Journal of Psychoeducational Assessment
2004, 22, 3-14

ASSESSING EARLY LITERACY SKILLS WITH THE MOUNTAIN SHADOWS PHONEMIC AWARENESS SCALE (MS-PAS)

Marley W. Watkins
Pennsylvania State University

Vicki A. Edwards
Deer Valley Unified School District
Phoenix, Arizona

Considerable evidence suggests that phonemic awareness is associated with the development of skilled reading. Consequently, it is recommended that beginning readers be assessed to ensure adequate development of phonemic awareness skills. When choosing an assessment method, reliability and validity, ease of administration and scoring, and cost-effectiveness should be considered. To meet these standards, a new phonemic awareness assessment measure, the Mountain Shadows Phonemic Awareness Scale (MS-PAS), was presented. This 20-item sound categorization task takes 20 minutes for an entire first-grade classroom. Reliability and validity data were good to excellent. It was recommended that the MS-PAS be used to identify students in need of explicit phonological awareness instruction or those who require more extensive individualized assessment.

The early stages of reading acquisition have been intensely investigated in recent years (Adams, 1990; Ball, 1993; Flynn & Rahbar, 1998; Rieben & Perfetti, 1991; Scanlon & Vellutino, 1996; Share & Stanovich, 1995). This research has yielded one consistent and clear conclusion: Phonemic awareness is strongly associated with the development of skilled reading. That is, children who understand that spoken words are composed of a series of discrete sounds that can be manipulated are more likely to become skilled readers than are children who are unable to hear and manipulate the individual sounds within words (Bus & van Ijzendoorn, 1999; Ehri, Nunes, Willows, Schuster, Yaghoub-Zadeh, & Shanahan, 2001; Snow, Burns, & Griffin, 1998).

Phonemic awareness gradually emerges during the preschool years. By kindergarten, evidence for its causal relationship with later reading success is overwhelming (Adams, 1990; Raz & Bryant, 1990; Scarborough, 2001; Share & Stanovich, 1995; Stanovich, 1986). Given this relationship, inclusion of phonemic awareness activities in kindergarten and first grade is frequently suggested (California State Board of Education, 1999; Mann, 1993; Scanlon & Vellutino, 1997). However, not all children develop strong phonemic awareness skills. Blevins (1997) reported that approximately 20% of children are affected by weak phonemic awareness skills that put them at increased risk for reading problems. Consequently, it has been recommended that teachers assess beginning readers to ensure proper development of phonemic awareness skills

Correspondence concerning this article should be addressed to Marley Watkins, Pennsylvania State University, Department of Educational and School Psychology and Special Education, 227 CEDAR Building, University Park, PA 16802. E-mail to mww10@psu.edu.

(Honig, Diamond, & Gutlohn, 2000; Zygouris-Coe, 2001). For example, Sodoro, Allinder, and Rankin-Erickson (2002) opined that "accurate assessment of phonological awareness is critical for teachers, educational specialists, and researchers" (p. 224) and the Consortium on Reading Excellence (1999) advised that a phonemic awareness screening test be administered "to all kindergarten children mid-year and to all first graders in the fall" (p. 15). More generally, it has been found that instruction guided by regular assessment facilitates student learning (Stecker & Fuchs, 2000).

Assessment of phonemic awareness has, however, been marked by wide variability. There has been little agreement on operationalization of the concept of phonemic awareness, and it has been measured by many different tasks that tap diverse aspects and levels of linguistic complexity (Adams, 1990; Sodoro et al., 2002; Stahl & Murray, 1994). Nevertheless, phonemic awareness tasks are generally found to be highly intercorrelated (Adams, 1990; Chafouleas, Lewandowski, Smith, & Blachman, 1997; Mann, 1993) and are thought to be described by only one or two factors (Stahl & Murray, 1994; Stanovich, Cunningham, & Cramer, 1984; Torgesen, Wagner, Bryant, & Pearson, 1992; Yopp, 1988).

