Journal of Psychoeducational Assessment
Volume 26 Number 3
September 2008 287-298

10.1177/0734282907306048
http://jpa.sagepub.com
hosted at
http://online.sagepub.com

Psychometric Evaluation of the Mountain Shadows Phonemic Awareness Scale With a Kindergarten Sample

Jason M. Nelson
University of Montana

This study investigated the psychometric properties of a group-administered early literacy measure, the Mountain Shadows Phonemic Awareness Scale (MS-PAS), using a kindergarten sample ($N = 213$). The MS-PAS was compared to the *Test of Phonological Awareness–Second Edition: Plus* (TOPA-2+). Results indicated excellent internal consistency for the MS-PAS, as well as adequate convergent and predictive validity. The MS-PAS was found to be an adequate measure for determining which students are unlikely to need additional instructional services in reading. The MS-PAS and TOPA-2+ had highly similar psychometric properties. Recommendations are provided for using the MS-PAS in a multistep screening process.

Keywords: *Mountain Shadows; phonological awareness; phonemic awareness; early literacy; reading screening*

Over the past three decades, researchers have discovered that the process of learning to read begins long before children enter school and before they are exposed to print (McCardle, Scarborough, & Catts, 2001). According to Torgesen (2002), one of the most important findings researchers have discovered during this time frame has been the impact of phonological awareness on reading development. Phonological awareness, which is the ability to sense and manipulate the sound structure of language, is one of the best predictors of later reading skills (Scarborough, 1998). Although a variety of factors are influential in the development of reading problems, phonological awareness difficulties are regarded as the primary deficit experienced by struggling readers (Stanovich & Siegel, 1994).

Once children fall behind in reading, returning them to a normal reading developmental trajectory is arduous. After about the age of 9 years, reading remediation efforts have limited success unless highly intensive instruction is provided (Lyon et al., 2001). On the other hand, preventive efforts and early reading intervention have shown more promising results. Several studies (e.g., Foorman, Francis, Fletcher, Schatschneider, & Mehta, 1998) have indicated that the majority of students who struggle in the early grades can be brought to

Author's Note: Correspondence concerning this article should be addressed to Jason M. Nelson, Department of Psychology, University of Montana, Skaggs Building 203, Missoula, MT 59812; e-mail: jason.nelson@umontana.edu.

average levels of word reading skills if they are provided early intervention in the form of phonemically explicit instruction.

The first step to prevention is early identification. A variety of early reading instruments have proliferated in response to the need for early identification. One such measure is the Mountain Shadows Phonemic Awareness Scale (MS-PAS; Watkins & Edwards, 1998). The MS-PAS is a group-administered early literacy measure that is available to the public free-of-charge via the World Wide Web at http://www.personal.psu.edu/mww10/Watkins3.html.

Remarkably similar to the MS-PAS is the Phonological Awareness Scale of the *Test of Phonological Awareness–Second Edition: Plus* (TOPA-2+; Torgesen & Bryant, 2004). The TOPA-2+ Phonological Awareness Scale consists of two sound categorization tasks that incorporate 20 items in which students look at pictures to determine objects that begin with the same or different phoneme as a word spoken by the examiner. Apart from using pictures of different objects, the 20 items of the TOPA-2+ Phonological Awareness Scale are exactly the same as those on the MS-PAS. A major difference between the TOPA-2+ and the MS-PAS is that the former must be purchased. In addition, the TOPA-2+ includes an extra subtest that measures letter-sound knowledge.

The purpose of the current study was to investigate the psychometric properties of the MS-PAS compared to those of the TOPA-2+ using a kindergarten sample. More specifically, the current study investigated the internal consistency and convergent, predictive, and classification validity of the two instruments. Previous published research (Watkins & Edwards, 2004) investigated the psychometric properties of the MS-PAS using a first-grade sample. Several authorities (e.g., Adams, 1990) have urged that students at risk for reading failure be identified in kindergarten. Consequently, information about the psychometric properties of the MS-PAS with this age group is important. Furthermore, no published studies have investigated the MS-PAS and TOPA-2+ concurrently to determine whether one is psychometrically superior to the other. Given the highly similar composition of the two instruments, it is possible they possess generally equivalent psychometric properties. Because school systems are often limited by budgetary constraints, it is important to know whether an instrument that is free-of-charge performs as adequately as one that is of substantial cost when used with large groups of students.

