



# Research Brief

August 2008

Applied Research Unit

## Evaluation of the Criterion-Related Validity of Montgomery County Public Schools Assessment Program in Primary Reading

Huafang Zhao, Ph.D. and Clare Von Secker, Ph.D.

### Executive Summary

This research brief provides evidence that the reading assessments developed by Montgomery County Public Schools (MCPS) and administered as part of the MCPS Assessment Program in Primary Reading (AP-PR) are valid measures of reading proficiency (have criterion-related validity). The standards set for reading proficiency with the AP-PR are aligned with those used to establish *TerraNova* Second Edition (TN/2) national reading benchmark (50<sup>th</sup> percentile). Students who meet the AP-PR benchmark by the end of Grade 2 are likely to be reading on grade level in Grade 3.

Three recommendations are:

1. Continue to administer the AP-PR to evaluate Grade 2 reading proficiency.
2. Use formative assessment data obtained from the fall and winter administrations of the AP-PR to monitor student likelihood of scoring at or above the national benchmark on TN/2 subtests in reading.
3. Use the results of the AP-PR to identify and provide appropriate instructional interventions for students who are at risk of reading below grade level by the end of Grade 3.

### Background

One milestone of the MCPS strategic plan, *Our Call to Action: Pursuit of Excellence* is that all students meet or exceed standards in reading by Grade 3 (MCPS, 2007). MCPS developed the AP-PR as part of their effort to monitor student attainment of this milestone.

The AP-PR is administered in the fall, winter, and spring of each academic year to students in kindergarten through Grade 2. The AP-PR provides formative assessment data that are used to monitor students' reading accuracy, oral comprehension, and written comprehension. The AP-PR end-of-year

benchmarks are summative assessments that indicate whether students have sufficient reading skills by the end of a school year to be successful at the next grade level.

The purpose of this research is to evaluate the criterion-related validity of the AP-PR as a measure of reading proficiency. Reading assessments have criterion-related validity if they have concurrent and predictive validity with other reading assessments (Messick, 1993). Evidence of the criterion-related validity of the AP-PR was found by determining how well AP-PR results correspond with those obtained by other reading assessments that are administered at about the same time (concurrent validity) or during a subsequent school year (predictive validity).

Two research questions were:

1. Does Grade 2 student attainment of the AP-PR end-of-year benchmark have high concurrent validity with attainment of the national reading benchmark for the TN/2 administered in spring of Grade 2?
2. Does Grade 2 student attainment of the AP-PR end-of-year benchmark have high predictive validity with proficiency levels on the Grade 3 Maryland School Assessment (MSA) Reading?

### Methodology

The sample used to evaluate concurrent validity (2007 cohort) included 8,037 Grade 2 students who took the AP-PR and the TN/2 in spring 2007. Students who received special education and English language learner (ELL) services are slightly underrepresented in the 2007 cohort because those students were more likely than other students to receive accommodations on the TN/2. Test takers who received a verbatim reading accommodation on the TN/2 reading subtest (1,072 students) were excluded because this accommodation may invalidate test results (CTB/McGraw-Hill, 2002).

The sample used to evaluate predictive validity (2006 cohort) included 8,206 students who took the Grade 2 AP–PR in spring 2006 and the Grade 3 MSA Reading in spring 2007. The relative proportions of students in the samples are similar to the proportions in the population of all Grade 2 students. The demographic compositions of the 2007 and 2006 cohorts are shown in Tables A1 and A2 (Appendix).

### *MCPS AP–PR*

Grade 2 AP–PR includes a Text Reading and Comprehension (TRC) book-level component and measures of reading accuracy, oral comprehension, and written comprehension (MCPS, 2005). Reading accuracy scores indicate whether students can read independently. Oral comprehension scores indicate how well students understand reading passages. Two written comprehension scores evaluate how well students are able to construct ideas and communicate their reading comprehension through writing.

To meet the Grade 2 AP–PR reading benchmark, students must meet four criteria. Those criteria are (1) read a book at TRC level M or higher; (2) read with at least 90% accuracy; (3) answer correctly at least four out of five literal or interpretive oral comprehension questions; and (4) score at least 2 out of a possible 3 points for each of two written response questions (Douglas, 2006).