Because phonemic awareness tasks are strongly related, it is important that a reliable and valid assessment method be selected from among the many possible operationalizations (Salvia & Ysseldyke, 2001). Additionally, it is vital that phonemic awareness assessments designed for classroom use be easy to administer and score (Alper, Ryndak, & Schloss, 2001). Finally, any classroom assessment must be cost-effective (Choate, Enright, Miller, Poteet, & Rakes, 1992). Given these criteria, a promising phonemic awareness assessment method is the sound categorization task developed by Bradley and Bryant (1985). In this task, children were presented with four pictures of objects, three of which rhymed and one that was the odd one out. After naming each of the pictures, children had to select the picture that did not belong with the others (did not rhyme). Ball (1993) pointed out that sound categorization tasks can be altered so that children categorize pictures by initial sounds or judge whether words share the same initial phoneme. Mann (1993) demonstrated that a group testing format is feasible and that the memory load can be reduced by accompanying the spoken words with pictures. Adams (1990) concluded that this type of sound categorization, which she labeled an oddity task, is the simplest phonemic awareness measure that retains substantial predictive validity for later reading achievement.

The Test of Phonological Awareness (TOPA; Torgesen & Bryant, 1994) is the only commercially available, group-administered sound categorization task. The TOPA-Kindergarten form was normed on 857 children in 10 states, and the TOPA-Early Elementary form used a normative sample of 3,654 children in 38 states. Coefficient alpha coefficients ranged from .87 to .91. Test-retest reliability for 69 first-grade children over an 8-week interval was .69. The TOPA-Kindergarten form given to 90 kindergarten children was moderately correlated ($r = .62$) with reading skills at the end of first grade.

Although the TOPA meets the criteria of reliability, validity, and ease of use, it is a commercial instrument that requires a financial commitment many schools cannot justify, especially for repeated measurements. Thus, a new

phonemic awareness assessment measure is presented that meets all four criteria: reliability, validity, ease of use, and cost-effectiveness.

METHOD

Participants

Nine consecutive cohorts of first-grade students ($n = 1,204$) tested in intact classrooms in a suburban, southwestern school served as participants. The number of students in each cohort is specified in Table 1. Most students were of White ethnic origin (91%), although students of Hispanic (4.3%) and African American (1.8%) ethnicity were also represented. Socioeconomic status was estimated to be lower middle to middle class because fewer than 10% of the students received free and reduced lunches. Boys and girls were relatively equally represented (49% versus 51%, respectively).

Instruments

The Mountain Shadows Phonemic Awareness Scale (MS-PAS; Watkins & Edwards, 1998) is a sound categorization task designed to assess young readers' phonemic awareness. It is a 20-item test designed to be administered to an entire first-grade class, though it may also be administered to kindergarten students in small groups of 6 to 8 students. Administration typically takes from 15 to 20 minutes. Following Ball (1993), both same and different sound categorizations are used. The 10 *same* items consist of a target picture and three response option pictures. The name for each picture is read aloud, and children are then asked to mark the picture that begins with the same sound as the stimulus word. This is followed by 10 *different* items where four pictures are presented, their names are read aloud, and children are instructed to mark the picture of the word that begins with a different sound than the other words. Both same and different components are preceded with practice items to ensure that children understand the task. The picture format reduces memory load and allows a purer assessment of phonemic awareness (Mann, 1993). A complete copy of the MS-PAS, including instructions and scoring forms, is available without cost on the web at http://espse.ed.psu.edu/spsy/Watkins/Watkins3.ssi.

The internal consistency reliability of the MS-PAS has been assessed in three previous studies. First, 63 students from central Pennsylvania were tested in intact classes with the MS-PAS at the beginning of first grade by Wyglinski (2000), who reported a coefficient alpha of .89. Second, 161 central Pennsylvania students were administered the MS-PAS in small groups at the end of kindergarten by Runge (2003), who found a coefficient alpha of .91. Both Pennsylvania samples were exclusively of White ethnic origin. Finally, a nationally representative sample of 4,112 students (49% girls) from the first three grades in the Republic of Trinidad and Tobago was tested in intact classes at the beginning of the school year with the MS-PAS by Worrell, Watkins, Runge, and Hall (2002), who reported a coefficient alpha of .89. The primary language of these students of African and East Indian ancestry was English.