Method

Participants

Participants were 213 kindergarten students enrolled in a midsized suburban public school system in a midwestern state. They were drawn from 10 classrooms and two schools. Participants' mean age was 5.45 years (SD = .52). The sample's ethnicity breakdown was as follows: White (89.7%; n = 191), African American (2.8%; n = 6), Hispanic (4.2%; n = 9), Asian American (.5%; n = 1), and Other (2.8%; n = 6).

Instruments

MS-PAS. The MS-PAS is an early literacy measure developed for use with kindergarten and first-grade students. Two scales, Same Initial Sounds (SIS) and Different Initial Sounds (DIS), make up the MS-PAS. The raw scores from these scales are summed to generate the Total Phonemic Awareness (TPA) score. Both the SIS and DIS tasks measure phonemic awareness by way of sound categorization tasks.

Watkins and Edwards (2004) found an alpha coefficient of .89 and test–rest reliability coefficient of .74 (12-16 week interval) for the MS-PAS. In addition, these researchers found predictive validity evidence for the MS-PAS, because it was moderately correlated with the Gates-MacGinitie Reading Test at the end of first grade and the Stanford Achievement Test at the end of second to sixth grade.

TOPA-2+. The TOPA-2+ is a group-administered, norm-referenced early literacy measure which consists of two versions, the Kindergarten version and the Early Elementary version. Only the Kindergarten version was used in the current study, and, therefore, is the only version described here. Two scales, Phonological Awareness and Letter Sounds, make up the TOPA-2+ Kindergarten version. The Phonological Awareness Scale is made up of the Initial Sound–Same (IS-S) and Initial Sound–Different (IS-D) tasks, which consist of the same type of items as those on the MS-PAS. The Letter Sounds subtest was excluded from the current study because it has no equivalent on the MS-PAS.

Torgesen and Bryant (2004) provided strong psychometric evidence for the TOPA-2+. The mean alpha and test–retest reliability coefficients for the Phonological Awareness Scale at the kindergarten level were .91 and .87, respectively. As evidence of validity, Torgesen and Bryant reported moderate-to-strong correlations between the TOPA-2+ and a variety of instruments purporting to measure the same or similarly related constructs.

Dynamic Indicators of Basic Early Literacy Skills (*DIBELS*). The kindergarten DIBELS (Good & Kaminski, 2002) are a collection of five, 1-minute probes that measure early literacy skills such as letter naming and phonemic segmentation. Four of the five DIBELS kindergarten measures were administered, including Initial Sound Fluency (ISF), Letter Naming Fluency (LNF), Phoneme Segmentation Fluency (PSF), and Nonsense Word Fluency (NWF).

Several published studies (e.g., Elliott, Lee, & Tollefson, 2001) on the DIBELS have indicated acceptable levels of test–retest, alternate forms, and interrater reliability. Research on the DIBELS has also indicated evidence of concurrent and predictive validity (Elliott et al. 2001; Kaminski & Good, 1996).

Texas Primary Reading Inventory-Phonemic Awareness (*TPRI-PA*) *screening test.* The TPRI-PA (Texas Education Agency and the University of Texas System, 2004-2006) consists of eight items that require students to blend phonemes. Evidence of internal consistency (r = .92) and concurrent validity with a range of phonological, orthographic, and reading tasks has been established for the kindergarten TPRI-PA (Foorman et al., n.d.).

Woodcock Johnson Tests of Achievement—Third Edition (WJ III) *Letter Word Identification (LW) and Word Attack (WA) subtests.* The WJ III (Woodcock, McGrew, & Mather, 2001) is a norm-referenced, standardized, individually administered test of multiple academic achievement domains. Only word reading and reading-related skills were of interest in the current study; therefore, only the LW and WA subtests were administered. The LW subtest measures skills in identifying letters, understanding letter–sound correspondence and recognizing real words. Examinees read nonsense words on the WA subtest. Difficulties on nonsense word reading tasks are hallmark deficits of many students with reading disabilities (Rack, Snowling, & Olson, 1992).