### *Concurrent Validity*

This analysis compared AP–PR subtest scores with scale scores obtained on the TN/2 reading, language, and language mechanics subtests administered to students in spring of Grade 2. The AP–PR and TN/2 were administered during approximately the same testing window. More than 90% of the items on the TN/2 subtests are aligned with the MCPS reading curriculum (MCPS, 2007). Table A3 (Appendix) shows descriptive statistics for the AP–PR and TN/2.

The TN/2 data compare performance of MCPS Grade 2 students with a national sample of Grade 2 students. The national average (national benchmark) for all students on each TN/2 subtest is the 50<sup>th</sup> normal curve equivalent (NCE). Grade 2 students who meet the AP–PR end-of-year benchmark were expected to score at or above the 50<sup>th</sup> NCE on the TN/2 reading subtest.

Concurrent validity was established by calculating the Pearson correlation coefficients for the AP–PR subtests with the TN/2 subtests in reading, language, and language mechanics. In addition, results were

examined to determine the classification accuracy for student attainment on the MCPS AP–PR and national reading benchmarks. Classifications were accurate if students attained both (i.e., met benchmark and scored at or above the 50<sup>th</sup> NCE) or did not attain both. Classifications were inaccurate if students met one benchmark but not the other.

### *Predictive Validity*

Predictive validity of the AP–PR was evaluated by comparing classification accuracy for the Grade 2 end-of-year benchmark with results of the MSA Reading administered in spring of Grade 3. Students who met the AP–PR Grade 2 end-of-year benchmark were predicted to attain scores of proficient or advanced on the Grade 3 MSA Reading. Students who did not meet the Grade 2 end-of-year benchmark were predicted to attain scores of basic on the Grade 3 MSA Reading.

## **Results**

### *Evidence of Concurrent Validity*

Overall, the correlations between the AP–PR and TN/2 provide evidence of concurrent validity for the two tests (Table 1). TN/2 reading scale scores were correlated with AP–PR TRC book level (0.57) and the accuracy subtest (0.40).

Table 1  
Correlation of 2007 Subtest Scores on the  
Grade 2 AP–PR and TN/2 Reading  
(2007 Cohort)

AP–PR Subtest	Correlation with TN/2 Reading Scale Score
TRC book level	0.57
Accuracy	0.40
Oral comprehension	0.15
Written Comprehension 1	0.17
Written Comprehension 2	0.16

The patterns of correlations observed for AP–PR and TN/2 reading tests were similar to those observed for two other TN/2 subtests that measure reading skills (Table 2). Both the TN/2 language and language mechanics subtest scale scores were highly correlated (greater than 0.50) with the AP–PR TRC book level moderately correlated with the accuracy subtest (greater than 0.30).

Table 2  
Correlation of 2007 Subtest Scores on the  
Grade 2 AP–PR and TN/2 Reading  
(2007 Cohort)

AP–PR Subtest	Correlation with TN/2 Scale Score	
	Language	Language Mechanics
TRC book level	0.58	0.51
Accuracy	0.40	0.32
Oral comprehension	0.15	0.12
Written Comprehension 1	0.20	0.18
Written Comprehension 2	0.18	0.15

The narrow score range for student performance on the AP–PR subtests for oral comprehension (0 to 5) and written comprehension (0 to 3) may account for their relatively low correlations with TN/2 scale scores in reading, language, and language mechanics. Everything else being equal, correlations among scores are lower when the range of scores is narrower (Glass & Hopkins, 1970). Most Grade 2 students, even those who did not meet the benchmark, achieved oral and written comprehension scores of 2 or higher.

Additional evidence of the concurrent validity of the AP–PR and the TN/2 was obtained by examining the consistency of student performance on the two tests. More than 75% of Grade 2 students who took the AP–PR and TN/2 reading subtest in 2007 were accurately classified as having met (59.9%) or not met (17.1%) the MCPS and national benchmarks (Table 3). Similar results were observed for classification accuracy between the AP–PR and the TN/2 subtests for language and language mechanics (not shown).