Wyglinski (2000) also assessed the test-retest reliability and concurrent validity of the MS-PAS among a sample of 63 central Pennsylvania first-grade

Table 1

Mean and Coefficient Alpha of Scores on the Mountain Shadows Phonemic Awareness Scale (MS-PAS) and its Correlation with Reading Tests for Each First-Grade Cohort across Elementary School

MS-PAS			*r* with Achievement (*n*)						*r* with Teacher Rating (*n*)		
Cohort (*n*)	Mean (*SD*)	Alpha	Grade 1[a]	Grade 2[b]	Grade 3[b]	Grade 4[b]	Grade 5[b]	Grade 6[b]	Grade 1	Grade 2	Grade 3
1 (120)	16.6 (4.1)	.89	.60 (117)	-	.33 (54)	.33 (54)	.28 (63)	.48 (63)	.58 (117)	.47 (88)	.42 (78)
2 (117)	16.0 (4.6)	.91	.67 (108)	-	.50 (68)	.57 (62)	.44 (59)	.55 (55)	.67 (108)	.57 (80)	.51 (72)
3 (153)	16.5 (3.6)	.85	-	-	.51 (103)	.34 (84)	.44 (75)	.31 (73)	.42 (140)	.44 (121)	.39 (102)
4 (131)	15.4 (4.8)	.91	-	-	.51 (80)	.51 (81)	.47 (74)	.40 (71)	.38 (122)	.35 (102)	.47 (88)
5 (155)	15.3 (4.7)	.89	-	.54 (124)	.49 (110)	.53 (100)	.57 (88)	-	.62 (94)	.52 (130)	-
6 (127)	15.1 (4.7)	.89	-	.42 (93)	.49 (83)	.44 (70)	-	-	.48 (119)	-	-
7 (126)	15.0 (4.4)	.87	.62 (123)	.62 (98)	.61 (81)	-	-	-	-	-	-
8 (146)	15.4 (4.8)	-	.55 (132)	.57 (109)	-	-	-	-	-	-	-
9 (129)	15.1 (4.6)	-	-	-	-	-	-	-	-	-	-
Total (1204)	15.6 (4.5)	.89 (929)	.59 (480)	.54 (424)	.51 (579)	.49 (451)	.47 (359)	.44 (262)	.51 (700)	.46 (521)	.43 (340)

[a] Gates-MacGinitie Reading Test total score.

[b] Stanford Achievement Test total reading score.

students. When the MS-PAS was repeated at a 2-week interval, the test-retest reliability was .88. When the MS-PAS and TOPA were administered in counterbalanced order, the two tests were strongly related (r = .91, corrected for restriction of range). Thus, the MS-PAS and TOPA appeared to be similar measures among this sample of students (Streiner & Norman, 1995).

Gates-MacGinitie Reading Test (MacGinitie & MacGinitie, 1989) total reading scores were taken at the end of first grade. In grades 2 through 6, reading achievement was measured with the total reading score from the Stanford Achievement Test (Harcourt Brace Educational Measurement, 1996) series. To provide an alternative method of measuring student reading achievement, teachers marked a visual grade level scale for each student. This single item reflected the teacher's judgment of each student's mastery of the reading curriculum. Visually, this item was displayed horizontally, with grade levels ranging from 2 years below to 2 years above the student's actual grade placement in quarter-year increments. Teachers were unaware of students' MS-PAS and group achievement test scores when completing these ratings.

Procedure

All first-grade classes were administered the MS-PAS during the first 6 weeks of school by the school's remedial reading teachers. Administration was standardized by following the instructions and procedures accompanying the MS-PAS. Data on educational achievement were later extracted from school records as each first-grade cohort progressed across the elementary school grades. Although there was some variability across time due to school and district policy changes, all but the ninth cohort received at least one standardized achievement measure in the subsequent 5 years. Additionally, each student's mastery of the school's reading curriculum was ranked by teachers in late spring of first, second, and third grades for the initial cohorts. The schedule of assessments for each cohort is reported in Table 1.

RESULTS

As reported in Table 1, summary performance levels on the MS-PAS were available for 1,204 first-grade students: mean correct was 15.6 with a standard deviation of 4.6 (M boys = 15.5, girls = 15.9). There was no statistically significant difference between the performance of boys and girls, t (927) = -.37, p = .21.