McGrew and Woodcock (2001) reported median alpha coefficients of .94 and .87 for the LW and WA subtests, respectively. Furthermore, these authors reported median test–retest reliability coefficients of .95 and .83 for the LW and WA subtests, respectively, when a 1-year interval between testing was used. Finally, McGrew and Woodock found evidence for the concurrent validity of the WJ III Basic Reading Skills Cluster, which is made up of the LW and WA subtests, because it was moderately to highly correlated with the Kaufman Test of Educational Achievement Reading Decoding Scale (r = .66) and the Wechsler Individual Achievement Test Basic Reading Scale (r = .82).

Classification validity is determined by calculating diagnostic accuracy statistics (e.g., sensitivity, specificity, and positive predictive power). In studies of diagnostic accuracy, screeners are judged in reference to a "gold standard," which is considered to be the best available evidence for the existence of a particular condition or characteristic (Kessel & Zimmerman, 1993). The screener in such studies is referred to as the index test, whereas the "gold standard" is referred to as the reference standard.

The LW subtest was used as the reference standard for determining the classification validity of the MS-PAS and TOPA-2+. When selecting a reference standard, two major decisions are made, including the selection of a measure that is suitable for judging the adequacy of the screening instrument and the determination of the level of performance on the reference standard that differentiates between adequate and inadequate reading (Jenkins, 2003). The LW subtest was selected as the current study's reference standard because of its excellent psychometric properties. Jenkins stated, "In the field of early intervention, the family of achievement tests developed by Richard Woodcock and associates come closest to a 'gold standard' criterion test of reading ability" (p. 2). A standard score of 90 was set as the performance standard on the LW subtest for distinguishing between adequate and inadequate reading skills. Several researchers (e.g., Speece, Mills, Ritchey, & Hillman, 2003) have used this cutoff as a performance standard in reading research.

Examiner Training

Six graduate students in school psychology were trained for 2 hrs on the administration and scoring of the MS-PAS and TOPA-2+ prior to data collection. Four examiners (two advanced psychology undergraduates and two graduate students in school psychology) were trained for 1 hr on the administration of the WJ III LW and WA subtests. To participate in the study, examiners were required to achieve 95% scoring accuracy with a preestablished scoring protocol prior to participating in the study. More specifically, the

examiners administered the measures to a graduate assistant who made a number of predetermined errors. Each examiner's protocol was then judged against the scored protocols with the predetermined errors. Examiner accuracy was calculated by dividing examiners' total number of correctly scored items by the total number of items and multiplying by 100. The average accuracy of the examiners was 100% for the three instruments.

Eight examiners (six advanced psychology undergraduates and two graduate students in school psychology) each received 12 hrs of training on the DIBELS and TPRI-PA. Using the same procedure mentioned above, the average scoring accuracy for the eight examiners was 99% for the DIBELS and 100% for the TPRI-PA.

Procedure

Testing occurred in two phases. The MS-PAS, TOPA-2+, DIBELS, and TPRI-PA were administered in mid-January, which is consistent with the time frame recommended for kindergarten reading screening (Snow, Burns, & Griffin, 1998). During a period of 2 days, the MS-PAS and TOPA-2+ were administered in small groups of two to five students. The administration of these instruments was counterbalanced, with half the participants taking the MS-PAS on the 1st day and the other half taking the TOPA-2+. On Day 2, those who took the MS-PAS the 1st day took the TOPA-2+ and vice versa. The DIBELS and TPRI-PA were then individually administered to participants over 5 days. During the second phase of testing (conducted in mid-May), each participant was individually administered the WJ III LW and WA subtests.

