Evaluation of the Montgomery County Public Schools Assessment Program: Grades 1 and 2 Reading

Office of Shared Accountability

October 2003

Elizabeth Cooper-Martin, Ph.D.
Theresa Alban, Ph. D.

OFFICE OF SHARED ACCOUNTABILITY

Dr. Wesley L. Boykin, Director
850 Hungerford Drive
Rockville, Maryland 20850
301-279-3448

Dr. Jerry D. Weast
Superintendent of Schools

Dr. Frieda K. Lacey
Chief of Staff

Table of Contents

Tables

Figures

Appendices

Executive Summary

This report provides the leadership of the school system with an evaluation of the MCPS Assessment Program-Primary Reading (MCPSAP-PR) for Grade 1 and Grade 2 students. The purpose of this evaluation is twofold—

1. To analyze student academic outcomes.
2. To examine the impact of the MCPSAP-PR on reading instruction in the primary grades and the refinements that have occurred, or need to occur, to the primary reading assessments.

In 2002–2003, 61% of Grade 1 students were able to meet benchmark performance. Students who did not meet the benchmark varied in reading levels from those reading at a kindergarten level to those who could read accurately at benchmark level but lacked adequate comprehension. All subgroups of students demonstrated comparable growth from fall to spring. However, there were statistically significant differences in the percentage of students able to reach benchmark based on race/ethnicity and on services received (i.e., English for Speakers of Other Languages (ESOL), Free and Reduced-price Meals (FARMS), and special education). The impact of poverty and second-language learning was evident, as only 24% of students receiving both ESOL and FARMS services were able to meet benchmark.

In Grade 2, 63 of students were able to meet benchmark performance, while those who did not had reading levels ranging from two grades behind to a level at or above benchmark but without adequate comprehension. All subgroups of students' demonstrated comparable growth from fall to spring, except for slower growth by special education students. Furthermore, there were statistically significant differences in the percentage of students able to reach benchmark based on race/ethnicity and services received. Once again, the impact of poverty and second-language learning was evident, with only 25% of students receiving both FARMS and ESOL services able to meet benchmark.

A sample of oral retelling and oral comprehension questions revealed strong inter-rater reliability that supports consistent use of scoring tools for MCPSAP-PR.

Several tests showed the usefulness of MCPSAP-PR results for predicting reading performance later in the year or in subsequent years. In each case, the usefulness of the MCPSAP-PR measures did not vary across racial/ethnic groups or support services received.

- Quarterly benchmarks established for both Grade 1 and Grade 2 were appropriate predictors for the annual benchmark.
- Annual benchmarks are strong predictors of performance in subsequent grades. More than 75% of Grade 1 students who met the kindergarten benchmark went on to meet the Grade 1 benchmark, and 85% of Grade 2 students who met the kindergarten benchmark went on to meet the Grade 2 benchmark. Additionally, 84% of Grade 2 students who met the Grade 1 benchmark went on to meet the Grade 2 benchmark.
- MCPSAP-PR measures have a statistically significant relationship with Comprehensive Test of Basic Skills (CTBS), which suggests that performance on MCPSAP-PR is predictive of performance on the reading and language subtests of the CTBS.

- MCPSAP-PR measures have a statistically significant relationship with MSA, suggesting that performance on MCPSAP-PR is predictive of performance on the MSA. Specifically, 90% of Grade 3 students who met the Grade 1 benchmark were able to score at the proficient level on the Grade 3 MSA, and 93% of Grade 3 students who met the Grade 2 benchmark were able to score at the proficient level on the Grade 3 MSA.

Teachers use MCPSAP-PR results most often to group students and to gain insights into specific student skills. The majority of teachers interviewed report that the assessments were worth the time it took to administer them. In keeping with the purposes of the MCPSAP-PR, teachers differentiate instruction during small-group lessons and perform checks for student understanding, particularly during whole-group lessons. Training to support implementation and analysis of the MCPSAP-PR occurs most often within individual schools in meetings with the reading specialist, staff development teacher, and peers.