Teacher ratings and standardized tests of reading achievement were strongly related. When collected simultaneously, the average correlation between teacher ratings and reading achievement test scores was .70. Although the relationship between teacher ratings and reading test scores attenuated across time, the predictive relationship remained robust across the elementary school years (.60, .57, .55, and .50 for reading test scores 1, 2, 3, and 4 years after teacher ratings). Thus, both methods seemed to tap components of reading achievement.

The mean performance of students on standardized reading tests and teacher ratings of reading achievement are provided in Table 2. Based on teacher ratings, student reading achievement paralleled actual grade placements (i.e., mean ratings of 1.9, 2.8, and 3.8 at the end of first, second, and

third grades, respectively). There were significant differences between ratings of boys and girls at all three grade levels, with girls receiving higher ratings although effect sizes were relatively small (see Table 2). In contrast, standardized reading test scores were not significantly different for boys or girls at any grade level.

Table 2
Mean (Standard Deviation) for Teacher Ratings of Reading Mastery in Grade Equivalent Scores and Reading Achievement in Normal Curve Equivalent Scores by Gender

Measure	*n*	Boys	Girls	Total
Teacher Rating Grade 1	722	1.85 (.52)	1.94 (.53)	1.89 (.53)*
Teacher Rating Grade 2	558	2.76 (.60)	2.88 (.54)	2.82 (.57)*
Teacher Rating Grade 3	367	3.72 (.61)	3.85 (.60)	3.78 (.61)*
Reading Test Grade 1[a]	500	58.1 (19.3)	56.9 (16.9)	57.6 (18.2)
Reading Test Grade 2[b]	432	61.0 (16.2)	60.0 (14.8)	60.5 (15.5)
Reading Test Grade 3[b]	600	57.8 (17.6)	58.2 (15.5)	58.0 (16.5)
Reading Test Grade 4[b]	474	60.8 (17.6)	60.7 (16.2)	60.8 (16.9)
Reading Test Grade 5[b]	382	58.8 (16.2)	59.4 (16.1)	59.1 (16.1)
Reading Test Grade 6[b]	282	60.2 (16.1)	63.6 (15.8)	61.9 (16.0)

[a] Gates-MacGinitie Reading Test total score.
[b] Stanford Achievement Test total reading score.
* $p < .05$.

Reliability

Internal consistency reliability for students tested with the MS-PAS in first grade was quantified by coefficient alpha. As detailed in Table 1, alpha coefficients ranged from .85 to .91 across seven cohorts, with a coefficient of .89 for the total sample of 929 students. Alpha coefficients were equivalent for boys and girls (.90 and .88, respectively) and for White, Hispanic, and African American students (.89, .91, and .89, respectively). A portion of the fifth cohort ($n = 115$) was also tested with the MS-PAS at the end of their kindergarten year. The coefficient alpha among this group of kindergarten students was .91. Test-retest reliability was also assessed among the 115 students of the fifth cohort who took the MS-PAS at the end of kindergarten and again at the beginning of first grade. The resulting stability coefficient was .74 across this 12- to 16-week interval.

Validity

The developmental growth of phonemic awareness skill was detected by the MS-PAS, as illustrated by its means and standard deviations across grade levels. One group of 115 students tested at the end of kindergarten ($M = 14.1$, $SD = 5.3$) and again at the beginning of first grade ($M = 15.5$, $SD = 4.6$) exhibited higher mean scores and lower variability across time, t (114) = 3.96, $p < .001$. Similarly, 376 students from the first, third, and eighth cohorts tested at the beginning of first grade ($M = 16.2$, $SD = 4.3$) and again at the beginning of second grade ($M = 19.4$, $SD = 1.5$) displayed higher mean scores and lower variability across time, t (375) = 15.5, $p < .001$. In contrast to this clear pattern of increasing MS-PAS scores across time, there were no significant differences between boys ($n = 52$, $M = 13.4$, $SD = 5.2$) and girls ($n = 63$, $M = 14.8$, $SD = 5.4$)

in kindergarten, t (113) = 1.4, p = .18, or second grade (n = 190 and 186, M = 19.4 and 19.4, SD = 1.5 and 1.5, respectively), t (374) = .12, p = .91.