Data Analyses

Reliability coefficients and correlations for determining convergent and predictive validity were calculated using SPSS version 14. Computer software developed by Watkins (2002) was used to calculate a variety of diagnostic accuracy statistics. The following describes each diagnostic accuracy statistic as discussed by Kessel and Zimmerman (1993): (a) Sensitivity indicates the percentage of participants with reading problems according to the reference standard (WJ III), who are identified as at-risk on the index tests (MS-PAS and TOPA-2+), (b) Specificity indicates the percentage of participants identified as having adequate reading skills on the WJ III, who are identified with the MS-PAS and TOPA-2+ as being at low risk, (c) Positive predictive power indicates the percentage of participants identified by the MS-PAS and TOPA-2+ as at-risk who are classified by the WJ III as having inadequate reading skills, (d) Negative predictive power indicates the percentage of participants identified by the MS-PAS and TOPA-2+ as at low risk, who are classified by the WJ III as having adequate reading skills, (e) False positive rate is the percentage of participants who are identified by the MS-PAS and TOPA-2+ as being at risk but who are classified as having adequate reading skills on the WJ III, (f) False negative rate is the percentage of participants who are identified as at low risk by the MS-PAS and TOPA-2+ but who are classified as having inadequate reading skills on the WJ III, (g) Correct classification rate indicates the proportion of adequate and inadequate readers correctly classified by the MS-PAS and TOPA-2+, and (h) Kappa indicates the level of agreement beyond random chance between the MS-PAS and TOPA-2+ and the WJ III.

Table 1
Correlations With TOPA-2+

	MS-SIS	MS-DIS	MS-TPA	TOPA-ISS	TOPA-ISD
MS-DIS	.74				
MS-TPA	.92	.94			
TOPA-ISS	.72	.66	.75		
TOPA-ISD	.71	.78	.81	.74	
TOPA-PA	.77	.79	.84	.90	.95

Note: MS-SIS = Mountain Shadows Same Initial Sounds scale; MS-DIS = Mountain Shadows Different Initial Sounds scale; MS-TPA = Mountain Shadows Total Phonemic Awareness (overall) scale; TOPA-ISS = Test of Phonological Awareness Initial Sounds Same scale; TOPA-ISD = Test of Phonological Awareness Initial Sounds Different scale; TOPA-PA = Test of Phonological Awareness Phonological Awareness (overall) scale.

Results

Reliability

To investigate the internal consistency reliability of the MS-PAS and TOPA-2+, Cronbach's alpha was calculated. An alpha coefficient of .91 was found for the overall scale of the MS-PAS and .90 for that of the TOPA-2+.

Validity

Convergent validity. The correlations displayed in Table 1 indicate the relationship between the scales of the MS-PAS and TOPA-2+. A high correlation (r = .84) was found between the overall scales of the two instruments.

As shown in Table 2, both instruments showed moderate correlations with other instruments purporting to measure similar constructs (i.e., the DIBELS and TPRI-PA). Not surprisingly, slightly lower correlations were found between the two instruments and the DIBELS LNF task. This finding is to be expected given the composition of the LNF task. Unlike the other DIBELS tasks, LNF is not a measure of phonological awareness but of letter-naming skills. Both the MS-PAS and TOPA-2+ exhibited highly similar correlations with the measures used to investigate convergent validity.

Predictive validity. Table 2 indicates that highly similar correlations were also found between the two instruments and those administered at the end of kindergarten (i.e., the WJ III LW and WA subtests). Moderate correlations were found between the MS-PAS and TOPA-2+ and the measures used to examine predictive validity. Neither instrument outperformed the other in terms of ability to predict later real nor nonsense word recognition.

Classification validity. As displayed in Table 3, four cutoff scores (15-18) of the MS-PAS and TOPA-2+ were examined to determine the accompanying diagnostic accuracy of

Table 2
Correlations With DIBELS, TPRI, and WJ III

	ISF	LNF	PSF	NWF	TPRI-PA	WJ LW	WJ WA
MS-PAS	.52	.48	.53	.58	.61	.57	.51
TOPA-2+	.56	.46	.54	.57	.59	.59	.54

Note: MS-PAS = Mountain Shadows Phonemic Awareness Scale; TOPA-2+ = *Test of Phonological Awareness–Second Edition: Plus*; ISF = Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Initial Sound Fluency; LNF = DIBELS Letter Naming Fluency; PSF = DIBELS Phoneme Segmentation Fluency; NWF = DIBELS Nonsense Word Fluency; TPRI-PA = Texas Primary Reading Inventory-Phonemic Awareness screening test; WJ LW = Woodcock Johnson Tests of Achievement Letter Word Identification subtest; WJ WA = Woodcock Johnson Tests of Achievement Word Attack subtest. ISF, LNF, PSF, NWF, and TPRI-PA were administered in mid-January (i.e., concurrently with the MS-PAS and TOPA-2+). WJ LW and WJ WA were administered in mid-May.

