPA total scores and Terra Nova Language ($r = .69$), Reading ($r = .58$), and Total composites ($r = .73$). An exploratory set of correlational analyses on 10 first grade children taking the SAPA and Terra Nova in the spring of their first grade year provided similar supportive outcomes (r's $> .85$).

Concurrent validation with popular phonological awareness measures. To provide a more direct and meaningful analysis of validity, we provided a direct test of concurrent validity of the SAPA as compared to the Emergent Literacy Survey (ELS, Pikulski, 1999) and phonological awareness subtests of the Dynamic Indicators of Early Literacy Skills (DIBELS, Good & Kaminski, 2002). Planned correlational analyses were used to explore the level of agreement shared with the SAPA and these established measures (see Table 3). As shown in the table, moderate to strong effects were demonstrated in the comparisons. This is particularly important to the assertion that subscale tasks from the 14 discrete tasks can be derived to mimic the broader reading measures offered in the past (see Cassady & Smith, 2003; 2004a; 2004b).

Reliability

The first issue of reliability we were concerned about was ensuring that the SAPA total score was a reasonable assessment of one broad construct of phonological awareness skill despite the use of 14 subscales. To test this, we first employed Cronbach's alpha statistic to estimate the internal consistency reliability on the original 13-subscale version. The result demonstrated that the SAPA produced patterns of responses that were highly consistent, $\alpha = .93$ (Cassady et al., 2002). Further exploration of each individual subscale revealed high levels of internal consistency for each subsequent subscale, as would be expected given the high overlap in skills for each item. In initial exploration of the items, those that were found to detrimentally impact the level of consistency within a particular subscale were removed and replaced with items that did not produce the same psychometric barriers. This revision process was focused primarily on acoustically problematic words such as the often confusing "r-controlled" words that do not follow standard phonic relationships by obscuring the separation between the vowel and the ending sound (e.g., car, purr, whir).

Table 3. *SAPA Subtest Correlations with Associated DIBELS & ELS Subscores*

	Emergent Literacy Survey						DIBELS		
SAPA Subtest	Rhyme	Blending Onset-Rimes	Beginning Sounds	Segmenting Onset-Rimes	Phoneme Blending	Phoneme Segmenting	Initial Sound Fluency	Phoneme Segmenting Fluency	Nonsense Word Fluency
Rhyme recognition									
Rhyme application	.51								
Oddity tasks: Beginning sounds			.67				.41		
Oddity tasks: Ending sounds									
Oddity tasks: Middle sounds									
Blending body-coda		.72			.70				
Blending onset-rimes		.64			.73				
Blending phonemes		.71			.73				
Segmenting onset-rimes		.		.43		.50	.61	.51	
Segmenting phonemes			.52	.53		.78		.64	
Phoneme deletion			.48	.76		.65		.51	
Phoneme sub: Beginning sounds			.36				.59		
Phoneme sub: Ending sounds									
Phoneme sub: Middle sounds									

Note. All values equal to or less than $r = .43$ are $p < .01$; all values greater than $r = .43$ $p < .001$

A second test of the structure of the SAPA was an exploratory factor analysis. The results from the principal components analysis demonstrated a dominant first factor that explained 54% of the variance (eigenvalue = 7.14). Maintaining the simple convention of accepting any factor with an eigenvalue exceeding 1.0 would lead to the acceptance of one other factor that explained an additional 10% of the variance (eigenvalue = 1.21). However, examination of the scree plot and eigenvalues in the exploratory factor analysis did not support employing that interpretational strategy. Furthermore, the values obtained in the varimax rotation component matrix revealed that the second factor was merely documenting a level of difficulty in the subscales. That is, the subscales that loaded on the second factor were simply the most difficult ones. This pattern supports two propositions. First, the SAPA meets criteria allowing for use as a reliable single-factor measure of phonological awareness. Despite our theoretical position that educators are best served by the discrete information afforded in examining the subtest performances, state and federal mandates generally seek global indicators of performance on phonological awareness, so the composite score is practically significant. Second, the presence of a weak second factor in the exploratory factor analysis supports our proposition that there are developmental gains across the scale (see Cassady et al, 2002).