As detailed in Table 1, predictive validity of the MS-PAS was assessed against teacher judgments and standardized reading achievement tests. MS-PAS scores at the beginning of first grade were moderately related to teacher ratings at the end of the first three grades (r = .51, .46, and .43, respectively). They were also moderately related to Gates-MacGinitie total reading scores taken at the end of first grade (r = .59). These robust relationships persisted across time. MS-PAS scores from the beginning of first grade were correlated with Stanford Achievement Test total reading scores from the end of second (r = .54), third (r = .51), fourth (r = .49), fifth (r = 47), and sixth (r = .44) grades.

Discriminative validity was evaluated by retrospectively analyzing the school's prereferral intervention records. All second- and third-grade teachers' reports for the first three cohorts were examined. Based on this review, students were classified into three categories: (a) no problem reported in either grade, (b) a minor problem reported in at least one grade that required the teacher to make classroom modifications but did not involve external resources, or (c) a major problem reported in at least one grade that necessitated such external resources as a special teacher (remedial reading or special education), cross-age peer tutor, or school psychologist. There were a total of 225 students with complete records. Of these, 161 (42% boys) had no problem, 33 (73% boys) experienced a minor problem, and 31 (65% boys) suffered a major problem. The validity of this tripartite categorization was confirmed by third-grade reading achievement scores (M = 64.9, 53.2, and 42.5 for no, minor, and major problem groups, respectively). These reading scores were significantly different across categories, F(2, 222) = 33.7, $p < .001$, with Bonferroni post hoc comparisons showing that all three groups were significantly different from each other. Likewise, these three groups significantly differed on first-grade MS-PAS scores (M = 17.9, 16.4, and 12.8 for no, minor, and major problem groups, respectively), F(2, 222) = 35.9, $p < .001$. Bonferroni post hoc comparisons again confirmed that all three groups were different significantly from each other.

The utility of MS-PAS scores in identifying individual students at severe risk for later reading difficulty was also explored. Those 31 students who experienced major academic problems in grade 2 or 3 were compared to the other 194 students. A receiver operating characteristic (ROC) curve (Figure 1) allowed the effectiveness of every cut score on the MS-PAS to be analyzed. Essentially, a ROC is a graph of the percentage of true positive decisions against the percentage of false positive decisions across all possible cutoff values (McFall & Treat, 1999). Overall, the ROC indicated that if one student were selected at random from the major problem group and another at random from the group without major problems, the MS-PAS would be 85% accurate in detecting the student with a major academic problem. Figure 1 reveals that a cut score of 17 produced a false positive rate of 32% and a true positive rate of 94%, the most discriminating for this sample of students. Unfortunately, students' first-grade MS-PAS scores were known to teachers so these results are only illustrative and should be validated locally to ensure applicability.

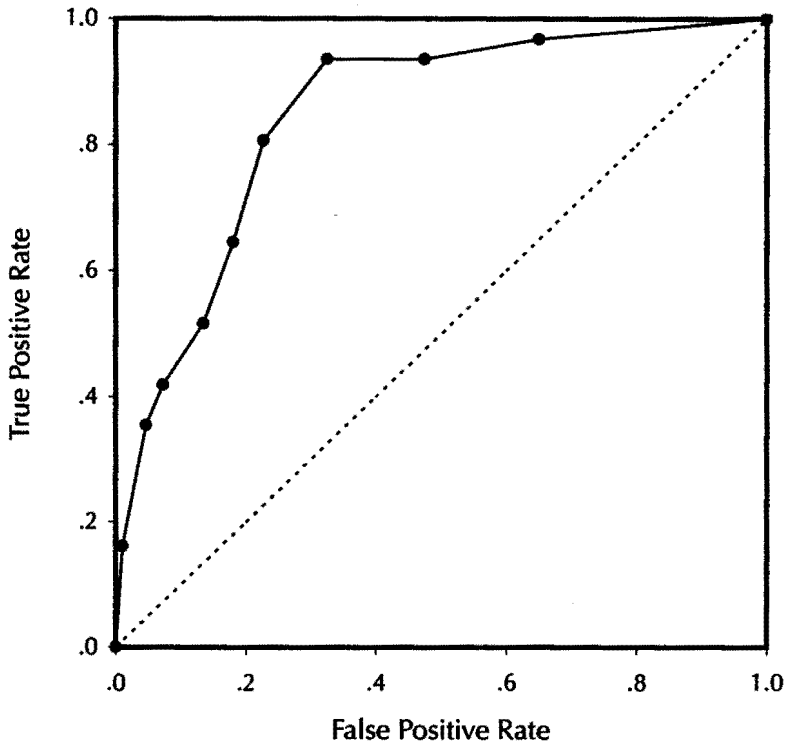

FIGURE 1. Receiver operating characteristic (ROC) curve for first-grade MS-PAS scores in identifying students who experienced severe academic problems in second or third grades.