each. Watkins and Edwards (2004) recommended a cutoff score of 17 when using the MS-PAS with first-grade students. They found this cutoff score produced a true positive rate of 94% and a false positive rate of 32%. Results of the current study indicated that this cutoff score produced a true positive rate (represented by sensitivity in Table 3) of 93% but a substantially higher false positive rate (55%). Additional analysis of this cutoff score indicated high negative predictive power (.99) but low positive predictive power (.12). This means that 99% of those who scored higher than 17 on the MS-PAS also scored higher than 90 on the WJ III LW subtest. Only 12% of those who scored 17 or lower on the MS-PAS also scored 90 or below on the WJ III LW subtest. The Kappa of .09 is indicative of low overall agreement beyond random chance between the MS-PAS and the WJ III.

Because the current study incorporated a kindergarten sample, it was necessary to investigate a higher cutoff score than 17. Increasing the cutoff score on the MS-PAS to 18 did not improve sensitivity and decreased specificity (see Table 3). The false positive rate for this cutoff score was 62%. Furthermore, improvements were not made in any of the other diagnostic accuracy statistics.

Lower cutoff scores were also investigated because of the substantial false positive rates of the above-mentioned cutoff scores. Examination of cutoff scores of 15 and 16 indicated a decrease in sensitivity but a slight reduction in false positive rate (see Table 3). False negative rates increased to 29% when a cutoff score of 15 was used as opposed to a false negative rate of 7% when a cutoff score of 17 was used. By and large, the other diagnostic statistics were not substantially affected by the change in cutoff scores.

Regarding the comparison of the MS-PAS and TOPA-2+, Table 3 shows that the examination of both instruments produced very similar diagnostic accuracy statistics. The most apparent difference between the measures resulted when a cutoff score of 15 was used. The TOPA-2+ at this cutoff score produced the same percentage of false positives (43%) as the MS-PAS but fewer false negatives (17% compared to 29%).

Table 3
Diagnostic Accuracy Statistics Using WJ III as Reference Standard

	MS (15)	TOPA (15)	MS (16)	TOPA (16)	MS (17)	TOPA (17)	MS (18)	TOPA (18)
Sensitivity	.71	.83	.86	.83	.93	.92	.93	1.0
Specificity	.57	.57	.53	.52	.45	.48	.38	.37
False positive rate	.43	.43	.47	.48	.55	.52	.62	.63
False negative rate	.29	.17	.14	.17	.07	.08	.07	.00
Positive predictive power	.12	.11	.13	.10	.12	.10	.11	.10
Negative predictive power	.96	.98	.98	.98	.99	.99	.99	1.0
Overall correct classification	.58	.59	.56	.54	.49	.50	.42	.41
Kappa	.08	.10	.11	.08	.09	.08	.07	.07

Note: MS = Mountain Shadows Phonemic Awareness Scale; TOPA = *Test of Phonological Awareness–Second Edition: Plus.*

Discussion

The aim of the current study was to investigate the psychometric properties of the MS-PAS. Specifically, the MS-PAS was compared to the TOPA-2+ in terms of its internal consistency reliability and convergent, predictive, and classification validity. Both instruments are potentially useful for the early identification of students at risk for reading failure, particularly because they can be group administered.

Results indicated that the internal consistency of both instruments was equivalent. The alpha coefficients of .91 and .90 for the MS-PAS and TOPA-2+, respectively, represent excellent reliability (Cicchetti, 1994). Watkins and Edwards (2004) found an alpha coefficient of .89 for the MS-PAS administered in the first grade. Similarly, Worrell, Watkins, Runge, and Hall (2002) found an alpha coefficient of .89 for the MS-PAS administered to first- to third-graders in the Republic of Trinidad and Tobago.

The MS-PAS and TOPA-2+ were strongly correlated ($r = .84$), suggesting that the two instruments measure a highly similar construct. Moderate correlations (.48 to .58) between the MS-PAS and DIBELS tasks provide further validity evidence. The correlations (.46 to .57) between the TOPA-2+ and the DIBELS were nearly identical to those between the MS-PAS and DIBELS.