Program Description

The MCPS Assessment Program-Primary Reading (MCPSAP–PR) is a locally developed assessment that provides formative information to help schools monitor students' progress in reading from prekindergarten through Grade 2. The stated goals of this assessment program are "to provide continuous confirmation of a student's reading development and some understanding of oral reading fluency, accuracy, and comprehension" (page 1, Office of Instruction and Program Development, 2002a). This program is a revision of Montgomery County's Early Childhood Assessment Program (ECAP), which was first introduced into 54 schools for Grade 2 during 1999-2000.

The MCPSAP-PR consists of two components—foundational reading skills and reading proficiency. The former are the building blocks that precede text reading; these skills are the focus of teaching and assessing in kindergarten. Students in Grades 1 and 2 who have not yet mastered these foundational skills are assessed in these six areas—

- Letter knowledge
- Print concepts
- Oral language
- Phonemic awareness
- Phonics
- Reading vocabulary

In Grades 1 and Grade 2, the focus of assessment and instruction shifts to text reading. The reading proficiency component of MCPSAP-PR assesses four areas—

- Accuracy of reading
- Reading behaviors
- Oral reading fluency
- Comprehension of the text

Accuracy of reading is assessed with a running record, which allows the teacher to record errors in word recognition as the student reads aloud. While the student is reading the text aloud, the teacher assesses reading behaviors—the application of foundational reading skills, such as

concepts about print, word recognition, and phonemic awareness second oral reading fluency—the student's use of expression and ability to divide the text into meaningful chunks.

Comprehension is measured in the manner that is most appropriate for the given reading level. For texts at the lowest levels of 3 to 9, the required comprehension task is oral retell, which allows the student to demonstrate story structure understanding of narration text. For texts at levels 10 to 16, the required comprehension tasks include oral comprehension questions and a written response. For texts at the highest levels of J to P, the required comprehension task is a written response.

The proficiency benchmarks for 2002–2003 for Grades 1 and 2 were set as follows. The Grade 1 target is reading a text at level 16 or higher with an accuracy rate of 90% or higher, along with a score of 80 to 100% on oral comprehension. The Grade 2 target is reading a text at level M or higher with an accuracy rate of 90% or higher, along with a score of 2 or 3 on the written response, representing partial or essential understanding.

Previous reports on ECAP presented by the MCPS Office of Shared Accountability (OSA) have documented the ongoing refinement of this assessment program since 1999. These reports also examine the reliability and validity of several components of ECAP (Raber, 2000 and OSA, 2002).

Scope of the Evaluation

The primary purpose of this evaluation is the analysis of student academic outcomes, which has been the primary focus of previous reports on ECAP. The evaluation then examines the impact of the MCPSAP-PR on reading instruction in the primary grades. Three essential questions guide the evaluation:

1. How has the MCPSAP-PR impacted instruction?
2. What refinement has occurred, or needs to occur, to the primary reading assessments?
3. What were the student academic outcomes in reading?

A variety of evaluation instruments and procedures were used to gather data for this study:

- Interviews with teachers to gather perceptions regarding implementation of the assessment program and its effects on classroom instruction and student learning;
- Observations of classroom instruction; and
- Analysis of systemwide assessments and achievement results from statewide assessments.

Major Evaluation Questions and Results

1. How has the MCPS Assessment Program Primary Reading impacted instruction?

Methodology

To support ongoing improvement and refinement of the MCPS-AP in both reading and mathematics, a representative sample of 19 elementary schools were identified as "in-depth study schools." These schools were selected to provide a range of demographic characteristics;

their participation is voluntary. Staff members who work with students in kindergarten through Grade 2 in these selected schools participated in group interviews twice during the year (winter and spring) to discuss implementation of the assessment program and accompanying curriculum. At the spring meeting, held in June 2003, 37 staff members responded to a written survey on the MCPSAP-PR. The respondents included 10 Grade 1 teachers and 8 Grade 2 teachers.

OSA staff conducted classroom observations of reading instruction in Grades 1 and 2 in 17 of the in-depth study schools from February to May 2003. The purpose of the observations was to determine how teachers were using the reading assessments to guide instruction. The observations included the following components:

- A set of pre-observation questions, completed by the teacher via e-mail, to provide background information on what would occur in the classroom on the day of the observation (e.g., Essential Questions, activities, expectations for student learning)
- Observation of a full block of reading instruction (approximately 90 minutes)
- A set of post-observation questions completed by the teacher one-on-one with the observer, over the phone, or via e-mail.