Construct validity was explored via a principal axis factor analysis of MS-PAS scores of 929 students from seven consecutive cohorts who took the MS-PAS at the beginning of first grade. Based on scree (Cattell, 1966) and parallel analysis (Horn, 1965) criteria, two factors were extracted that accounted for 35% of the total variance. Following both oblique and orthogonal rotation, the same and different item types were cleanly separated into two factors. These results are illustrated in Table 3 by the pattern matrix from a Promax rotation. The two factors were correlated at .67, demonstrating overlapping but not identical coverage of phonemic awareness by same and different item types. However, conceptual formulations of phonemic awareness (Adams, 1990) and reduced reliability coefficients (-.07 to -.09) suggested that the MS-PAS total scale should be used rather than the MS-PAS item-type scales.

CONCLUSION

It has been estimated that 20% to 25% of students do not acquire proficient phonemic awareness skills without direct instruction (Adams, 1990; Blevins, 1997). Fortunately, phonemic awareness can be developed through instruction and doing so enhances children's subsequent reading skills (Bus & van Ijendoorn, 1999; Ehri et al., 2001; National Reading Panel, 2000). It is vital, however, to identify those students who do not spontaneously acquire phone-

mic awareness skills and provide them with explicit instruction because "'catching up' [is] all but impossible for students on a low developmental reading trajectory" (Good, Simmons, & Smith, 1998, p. 48).

Table 3
Pattern Coefficients from Principal Axis Factoring with Promax Rotation of 20 MS-PAS Items among 929 First-Grade Students

Item	Factor I	Factor II
Same 1	**.55**	-.02
Same 2	**.55**	-.04
Same 3	**.53**	.09
Same 4	**.61**	-.08
Same 5	**.54**	-.09
Same 6	**.53**	.06
Same 7	**.53**	.04
Same 8	**.46**	.12
Same 9	**.60**	.00
Same 10	**.49**	.13
Different 1	.05	**.66**
Different 2	.09	**.59**
Different 3	-.09	**.65**
Different 4	.04	**.57**
Different 5	.07	**.53**
Different 6	.04	**.57**
Different 7	.11	**.51**
Different 8	-.07	**.57**
Different 9	-.02	**.69**
Different 10	-.07	**.73**

Note.—Salient coefficients (≥ .40) in bold.

Given this fundamental relationship between phonemic awareness and reading, it has been recommended that a phonemic awareness screening test be administered to all kindergarten and first-grade students (Consortium on Reading Excellence, 1999). Measurement specialists suggest that tests should demonstrate reliability coefficients of .70 (Kline, 1998) to .90 (Salvia & Ysseldyke, 2001) if they are to be used to make important decisions about individuals. For screening, reliability coefficients of .80 have been recommended (Salvia & Ysseldyke, 2001). Streiner and Norman (1995) suggested that internal consistency reliability coefficients should exceed .80 and stability coefficients should exceed .50. The MS-PAS meets these standards. Using the guidelines promulgated by Cicchetti (1994), internal consistency reliability of the MS-PAS is categorized as good to excellent. Given that phonemic awareness is developmental and changes in scores across time represent both development and test error, the MS-PAS demonstrated substantial short-term stability when compared to similar instruments (Chafouleas et al., 1997; Torgesen & Bryant, 1994).

The predictive validity of the MS-PAS was very similar to other phonemic awareness tests when the criterion was end-of-first-grade reading achievement test scores (near .60; Stanovich et al., 1984; Torgesen & Bryant, 1994; Yopp, 1995). It demonstrated equivalent predictive accuracy when the criterion was teacher ratings of reading proficiency. Three-year prediction of reading test

scores was somewhat lower than reported by Yopp (1995) but was still significant. Scarborough (1998) summarized 27 phonemic awareness studies and reported a mean correlation of .46 between phonemic awareness and future reading. The relationship between MS-PAS and later reading scores ranged from .44 to .59. Thus, the MS-PAS exhibited substantial predictive validity.