Both the MS-PAS and TOPA-2+ administered at midkindergarten showed moderate correlations (.57 and .59, respectively), with a measure of real word reading ability at the end of kindergarten. This evidence of predictive validity is similar to that of Watkins and Edwards (2004), who found the MS-PAS administered at the beginning of the first grade moderately correlated ($r = .59$) with the Gates-MacGinitie Reading Test at the end of the first grade. Watkins and Edwards also investigated the ability of the MS-PAS administered in the first grade to predict reading ability in the later grades and found moderate correlations in the second ($r = .54$), third ($r = .51$), fourth ($r = .49$), fifth ($r = .47$), and sixth (.44) grades. As is clear by the trend of these correlations, the relationship between phonological awareness and word reading decreases with time. It is important to keep this trend in mind when comparing the results of the current study with those of large-scale meta-analyses of the

relationship between phonological awareness and word reading, which have found lower mean or median correlations (e.g., $r = .40$ in Hammill, 2004). These meta-analyses incorporated studies that used samples with broad age ranges. For example, the mean age of participants in the Hammill meta-analysis was 10 years with a range of 5-43 years.

Although the establishment of both convergent and predictive validity is valuable, classification validity provides the strongest evidence for the utility of early identification measures (Jenkins, 2003). Watkins and Edwards (2004) recommended using a cutoff score of 17 on the MS-PAS for first graders. In their study, this cutoff score produced a true positive rate of 94% and a false positive rate of 32%. Results from the current study, when this cutoff score was used with kindergarteners, indicate a nearly identical true positive rate (93%) but a higher false positive rate (55%). The lower false positive rate of the Watkins and Edwards study compared to the current study is likely explained by their use of older participants. With age, prediction of later reading outcomes becomes more accurate (O'Connor & Jenkins, 1999). In addition, a different reference standard was used in the Watkins and Edwards study. Jenkins noted that it is common for different diagnostic accuracy statistics to occur, depending on the reference standard used.

Sensitivity (true positive rate) and specificity (true negative rate) can be misleading because these statistics do not take into account the base rate of the predicted attribute (Landau, Milich, & Widiger, 1991). Because positive and negative predictive power statistics do incorporate base rates, they have been deemed more useful than sensitivity and specificity (Kessel & Zimmerman, 1993). Results from the current study indicate evidence of excellent negative predictive power (.99) of the MS-PAS using a cutoff score of 17. Ninety-nine percent of participants who achieved a score above 17 on the MS-PAS exhibited adequate word reading skills at the end of kindergarten; however, this cut-score produced a low positive predictive power index (.12). Positive predictive power of .70 is regarded as acceptable for screening measures (Glascoe et al., 1992). According to McDermott et al. (1995), negative predictive power is indicative of the utility of a test as an exclusion criterion, whereas positive predictive power reveals the utility of a test as an inclusion criterion. Hence, the results of the current study suggest that the MS-PAS using a cutoff score of 17 is excellent at providing guidance as to which students are not in need of secondary interventions but lacking in its ability to identify accurately those who do need more intensive reading instruction.

Three additional cutoff scores (15, 16, and 18) were examined to determine whether diagnostic accuracy could be improved. Using a cutoff score of 18 did not improve sensitivity and produced more false positives. Cutoff scores of 15 and 16 resulted in lower sensitivity but only slightly improved specificity. Positive and negative predictive power remained virtually the same across all cutoff scores examined, that is, positive predictive power was low, and negative predictive power was high. The diagnostic accuracy statistics of the TOPA-2+ for the various cutoff scores were highly similar to those of the MS-PAS.

A possible explanation of the low positive predictive power of the MS-PAS is its inclusion of only one type of phonemic awareness task (i.e., sound categorization using same and different initial sounds). Phonemic awareness is a multidimensional skill best described as being on a continuum from basic to complex. Although initial sound identification and manipulation is a necessary and valuable skill on the continuum, other more sophisticated

phonemic awareness skills such as segmenting all the sounds in words and blending sounds are also highly important in the development of fluent and accurate word recognition (Adams, 1990). Perhaps the addition of a broader range of phonemic awareness tasks would have resulted in higher positive predictive power. The addition of these tasks, however, would not be conducive to group administration.