Appendix A contains each of these data/collection instruments. The reading observation protocol was designed to collect information on checks for student understanding, differentiation of instruction, and informal assessments. The observer recorded how frequently these practices occurred during each whole-group, small-group, or individual activity throughout the entire instructional block. The post-observation questions addressed teacher use of reading assessments and also professional development to support implementation of the assessment program.

OSA staff visited a total of 68 classrooms—34 in Grade 1 and Grade 2. They conducted post-observation interviews with all Grade 1 teachers and with 31 of the Grade 2 teachers.

At the spring meeting of the in-depth study schools, teachers and other staff rated their experience in using the data from the MCPSAP-PR with a 5-point scale that labeled 1 as Useless, 3 as Moderately Useful, and 5 as Highly Useful (see Appendix B).

Analysis

An analysis of the frequency of observed behaviors and interview responses was then completed. A one-way ANOVA with post-hoc contrasts was used to test for differences in frequencies of the practices teachers used to check for student understanding.

Results

Teacher Use of Reading Assessments: Overall. As part of the post-observation interviews, OSA staff questioned teachers about their use of the assessments during the observed lesson. The data in Table 1 show that the three most common uses were grouping students, insights into student skills, and comprehension.

Table 1. Teacher use of MCPSAP-PR

	Teachers (N=65)	
	N	%
Grouping	30	46
Insights into specific student skills	27	42
Comprehension	17	26
Writing	5	8
General, Weaknesses to reinforce	11	17
Other uses	11	17
Don't use	8	12

Note: Columns do not sum to 100% because teachers could give more than one response.

In a related question, teachers reported how they formed small groups for the observed lesson or a recent lesson (see Table 2). Compared with the previous question, reported additional teachers using the MCPSAP-PR for forming groups. The two most common approaches—MCPSAP-PR and reading levels—are similar, as the assessments provide a student's reading level.

Table 2. How Teachers Formed Small Groups for Reading

	Teachers (N=65)	
	N	%*
MCPSAP-PR	38	59
Reading level	35	54
Running records, other informal assessments	15	23
Need	9	14
Comprehension level	5	8
Other	7	11
Don't teach small groups	2	3

Note: Columns do not sum to 100% because teachers could give more than one response.

Teachers also gave their opinions on whether the MCPSAP-PR assessments are worth the time it takes to administer them. The majority of teachers (48 or 74% of the total) reported that the assessments were worthwhile; only four teachers (6%) felt they were not worth the time it takes. The remaining 13 teachers were not sure, did not administer the tests, or did not answer.

The most frequently reported benefit of the MCPSAP-PR (noted by 36 or 55% of the teachers) was that the assessments showed them where students are, what they can do, and what they need. As one teacher stated, with the assessments, "I'm not shooting in the dark." Fewer teachers mentioned the benefits of showing growth (9 teachers or 14%) and of providing uniformity across classrooms or across the county (6 teachers or 9%).

The most commonly reported problem of the MCPSAP-PR (given by 35% of the teachers) was the amount of time required, especially for tests of foundational skills. Some teachers noted the importance of release time to conduct the assessments and others believed some parts, such as running records, are not necessary. Fewer teachers mentioned that they wanted a different type

of assessment (5 teachers or 8%) and that there was not a wide enough selection of books (3 teachers or 5%).

Teachers the in-depth schools who participated in a survey on the MCPSAP-PR offered views that support the earlier finding that many teachers find the assessments useful for grouping students (see Table 3). The findings from the survey also suggest that the assessment data are at least moderately useful for a variety of conferences (e.g., with parents) and discussions (e.g., across grades).

Table 3. Teachers' Experience Using Data from MCPSAP-PR

Purpose (number of responses)	Scale Response					
	1 & 2 (1=Useless)		3 (3=Moderately Useful)		4 & 5 (5=Highly Useful)	
	n	%	n	%	n	%
Differentiating instruction/flexible grouping (n=16)	0	0	0	0	16	100
Grade level discussion (n=18)	0	0	2	11	16	89
EMT (n=15)	1	7	3	20	11	73
Parent conferencing (n=18)	1	6	6	33	11	61
Principal/teacher conferencing (n=17)	1	6	7	41	9	53
Cross-grade level discussion (n=15)	3	20	5	33	7	47

Note: Some teachers did not answer all items.