Bowey (1995) suggested that children who scored below 90% on a phonological oddity task were more likely to become poor readers. In agreement with Bowey (1995), a cut score of 17 produced optimal diagnostic accuracy in this sample. A score of 17 or less on the MS-PAS identified 94% of the first-grade students who later experienced major academic problems in second or third grade. As befits a screening test, this cut score missed only 2 of the 31 students who later developed major academic problems. This true positive rate has been characterized as excellent by Cicchetti (2001).

The MS-PAS has demonstrated adequate reliability and validity among several samples of students. It produces equivalent results for boys and girls, and its diagnostic accuracy appears excellent. Its group administration format and simple scoring rules make it easy to use and time-efficient. Its authors' release of the test for noncommercial use makes it cost-efficient. Thus, the MS-PAS is a promising new tool for classroom teachers and other educational professionals charged with assessing the early reading skills of young children. Following the recommendation of Bowey (1995) and the current results, a 90% cut score (i.e., raw score = 17) on the MS-PAS is recommended for identification of beginning-first-grade students who can profit from explicit instruction or who require individualized assessment.

Like all investigations, however, this study was imperfect. The foremost limitation was the restricted variability in ethnic background and socioeconomic status (SES) of the participants. Although the MS-PAS exhibited good reliability with students of East Indian and African ancestry in the Republic of Trinidad and Tobago, there is little psychometric evidence regarding its use among American-born students of non-White ethnic background. Likewise, evidence of reliability and validity among lower SES students is limited. More research among these populations should be conducted. Additionally, the establishment of specific cut scores should be validated locally to ensure applicability.

REFERENCES

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

Alper, S., Ryndak, D. L., & Schloss, C. N. (2001). *Alternative assessment of students with disabilities in inclusive settings.* Boston: Allyn and Bacon.

Ball, E. W. (1993). Assessing phoneme awareness. *Language, Speech, and Hearing Services in Schools, 24,* 130-139.

Blevins, W. (1997). *Phonemic awareness activities for early reading success.* New York: Scholastic.

Bowey, J. A. (1995). On the contribution of phonological sensitivity to phonological recoding. *Issues in Education, 1,* 65-69.

Bradley, L., & Bryant, P. (1985). *Rhyme and reason in reading and spelling.* Ann Arbor: University of Michigan Press.

Bus, A. G., & Van Ijzendoorn, M. H. (1999). Phonological awareness and early reading: A meta-analysis of experimental training studies. *Journal of Educational Psychology, 91,* 403-414.

California State Board of Education. (1999). *Reading/language arts framework for California public schools kindergarten through grade twelve.* Sacramento, CA: Author.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1,* 245-276.

Chafouleas, S. M., Lewandowski, L. J., Smith, C. R., & Blachman, B. A. (1997). Phonological awareness skills in children: Examining performance across tasks and ages. *Journal of Psychoeducational Assessment, 15,* 334-347.

Choate, J., Enright, B., Miller, L., Poteet, J., & Rakes, T. (1992). *Curriculum-based assessment and programming.* Boston: Allyn and Bacon.

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

Cicchetti, D. V. (2001). The precision of reliability and validity estimates revisited: Distinguishing between clinical and statistical significance of sample size requirements. *Journal of Clinical and Experimental Neuropsychology, 23,* 695-700.

Consortium on Reading Excellence. (1999). *Assessing reading: Multiple measures for kindergarten through eighth grade.* Novato, CA: Arena Press.

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.

Flynn, J., & Rahbar, M. H. (1998). Kindergarten screening for risk of reading failure. *Journal of Psychoeducational Assessment, 16,* 15-35.

Good, R. H., Simmons, D. C., & Smith, S. B. (1998). Effective academic interventions in the United States: Evaluating and enhancing the acquisition of early reading skills. *School Psychology Review, 27,* 45-56.

Harcourt Brace Educational Measurement (1996). *Stanford Achievement Test.* San Antonio, TX: The Psychological Corporation.