The results of the current study are consistent with other studies examining the diagnostic accuracy of early reading instruments. In general, predicting good reading outcomes is much easier than predicting poor ones (Scarborough, 1998). Although correlational studies often indicate moderate correlations between early indicators (e.g., phonological awareness) and later reading skills, the majority of the variance is accounted for by good readers (Felton, 1992). Therefore, when the diagnostic accuracy of early reading measures is examined, low positive predictive power and high negative predictive power occurs with regularity. Because positive predictive power is often low, it is also common for early reading measures to produce high false positive rates. For example, O'Connor and Jenkins (1999) found false positive rates ranging from 47% to 70%, when they adjusted their cutoff scores to identify nearly all participants at risk.

As is apparent from the current study and others with similar aims, the business of developing an early reading measure that is diagnostically accurate is a difficult one. Speece (2005) described such pursuits as attempts at "hitting the moving target known as reading development" (p. 487). Of course, the challenges of early assessment are not restricted to the domain of reading. During early childhood, children are often rapidly developing the skills that adults are attempting to measure and are often doing so according to a time line that is not uniform. Accordingly, the assessment of young children is an inherently inaccurate undertaking (Shepard, 1997).

Limitations

The results of this study should be interpreted in light of the following limitations. Generalizability should be considered in relation to the demographic characteristics of the current study's sample. More than 90% of participants were White, and all were drawn from the same area of the country. In addition, the period of time for examining predictive validity was minimal. A final limitation is inherent in all diagnostic accuracy studies because of their incorporation of a "gold standard." The assumption is that the reference standard is 100% accurate; however, as Macmann and Barnett (1999) stated, "Given the indefinite nature of the constructs measured, 'true' criterion status is known only to God (and even She or He may have questions)" (p. 525).

Implications and Recommendations for Practice

The results of the current study indicate that if the MS-PAS is used, the screening process should not end after its administration. Doing so would likely lead to an excessive number of false positive cases, which doubtless would result in a dilution of instructional services for those who truly need intensive, explicit, and systematic reading instruction. The current study provides strong evidence that the MS-PAS has utility as an exclusionary measure because of its adequate negative predictive power but not an inclusionary measure

because of its poor positive predictive power. Perhaps a multistep process could be implemented in which the MS-PAS is used to exclude as a first step those who are not at risk. Following universal screening, more specific and thorough assessment could be conducted with those who are not excluded. These assessments would likely need to be individually administered and could include more sophisticated phonemic awareness tasks. The assessment of other reading-related skills, such as rapid automatized naming and letter identification, may also produce more accurate identification of at-risk readers.

A cutoff score of 17 on the MS-PAS for determining risk status is recommended when used with kindergarteners. Although this cutoff score produced a substantial number of false positives, it also resulted in high sensitivity. Examinations of alternative cutoff scores did not substantially improve the diagnostic accuracy characteristics of the MS-PAS. Providing the test user bears in mind the need to conduct follow-up assessments with those identified as at-risk, the MS-PAS can be a valuable tool within a multistep and multimethod early identification procedure.

References

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*, 284-290.

Elliott, J., Lee, S. W., & Tollefson, N. (2001). A reliability and validity study of the Dynamic Indicators of Basic Early Literacy Skills–Modified. *School Psychology Review, 30*, 33-49.

Felton, R. H. (1992). Early identification of children at risk for reading disabilities. *Topics in Early Childhood Special Education, 12*, 212-230.

Foorman, B. R., Fletcher, J. M., Francis, D. J., Carlson, C. D., Chen, D., Mouzaki, A., et al. (n.d.). *Technical report: Texas Primary Reading Inventory* (1998 ed.). Retrieved January 13, 2007, from http://www.tpri.org/Documents/19971998TechnicalManual.pdf

Foorman, B. R., Francis, D. J., Fletcher, J. M., Schatschneider, C., & Mehta, P. (1998). The role of instruction in learning to read: Preventing reading failure in at-risk children. *Journal of Educational Psychology, 90*, 37-55.

Glascoe, F. P., Byrne, K. E., Ashford, L. G., Johnson, K. L., Chang, B., & Strickland, B. (1992). Accuracy of the Denver-II in developmental screening. *Pediatrics, 89*, 1221-1225.

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

Hammill, D. D. (2004). What we know about the correlates of reading. *Exceptional Children, 70*, 453-468.

Jenkins, J. R. (2003, December). *Candidate measures for screening at-risk students*. Paper presented at the National Research Center on Learning Disabilities Responsiveness-to-Intervention Symposium, Kansas City, MO.