The primary benefit provided by the SAPA with respect to reliability is the control over the standardization of presentation methods. The standard presentation of pre-recorded digitally-controlled spacing in the phonological stimuli overcomes the problem encountered with both the DIBELS and ELS being administered by multiple individuals. For instance, use of schwa in blending tasks is known to affect the performance of young children in phonological awareness screening measures (Murray, Brabham, Villaume, & Veal, 2002). Relying on each individual administrator to accurately "cut" the schwa ("mmm" – "ix") or leave it intact ("muh" – "ix") correctly on each item is risky. Furthermore, we have control over the spacing among all phonemic units with the SAPA. In our assessment, all phonemic units are separated with a 1 second interval. This spacing is controlled and consistent across all administrators, items, and individuals' experiences. Spacing the three phonemes in a standard C-V-C phoneme blending task at 1 sec intervals evenly ensures that the task measured is truly phoneme blending, as opposed to the errors we have observed in administration of the ELS and other assessments where test administrators present the stimuli with irregular spaces. For example, if one tester provides "CAT" in such a way as to provide an abbreviated pause after the middle sound (a common mistake when working quickly), it is likely that the task becomes more similar to a blending onset-rime activity (/c/ /at/ rather than /c/ /a/ /t/). Clearly, accuracy in measurement and confidence in the validity of data-driven conclusions are lost whenever there are such variations in the assessment materials.

Procedural Benefits

Our attempt to identify the advantages and liabilities to using a computer-assisted phonological awareness tool has presented three themes as central advantages. Test administrators who delivered both the SAPA and at least one other phonological awareness test delivered in a traditional format (including the original version of the SAPA that was delivered orally) provided feedback on the process of testing emergent readers. The test administrators reported advantages in training, pragmatics of testing, and greater confidence in the accuracy of testing procedure.

Training advantages. To directly assess the method of presentation, we interviewed test administrators who delivered the SAPA items orally and those who used the SAPA on the computer. Our discussions revealed that for computer-literate individuals training time for the multimedia version of the SAPA was dramatically shorter than for the oral presentation. Indeed, those individuals who were able to access the testing materials by double-clicking on the SAPA icon were able to be proficient in test administration within 15 minutes. Conversely, training the test administrators the appropriate pronunciation and spacing for over 100 items (all practice and test items for the 14 subtests) generally required 2 1-hour training sessions accompanied by follow-up sessions to verify procedures and answer questions.

An additional benefit of the simplified process of delivering the test content that we have confirmed in our use of the SAPA is that the level of expertise required to confidently deliver the testing materials is lower for the multimedia presentation format employed by the SAPA. Many tests of emergent literacy skills are simply too complex for even teachers or classroom aides to deliver without involved professional development and technical assistance (Caldwell, 2002). The use of technical jargon (however basic or simple) in traditional administration instructions hampers the average user from being able to meet stringent administration policies. However, relaxing the technical precision in test administration inserts several threats to test validity and reliability. All these concerns have been alleviated with the multimedia presentation, as the computer provides consistent presentation of all testing materials every time. To test this prospect, undergraduate pre-service teaching majors were asked to learn to use the SAPA. The junior-level future teachers had no difficulty accessing the items and were successful in learning the rules of administration within the standard 15-minute training period.

Test administration facilitation. Test administrators reported satisfaction with the SAPA with regard to easing the labor of phonological awareness testing, finding the process to be less "jumbled." Contrasted with traditional testing procedures, the SAPA requires no ancillary materials (e.g., picture cards or stopwatch) that are common in phonological awareness testing.