Honig, B., Diamond, L., & Gutlohn, L. (2000). *Teaching reading sourcebook for kindergarten through eighth grade.* Novato, CA: Arena Press.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30,* 179-185.

Kline, P. (1998). *The new psychometrics: Science, psychology, and measurement.* London: Routledge.

MacGinitie, W. H., & MacGinitie, R. K. (1989). *Gates-MacGinitie reading tests.* Chicago: Riverside.

Mann, V. A. (1993). Phoneme awareness and future reading ability. *Journal of Learning Disabilities, 26,* 259-269.

McFall, R. M., & Treat, T. A. (1999). Quantifying the information value of clinical assessments with signal detection theory. *Annual Review of Psychology, 50,* 215–241.

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.* Bethesda, MD: National Institute of Child Health and Human Development.

Raz, I. S., & Bryant, P. (1990). Social background, phonological awareness and children's reading. *British Journal of Developmental Psychology, 8,* 209-225.

Rieben, L., & Perfetti, C. A. (Eds.). (1991). *Learning to read: Basic research and its implications.* Hillsdale: Erlbaum.

Runge, T. J. (2003). *The dimensionality of phonemic awareness among children at the end of kindergarten.* Unpublished dissertation, The Pennsylvania State University, University Park.

Salvia, J., & Ysseldyke, J. E. (2001). *Assessment* (8th ed). New York: Houghton Mifflin.

Scanlon, D. M., & Vellutino, F. R. (1996). Prerequisite skills, early instruction, and success in first-grade reading: Selected results from a longitudinal study. *Mental Retardation and Developmental Disabilities Research Reviews, 2,* 54-63.

Scanlon, D. M., & Vellutino, F. R. (1997). A comparison of the instructional backgrounds and cognitive profiles of poor, average, and good readers who were initially identified as at risk for reading failure. *Scientific Studies of Reading, 1,* 191-215.

Scarborough, H. S. (1998). Early identification of children at risk for reading disabilities: Phonological awareness and some other promising indicators. In B. K. Shapiro, P. J. Accardo, & A. J. Capute (Eds.), *Specific reading disability: A view of the spectrum* (pp. 75-119). Timonium, MD: York Press.

Scarborough, H. S. (2001). Connecting early language and literacy to later reading (dis)abilities: Evidence, theory, and practice. In S. B. Neuman & D. K. Dickinson (Eds.), *Handbook of early literacy research* (pp. 97-110). New York: Guilford.

Share, D. L., & Stanovich, K. E. (1995). Cognitive processes in early reading development: Accommodating individual differences into a model of acquisition. *Issues in Education, 1,* 1-57.

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

Stahl, S. A., & Murray, B. A. (1994). Defining phonological awareness and its relationship to early reading. *Journal of Educational Psychology, 86,* 221-234.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21,* 360-406.

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.

Stecker, P. M., & Fuchs, L. S. (2000). Effecting superior achievement using curriculum-based measurement: The importance of individual progress monitoring. *Learning Disabilities: Research & Practice, 15,* 128-134.

Streiner, D. L., & Norman, G. R. (1995). *Health measurement scales: A practical guide to their development and use* (2nd ed.). New York: Oxford University Press.

Torgesen, J. K., & Bryant, B. R. (1994). *Test of Phonological Awareness: Examiner's manual.* Austin, TX: ProEd.

Torgesen, J. K., Wagner, R. K., Bryant, B. R., & Pearson, N. (1992). Toward development of a kindergarten group test for phonological awareness. *Journal of Research and Development in Education, 25,* 113-119.

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

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.

Wyglinski, L. A. (2000). *A study of the reliability and validity of the Test of Phonological Awareness and the Mountain Shadows Phonemic Awareness Scale.* Unpublished master's thesis, The Pennsylvania State University, University Park.

Yopp, H. K. (1988). The validity and reliability of phonemic awareness tests. *Reading Research Quarterly, 23,* 159-177.

Yopp, H. K. (1995). A test for assessing phonemic awareness in young children. *The Reading Teacher, 49,* 20-29.

Zygouris-Coe, V. (2001). *Phonemic awareness* (Document No. 2-001). Orlando, FL: Family Literacy and Reading Excellence Center, University of Central Florida.