Teacher Use of Reading Assessments: Differentiation of Instruction

As suggested by the survey just reviewed, an important use of MCPSAP-PR data is for differentiation of instruction. Observations were used to study differentiation in the classroom. For each activity during the observed classes, OSA staff recorded the classroom organization (e.g., whole group), the purpose or focus of the activity, a description of the activity, and whether teachers differentiated instruction during that activity, either by giving alternative instructions for the same assignment or by giving alternative activities or instructions.

Of 68 classes observed, 67 included a whole-group lesson. Only 17 (25%) of the teachers did one or both of the differentiating activities in a whole-group lesson.

Of 68 classes observed, 18 had one or more activities during which students worked individually. As for whole-group lessons, only about one quarter of the teachers (four or 22%) did one or both of the differentiating activities during individual activities.

Differentiation during small-group lessons was more common. Of 68 classes observed, 62 had one or more small group lessons. Half of the teachers (31 of 62) included one or both of the differentiating activities within a small-group. The frequencies of differentiating activities did not vary between Grades 1 and 2, for whole group, individual, or small groups.

In 58 classes, more than one small group was observed. The purpose and activity descriptions of each small group were reviewed, along with any description of activities in the pre-observation questionnaire, to identify whether teachers gave different activities to different small groups, if so, and whether the activities varied according to student ability. Among the 58 teachers with two or more small groups, the majority (44 teachers, 76%), gave a different activity to each small group. However, 12 teachers (21%) did not differentiate and two teachers were not classified, due to insufficient notes on the observations. Nearly all teachers with small groups (56 or 90%) differentiated between small groups or differentiated within the group or both.

Teacher Use of Reading Assessments: Instructional Strategies

A key purpose of the MCPSAP-PR is to assist teachers in monitoring students' progress and to make instructional decisions. Additionally, the Instructional Guides for Reading/Language Arts for both Grade 1 and Grade 2 encourage teachers to do daily running records, as a type of ongoing assessment. To see how teachers apply ongoing assessment, two instructional strategies were examined in classroom observations—teachers' use of informal assessments in between the windows for the MCPSAP-PR and their use of checks for student understanding. Additionally, in the post-observation interview, teachers described any special strategies they used for students who were performing below the quarterly benchmark for reading.

During the 68 observed classes, 15 teachers (22%) took at least one running record and 21 teachers (31%) took anecdotal notes on at least one student. Combining these two assessment strategies, a total of 28 teachers (41%) performed at least one type of recommended ongoing assessment.

Observers recorded how frequently teachers used each of the eight checks for student understanding during each observed activity. Teachers used at least one check for student understanding during every observed activity, except for one small group. The total frequency of each check for student understanding was calculated separately for each of the three types of activity—whole group, small group, and individual. A mean was, then calculated for each type of activity by dividing the total for each check for student understanding by the number of occurrences of that activity (e.g., 98 for whole group). Across all activities, the two most frequently used checks were asking questions that require a multiple-word response and asking questions that require a single-word response (see Table 4). The overall use of checks for understanding was higher for whole group activities than for small groups; this difference was statistically significant ($F(2,283) = 5.78$, $p < .01$).

Table 4. Observations of Teacher Checks for Student Understanding

	Whole-Group (N=98)	Small-Group (N=163)	Individual (N=25)
Teacher check for student understanding	**Mean**	**Mean**	**Mean**
Asks question that requires multiple-word response	6.1	4.6	4.6
Asks question that requires single-word response	4.4	3.5	3.5
Asks student to clarify thinking or justify response	2.0	1.3	0.4
Repeats instruction when necessary for student understanding	1.2	0.9	1.7
Uses every student response technique	0.7	0.7	0.1
Uses instructional activity to check for understanding	0.7	0.4	0.2
Elicits questions from students	0.6	0.3	0.4
Uses exit cards	0.1	0.1	0
All checks	15.8	11.8	10.4

The third instructional strategy studied concerned strategies to help students who were performing below the quarterly benchmark for reading. Teachers self-reported the strategies used; the totals are shown in Table 5. The most commonly reported strategy for first grade teachers was to have such students meet with a specialist (e.g., reading specialist, reading intervention teacher, ESOL teacher, instructional assistant, or volunteer). The most commonly reported strategy for second grade teachers was to use a specific instructional strategy such as modeling, picture walk, vocabulary preview, or familiar books; the strategy varied across teachers. Another frequently reported strategy was to spend more time with the student; examples include extra meetings, longer meetings, and daily meeting.