Consistent with the proposed advantage of CoPS (Singleton et al., 2000), administrators also claimed they were able to more efficiently move from one student to the next with the multimedia version. As the equivalent of a menu-driven DVD, starting a new student does not require reorganizing materials or turning back through a test booklet to find the correct starting point. All testing commences from the main menu screen which is always active.

Consistency in testing procedures. Without doubt, the single greatest contribution offered by the SAPA's multimedia presentation format is the consistency gained across individual testing sessions. Our primary concerns with traditional phonological awareness assessment rests in the pronunciation and pacing of the test stimuli. Standardized testing procedures require consistency in order to ensure that comparisons made among students are reasonable. With every additional test administrator that is used to orally deliver phonological awareness test materials, there is an additional level of variation in the test materials. Thus, there is a lower degree of confidence that the data are meaningful and accurate when comparisons are made.

The test administrators using the SAPA reported being more confident that they were "doing it right." Traditional tests of phonological awareness are often complex and be confusing. Those test administrators using both the SAPA and DIBELS found the DIBELS to be significantly more difficulty to learn and deliver. This difficulty attribution arose primarily from the rules underlying timed administration and how to determine final scores. Our experience with the DIBELS in particular demonstrated that the instructions, manuals, and online materials were all necessary for the administrators to gain confidence and skill in delivery. However, we continued to see instances of inconsistency in administration when looking to issues such as pace, tone, and use of phonemic conventions that are not addressed explicitly on most administration manuals. For instance, the DIBELS materials do not provide clear indication to the administrator whether to provide the blending task items with the schwa sound (e.g., "/muh/ /at/" or "/mmm/ /at/"). This is clearly an important phonological point, as Murray et al. (2002) reported that presenting onset-rime blending stimuli with the schwa was consistently easier for students to complete than when the schwa was removed. Further evidence of the importance of this administrative technique comes from the published scoring criteria for the DIBELS (Good & Kaminski, 2002), which explicitly allows for children to receive credit for responses that have the additional phonemic content in segmentation tasks.

This lack of specificity in the test administration guidelines is by no means exclusive to the DIBELS. In a review of several phonological awareness test protocols, we found none that provided detailed information on the pace of deliver for blending tasks (defined as 1-sec in our materials), explicit directions for use of schwa, and detailed pronunciation guidelines. In fact, users of the DIBELS are provided with a pronunciation guide for the phonemic units to be delivered, but are instructed that regional dialects may dictate deviation from those standard pronunciations. Naturally, dialectic differences can lead to problematic results as children's performance levels may be suppressed when they are presented with auditory pronunciations that do not match their standard expectation for the target words. However, this threat exists at an exponentially higher level when the dialectical variation is left to each individual administrator.

CONCLUSIONS

The presence of phonological awareness assessment in educational settings is likely to continue to increase, as is the scrutiny of the assessment methods used to demonstrate proficiency in this set of foundational reading skills given the level of federal funding tied to demonstrating successful gains in this domain. It is our assertion that it is necessary that educators, evaluators, and researchers make use of phonological awareness measures that address the two primary threats to validity we have discussed here. That is, phonological awareness measures should focus on discrete tasks to provide more precise measurement of the development of phonological processing skills using a protocol that eliminates variability across testing situations. Again, although we propose that our own assessment tool meets these criteria, the SAPA is provided in this analysis as a mere example. Any existing measure of phonological awareness could be revised to overcome the second threat to validity (standard presentation) offered in our discussion.

REFERENCES

Adams, M. J., Foorman, B. R., Lundberg, I., & Beeler, T. (1998). *Phonemic awareness in young children*. Baltimore: Brookes.

Bracken, B. A., & McCallum, R. S. (1998). *The universal nonverbal intelligence test.* Itasca, IL: Riverside Publishing.

Caldwell, J. S. (2002). *Reading assessment: A primer for teachers and tutors.* New York: Guilford.