Table 3  
Classification Consistency of Grade 2 Student  
Reading Benchmark Status from Results of  
AP–PR and TN/2 Administered in 2007

MCPS AP–PR Benchmark Status	TN/2 Reading National Benchmark <sup>a</sup>	
	Below 50 <sup>th</sup> NCE	At or Above 50 <sup>th</sup> NCE
Not Met	17.1%	9.4%
Met	13.6%	59.9%

Note. Accuracy Rate = 77.0% (17.1% + 59.9%)

<sup>a</sup> 50<sup>th</sup> NCE

Students who met one benchmark but not the other were slightly more likely to meet the AP–PR

benchmark than the national benchmark. In 2007, 13.6% of the students who met the AP–PR benchmark scored at or below the 50<sup>th</sup> NCE in reading, versus 9.4% of students who scored at or above the 50<sup>th</sup> NCE in reading but did not meet the AP–PR benchmark. This suggests that the MCPS AP–PR benchmark is slightly lower than the 50<sup>th</sup> NCE on the TN/2 reading subtest.

#### *Evidence of Predictive Validity*

Student performance on the Grade 2 AP–PR benchmark correctly predicted Grade 3 MSA Reading proficiency for more than 75% of students who took both tests (Table 4). When predictions were incorrect, students were more likely to perform better than predicted on the Grade 3 MSA Reading. This suggests that the AP–PR benchmark is predictive of proficiency on the Grade 3 MSA Reading.

Overall, more than 20% of students who did not meet the Grade 2 AP–PR reading benchmark attained scores of proficient or advanced on the Grade 3 MSA Reading. Only 3.8% of students who met the AP–PR benchmark performed worse than expected and attained scores of basic on the MSA. The same patterns observed for all students combined were observed for students by racial/ethnic group.

Table 4  
Accuracy of Grade 2 AP–PR Predictions of  
Grade 3 MSA Reading Proficiency  
by Race/Ethnicity

Demographic Group	N Took AP–PR and MSA	Prediction Accuracy <sup>a</sup>		
		% Correctly Predicted	% Scored Better	% Scored Worse
All Students	8,602	75.3	20.9	3.8
African Am.	1,822	69.3	23.4	7.3
Asian Am.	1,271	80.0	17.6	2.4
Hispanic	1,487	64.3	29.7	6.1
White	3,601	81.2	17.2	1.5

Note. American Indian students are included with all students but are not reported separately.

<sup>a</sup> Row totals may not sum to 100% due to rounding.

The Grade 2 AP–PR end-of-year benchmark predictions of Grade 3 MSA Reading proficiency were more accurate for Asian American (80.0%) and White (81.2%) students than for African American (69.3%) and Hispanic (64.3%) students. One reason for this difference is that more African American and Hispanic students received Grade 3 MSA Reading

scores that were close to the MSA cut scores for reading proficiency. Measurement errors are expected to be higher for students whose scores were close to the cut score.

### Conclusions

The AP–PR is a high-quality formative assessment that promotes student achievement by providing timely feedback about students’ reading progress and readiness. The criterion-related validity of the assessments was evident from the high levels of consistency in student performance on the AP–PR with other reading measures. The consistencies demonstrate that the AP–PR meets two standards for criterion-related validity, namely, concurrent validity and predictive validity.

The reasonably high correlations between AP–PR subtest scores and TN/2 scale scores for subtests in reading, language, and language mechanics provide evidence of concurrent validity. Both tests are aligned with the MCPS curriculum and are administered in spring of Grade 2. The standards used by MCPS AP–PR to establish reading proficiency are aligned with those used to establish the national benchmark (50<sup>th</sup> NCE) in reading. Students who met the AP–PR Grade 2 end-of-year benchmark in reading also were likely to score at or above the 50<sup>th</sup> NCE on TN/2 subtests in reading, language, and language mechanics.

The high correlation in the results obtained by the MCPS AP–PR administered in Grade 2 and the MSA Reading administered in Grade 3 provided evidence of predictive validity. Students who met the AP–PR Grade 2 end-of-year benchmark in reading also were likely to score at the level of proficient or higher on the Grade 3 MSA Reading.

Performance on the AP–PR end-of-year benchmark is predictive of whether students will be reading on grade level by Grade 3. However, the AP–PR benchmark may underestimate Grade 3 MSA reading proficiency for some students. Early identification of students at risk for poor performance and effective delivery of curriculum and instruction is likely to result in higher than predicted third grade achievement.

Teachers and other educators are reminded that although end-of-year benchmark is a summative measure, other factors, such as effective implementation of a curriculum with clear standards taught by highly qualified teachers who have received focused professional development, will

influence academic outcomes. Thus, although the AP–PR has high predictive validity, it is not a conclusive forecast of future student achievement.