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

Kessel, J. B., & Zimmerman, M. (1993). Reporting errors in studies of diagnostic performance of self-administered questionnaires: Extent of the problem, recommendations for standard presentation of results, and implications for the peer review process. *Psychological Assessment, 5*, 395-399.

Landau, S., Milich, R., & Widiger, T. A. (1991). Predictive power methods may be more helpful for making a diagnosis than sensitivity and specificity. *Journal of Child and Adolescent Psychopharmacology, 1*, 343-351.

Lyon, G. R., Fletcher, J. M., Shaywitz, S., Shaywitz, B., Torgesen, J. K., Wood, F., et al. (2001). Rethinking learning disabilities. In C. Finn, A. Rotherham, & C. Hokanson (Eds.), *Rethinking special education for a new century* (pp. 259-287). Washington, DC: Thomas B. Fordham Foundation.

Macmann, G. M., & Barnett, D. W. (1999). Diagnostic decision making in school psychology: Understanding and coping with uncertainty. In C. R. Reynolds & T. B. Gutkin (Eds.), *Handbook of school psychology* (3rd ed., pp. 519-548.). New York: John Wiley.

McCardle, P., Scarborough, H. S., & Catts, H. W. (2001). Predicting, explaining, and preventing children's reading difficulties. *Learning Disabilities Research and Practice, 16*, 230-239.

McDermott, P. A., Watkins, M. W., Sichel, A. F., Weber, E. M., Keenan, J. T., Holland, A. M., et al. (1995). The accuracy of new national scales for detecting emotional disturbance in children and adolescents. *Journal of Special Education, 29*, 337-354.

McGrew, K. S., & Woodcock, R. W. (2001). *Woodcock-Johnson III technical manual*. Itasca, IL: Riverside.

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

Rack, J. P., Snowling, M. J., & Olson, R. K. (1992). The nonword reading deficit in developmental dyslexia: A review. *Reading Research Quarterly, 27*, 28-53.

Scarborough, H. S. (1998). Early identification of children at risk for reading disabilities: Phonological awareness and some other promising predictors. In B. K. Shapiro, P. J. Accardo, & A. J. Capute (Eds.), *Specific reading disability: A view of the spectrum* (pp. 75-107). Timonium, MD: York Press.

Shepard, L. (1997). Children not ready to learn? The invalidity of school readiness testing. *Psychology in the Schools, 34*, 85-97.

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

Speece, D. L. (2005). Hitting the moving target known as reading development: Some thoughts on screening children for secondary interventions. *Journal of Learning Disabilities, 38*, 487-493.

Speece, D. L., Mills, C., Ritchey, K. D., & Hillman, E. (2003). Initial evidence that letter fluency tasks are valid indicators of early reading skill. *Journal of Special Education, 36*, 223-233.

Stanovich, K. E., & Siegel, L. S. (1994). The phenotypic performance profile of reading-disabled children: A regression-based test of the phonological-core variable-difference model. *Journal of Educational Psychology, 86*, 24-53.

Texas Educationa Agency and the University of Texas System (2004-2006). *Texas Primary Reading Inventory*. Austin, TX: Texas Reading Instruments.

Torgesen, J. K. (2002). The prevention of reading disabilities. *Journal of School Psychology, 40*, 7-26.

Torgesen, J. K., & Bryant, B. R. (2004). *Test of Phonological Awareness–Second Edition*: *Plus*. Austin, TX: Pro-Ed.

Watkins, M. W. (2002). Diagnostic Utility Statistics [Computer software]. State College, PA: Ed & Psych Associates.

Watkins, M. W., & Edwards, V. A. (1998). *Mountain Shadows Phonemic Awareness Scale*. Unpublished assessment instrument.

Watkins, M. W., & Edwards, V. A. (2004). Assessing early literacy skills with the Mountain Shadows Phonemic Awareness Scale (MS-PAS). *Journal of Psychoeducational Assessment, 22*, 3-14.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement*. Itasca, IL: Riverside.

Worrell, F. C., Watkins, M. W., Runge, T. J., & Hall, T. E. (2002). Pre-reading skills in Trinidad and Tobago students in the first three years of school. *Caribbean Curriculum, 9*, 1-19.