Cassady, J. C., Smith, L. L., Bauserman, K., Jordan, F., Walker, C. A., & Popplewell, S. R. (2002). *Developmental Models of Phonological and Phonemic Awareness: A Comparison and Reformulation.* Paper presented at the National Reading Conference 2002 Annual Meeting, Miami, FL.

Cassady, J. C., & Smith, L. L. (2003). Development of phonological awareness: The trouble with middle sounds. *National Reading Conference Yearbook, 52,* 139-149.

Cassady, J. C., & Smith, L. L. (2004a). Acquisition of blending skills: Comparisons among body-coda, onset-rime, & phoneme blending tasks. *Reading Psychology, 25(4),* 261-272.

Cassady, J. C., & Smith, L. L. (2004b). The impact of a reading-focused integrated learning system on phonological awareness in kindergarten. *Journal of Literacy Research35,* 947-964.

Ehri, L., Nunes, S., Willows, D. Schuster, B. B., Yaghoub-Zadeh, Z. & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly, 36,* 250-87.

Gibbs, S. (2003). Do pictures make a difference? A test of the hypothesis that performance in tests of phonological awareness is eased by the presence of pictures. *Educational Psychology in Practice, 19,* 219-228.

Good, R. H., & Kaminski, R. A. (Eds.) (2002). *Dynamic indicators of basic early literacy skills* (6th Ed.). Eugene, OR: Institute for the Development of Education Achievement. Available http://dibels.uoregon.edu/. Accessed February 6, 2005.

Gordinier, C. L., & Foster, K. (2004). What stick is driving the Reading First hoop? *Childhood Education, 81,* 940.

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

Lane, H. B., Pullen, P. C., Eisele, M. R., & Jordan, L. (2002). Preventing reading failure: Phonological awareness assessment and instruction. *Preventing School Failure, 46,* 101-106.

Murray, B. A., Brabham, E. G., Villaume, S. K., & Veal, M. (2002, December). The effect of three segmentation options on ease of blending for prealphabetic and partial alphabetic readers. Paper presented at the 2002 National Reading Conference, Miami, FL.

Puolakanaho, A., Poikkeus, A., Ahonen, T., Tolvanen, A., & Lyytinen, H. (2003). Assessment of three-and-a-half-year-old children's emerging phonological awareness in a computer animation context. *Journal of Learning Disabilities, 36,* 416-423.

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

Sodoro, J., Allinder, R. M., & Rankin-Erickson, J. L. (2002). Assessment of phonological awareness: Review of methods and tools. *Educational Psychology Review, 14,* 223-260.

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.

Thomas-Tate, S., Washington, J., & Edwards, J. (2004). Standardized assessment of phonological awareness skills in low-income African-American first graders. *American Journal of Speech-Language Pathology, 13,* 182-190.

Treiman, R. (1985). Onsets and rimes as units of spoken syllables: Evidence from children. *Journal of Experimental Child Psychology, 39,* 161-181.

Treiman, R., & Zukowski, A. (1996). Children's sensitivity to onses, rimes, and phonemes. *Journal of Experimental Child Psychology, 61,* 193-215.

Yopp, H.K. & Yopp, R.H. (2000). Supporting phonemic awareness development in the classroom. *The Reading Teacher, 54,* 130-143.

Note

The contents of this article were developed under a subgrant from the U.S. Department of Education. However, such contents do not necessarily represent the policy of the U.S. Department of Education, and you should not assume endorsement of the Federal Government, Project Funding Authority; Part B, IDEA Project Number 14204-001-DY08.

Citation

Cassady, Jerrell C., Smith, Lawrence L. & Huber, Linda K. (2005). Enhancing Validity in Phonological Awareness Assessment through Computer-Supported Testing. *Practical Assessment Research & Evaluation*, 10(18). Available online: http://pareonline.net/getvn.asp?v=10&n=18

Corresponding Author

Jerrell C. Cassady
Associate Professor of Psychology
Department of Educational Psychology
Ball State University
Muncie, IN 47306
765-285-8522
jccassady@bsu.edu