Table 5. Teacher Strategies for Low-Performing Students

	Grade 1 (N=32)		Grade 2 (N=29)	
	n	**%**	**N**	**%**
Meet with specialist, IA, and/or volunteer	17	53	6	21
Adopt various instructional strategies	15	47	23	79
Work with them more intensively	15	47	9	31
Meet with them one on one, in pairs, or in smaller groups	10	31	4	14
Involve parents of the student	8	25	2	7
Word work, sight words	5	16	6	21
Other	4	13	2	7
None, no answer	2	6	0	0

Note 1: Two Grade 1 teachers and two Grade 2 teachers did not have any children performing below the quarterly benchmark.

Note 2: Column does not sum to 100% because teachers could give more than one response.

Professional Development to Support Implementation of the Assessment Program

During the post-observation interview, teachers reported on three aspects of professional development related to the implementation of the MCPSAP-PR. The first concerned training to implement and score the reading assessments. Out of 64 teachers who responded, about two thirds (69%) reported receiving such training within their school, including 19 teachers (30%) who mentioned working with their team and/or reading specialist. Only two teachers (3%) reported any training with other schools. The remaining 18 teachers did not receive any training this year or did not score the assessments.

Teachers also reported on the training they received to adjust instruction based on the data from the reading assessments. Among the 59 teachers who responded, the majority (38 or 64%) did receive such training. The most common method (reported by 32 teachers or 54%) was by way of meetings with their reading specialist, staff development teacher, and/or other teachers. A little over one fourth of the teachers interviewed (17 or 29%) wanted more training on this topic.

The third aspect of professional development concerned resource and support materials available to support teachers' use of assessment data. As seen in Table 6, the 60 teachers who responded mentioned a variety of materials, including other professional staff in their building (e.g., reading specialists and teammates) and materials produced by MCPS (i.e., OCIP guides and IMS reports). Only 10 teachers (17% of the 60 respondents) felt that they needed more resources to support their use of the reading assessment data.

Table 6. Frequency of Teacher Reports on Resources to Support Use of Reading Assessment Data

	Teachers (N=65)	
	n	**%**
Reading specialist, staff development teacher, resource teacher	24	37
OCIP instructional guides	15	23
Teammates	11	17
IMS reports	6	9
Level books	6	9
Other	20	31
None, no answer	12	18

Note: Column does not sum to 100% because some teachers gave more than one response.

2. *What refinement has occurred, or needs to occur, to the primary reading assessments?*

Reliability of Oral Retelling and Oral Comprehension Questions

Oral retelling was a new component of the 2002–2003 MCPSAP-PR and, beginning with this year, oral comprehension questions were used as a benchmark measure for comprehension. The reliability for these two measures has not been examined. Therefore, the inter-rater reliability of the oral retelling and oral comprehension components was examined, as follows.

Method. From each of the in-depth study schools, OSA requested information on 5 to 10 students who were in kindergarten or Grade 1 and who did oral retell for books at levels 3–9 and also on 10 to 15 students who were in Grade 1 and who did oral comprehension for books at levels 10–16. Participating schools received tape recorders, microphones, and blank tapes. Teachers were asked to record the child's total response for either oral retell or oral comprehension questions and to copy the scoring sheet for that student.

Following receipt of the tapes and scoring sheets from the schools, a trained observer listened to the tape recording of each student's response and scored it, without reference to the teacher score. Results for the observer's scoring were compared with the teachers' scoring to determine inter-rater reliability.

Results. Eighteen schools submitted 282 responses comprising of 89 for oral retell and 193 for oral comprehension questions. However, six oral retell and 15 oral comprehension questions could not be re-scored due to a damaged tape or no identification of the student on the tape.