### Recommendations

1. MCPS should continue to administer the AP–PR to monitor Grade 2 reading proficiency.
2. Formative assessment data should be used to guide decisions about how to modify instruction and implement interventions as so to best meet the learning needs of individual students.
3. Teachers should use summative AP–PR data to identify students who are at risk of reading below grade level in Grade 3.

### References

- CTB/McGraw-Hill. (2002). *Classroom connections to TerraNova, the second edition*. Monterey, CA: Author.
- Douglas, A.R. (2006). *Results for the 2006 spring reading benchmark assessments*. Rockville, MD: Montgomery County Public Schools.
- Glass, G. & Hopkins, K. (1970) *Statistical methods in education and psychology* (2<sup>nd</sup> edition). Needham Heights, MA, Allyn and Bacon.
- Messick, S. (1993). *Educational measurement* (3rd edition). Phoenix: Oryx Press.
- MCPS. (2005). *AP–PR: Reading 3D supplementary guide*. Rockville, MD: Author.
- MCPS. (2007). *Our call to action: pursuit of excellence*. Rockville, MD: Author.

### Acknowledgements

We greatly appreciate the collegial support and valuable comments provided for this study by Ms. Ann Bedford, Ms. Sophie Kowzun, Ms. Margaret Gheen, and Mr. Steven Fink.

Table A1  
 Number and Percentage of MCPS Grade 2 Students Who Took the  
 MCPS AP–PR and TN/2 in Spring 2007 by Demographic Group  
 (2007 Cohort)

Demographic Group	<i>N</i> Students Took Both Tests	% Test Takers
All	8,037	100.0
Female	3,986	49.6
Male	4,051	50.4
African American	1,852	23.0
Asian American	1,260	15.7
Hispanic	1,331	16.6
White	3,561	44.3
FARMS	1,921	23.9
Special Education <sup>a</sup>	477	5.9
ELL <sup>a</sup>	594	7.4

*Note:* American Indian students are included with all students but are not reported separately.

<sup>a</sup> Students who received these services were more likely than other groups to receive verbatim read-to accommodations. Students who received verbatim read-to accommodations on the TN/2 reading test were excluded from the 2007 cohort because the accommodations may invalidate test results.

FARMS=Free and Reduced-price Meals System

ELL=English Language Learners

Table A2  
 Number and Percentage of MCPS Students Who Took the  
 MCPS Grade 2 AP-PR in Spring 2006 and the Grade 3 MSA Reading in Spring 2007  
 by Demographic Group  
 (2006 Cohort)

Demographic Group	<i>N</i> Students Took Both Tests	% Test Takers
All	8,206	
Female	3,987	48.6
Male	4,219	51.4
African American	1,822	22.2
Asian American	1,271	15.5
Hispanic	1,487	18.1
White	3,601	43.9
FARMS	2,199	26.8
Special Education	785	9.6
ELL	610	7.4

*Note:* American Indian students are included with all students but are not reported separately.

FARMS=Free and Reduced-price Meals System

ELL=English Language Learners

Table A3  
 Descriptive Statistics for the AP-PR and TN/2 Tests  
 Administered to MCPS Grade 2 Students in Spring 2007  
 (2007 Cohort)

Test and Subtest Score	N Scores	Minimum Score	Maximum Score	Mean	SD
<b>AP-PR</b>					
TRC book level	7,492	3	38 <sup>a</sup>	28.7 <sup>a</sup>	6.6
Accuracy	7,492	55	100	97.8	2.4
Oral comprehension	7,492	0	5	4.6	0.7
Written Comprehension 1	6,861	0	3	2.1	0.7
Written Comprehension 2	6,861	0	3	2.0	0.7
<b>TN/2</b>					
Reading scale score	8,037	423	722	628.0	36.9
Language scale score	8,037	424	706	632.9	39.9
Language Mechanics scale score	7,988	445	695	624.1	35.6

<sup>a</sup> MCPS reports TRC book levels as both numbers and letters. For descriptive purposes, alphanumeric values for TRC book levels were converted to numeric values. TRC book level M corresponds to numeric value 27. This highest TRC book level, P, corresponds to numeric value 38.