For oral retell, the possible scores were 0, 1, 2, and 3. The teacher scores and the second scores were the same for 70 responses, representing 84% of the total. For the remaining 13 papers, the teacher scores and second scores were within one point of each other. Thus, 100% of the second scores were within one point of the first score.

For oral comprehension questions, the possible scores were 0, 1, 2, 3, 4, or 5. The teacher scores and the second scores were identical for 121 responses, representing 68% of the total. For the remaining responses, the teacher scores and second scores were within one point for 51 (29% of the total) and within two points for 5 responses (3% of the total). Thus, 97% of the double-scored responses were within one point. These results indicate a consistent use of scoring guidelines for oral retell and oral comprehension questions and suggest that MCPSAP-PR results are reliable measures of reading performance.

Reliability of Written Responses

Inter-rater reliability of written responses to comprehension questions has been examined in previous years through structured scoring workshops conducted by OSA and OCIP. As this component of the assessment program moves into its fourth year of implementation, it was determined that schools would be responsible for monitoring inter-rater reliability in scoring written comprehension responses.

Method. Administrative guidelines for MCPSAP-PR require that two teachers score each student's written response. For the written responses, there were four possible score points—3, representing "essential" understanding; 2, representing "partial" understanding; 1, representing "minimal" understanding; and 0, representing "no" understanding or no response. Scoring was holistic, meaning that the student's responses to all of the questions were taken into account. If the teachers disagreed on the score, a third teacher scored the paper. If none of the three scorers agree, the paper was scored through arbitration. All scoring forms for each paper that required three or four scores were sent to OSA after each assessment window.

Results. Overall, the written responses were scored consistently within schools, at each of the three windows (see Table 7). At each assessment window, the two scores were the same for at least 80% of the written responses and the scores were within one point of each other for at least 98% of the papers. These findings continue to support the reliability of scores reported for the MCPSAP-PR.

Table 7. Number and Percentages of Disagreements Between First and Second Scores on Written Responses for Fall, Winter and Spring Assessments

	Fall 2002 (N=8958)		Winter 2003 (N=8616)		Spring 2003 (N=16,547)	
	n	%	n	%	n	%
Responses without disagreements	7362	82	7556	88	14300	86
Responses with disagreements within one point	1513	17	987	11	1977	12
Responses with disagreements greater than one point	83	1	73	1	270	2

Internal Predictive Validity: Quarterly Benchmarks

For 2002–2003, OCIP added reading targets for the end of each quarter for Grades 1 and 2. This section examines whether these benchmarks are appropriate predictors for the annual benchmark and whether there are differences in the usefulness of these benchmarks based on racial groups or support services received.

Method. The analytical approach was to examine the relationship between whether or not students met the quarterly benchmarks and whether or not they met the end-of-year benchmark. However, the assessment data is available only for fall and winter, not for each quarter. Therefore, it was necessary to determine what benchmark would be appropriate for these two windows. Because the fall 2002 assessment window was at the beginning of the first quarter, its benchmark was set to the annual benchmark for the prior grade—for Grade 1, reading a book at level 3 or higher with a running record of 90% or higher and for Grade 2, reading a book at level 16 or higher with a running record of 90% or higher; and adequate written comprehension (i.e., a score of 2 or 3 which represents partial or essential understanding). Because the winter 2003 assessment was done during January and the second quarter concludes at the end of that month, the benchmark for winter 2003 was set to equal that for the end of the second quarter—reading a book at level 8 or higher, with a running record of 90% or higher and oral comprehension of 80% or higher for Grade 1 and reading a book at level K or higher, with a running record of 90% or higher and adequate written comprehension (i.e., a score of 2 or 3) for Grade 2.

For an interim benchmark to be a useful predictor, a high percentage of students who meet it also should have met the end-of-year benchmark; and a high percentage of students who did not meet the interim benchmark also should fail to meet the end-of-year benchmark. A phi coefficient (which ranges in value from 0 to 1) was used to measure this relationship.

Results. As seen in Table 8, more than 75% of Grade 1 students who met the reading benchmark for fall 2002 were able to meet the end-of-year benchmark and 81% of the students who met the winter 2003 benchmark were able to meet the annual benchmark. The phi coefficient equals .45 for the fall benchmark and .49 for the winter benchmark; both are statistically significant ($P < .0001$).

Table 8. Cross-Tabulation of Grade 1 Students' Performance on Fall or Winter Benchmark and End-of-Year Benchmark

	Did not meet end-of-year benchmark		**Met end-of-year benchmark**	
All students who took fall 2002 and spring 2003 assessments	**n**	**% of total**	**n**	**% of total**
Did not meet fall benchmark (n=2590)	1849	71	741	29
Met fall benchmark (n=6535)	1493	23	5042	77
All students who took fall 2002, winter 2003, and spring 2003 assessment	**n**	**% of total**	**n**	**% of total**
Did not meet winter benchmark (n=3840)	2592	68	1248	32
Met winter benchmark (n=3439)	650	19	2789	81

Phi coefficients were calculated for each of the five racial/ethnic groups and for each of the groups with and without special education, ESOL, and FARMS services. The phi coefficient for each reported group was statistically significant. (These values are reported in Table 29, Appendix C.) The numbers and percentages within each group that met the fall benchmark and also met the annual benchmark, and that met the winter and also met the annual benchmark are in Table 9. These findings support the appropriateness of the proposed fall and winter benchmarks as predictors for the annual benchmark in Grade 1 and, by extension, the appropriateness of the quarterly benchmarks on which the interim benchmarks were based. There are no differences in the usefulness of these benchmarks as predictors, based on racial group or support services received.

Table 9. Number and Percentage of Students Who Met Reading Benchmarks at Fall or Winter Assessments and Met End of Grade 1 Reading Benchmark, by Demographic Group

Student Group (Students with fall 2002 & spring 2003 scores)	Students who met fall benchmark	Students who met fall and annual benchmarks		Students who met winter benchmark	Students who met winter and annual benchmarks	
	n	n	% of all who met fall benchmark	n	N	% of all who met winter benchmark
All Students						
	6535	5042	77	3439	2789	81
Race/Ethnicity						
American Indian	11	*		19	*	
Asian American	1140	895	79	544	442	81
African American	1162	824	71	755	589	78
White	3204	2658	83	1473	1261	86
Hispanic	1010	652	65	656	491	75
Special Education						
IEP	360	224	62	171	125	73
Non-IEP	6175	4818	78	3268	2664	82
ESOL Services						
ESOL	325	175	54	217	135	62
Non-ESOL	6210	4867	78	3222	2654	82
FARMS Services						
FARMS	1183	753	64	759	566	75
Non-FARMS	5352	4289	80	2680	2223	83

* Cell size too small to report

The data for the Grade 2 benchmarks are in Tables 10 and 11. As seen in Table 10, 89% of second graders who met the reading benchmark for fall 2002 also were able to meet the end-of-year benchmark and the same percentage of students who met the winter 2003 benchmark also met the annual benchmark. The phi coefficient equals .45 for the fall benchmark and .56 for the winter benchmark; both are statistically significant ($P < .0001$).

Table 10. Cross-Tabulation of Grade 2 Students' Performance on Fall or Winter Benchmark and End-of-Year Benchmark

	Did not meet end-of-year benchmark		Met end-of-year benchmark	
All students who took fall 2002 and spring 2003 assessments	**n**	**% of total**	**n**	**% of total**
Did not meet fall benchmark (n=5676)	3077	54	2599	46
Met fall benchmark (n=3919)	414	11	3505	89
All students who took fall 2002, winter 2003 and spring 2003 assessment	**n**	**% of total**	**n**	**% of total**
Did not meet Winter benchmark (n=4054)	2779	68	1275	32
Met winter benchmark (n=2363)	263	11	2100	89

The analysis was repeated for each of the five racial/ethnic groups and for the groups with and without special education, ESOL, and FARMS services. The phi coefficient for each group that could be reported was statistically significant (see Table 29, Appendix C). The numbers and percentages within each group that met the fall benchmark and were able to meet the annual benchmark, and that met the winter benchmark and were able to meet the annual benchmark are outlined in Table 11. These findings support the appropriateness of the proposed fall and winter benchmarks as predictors for the annual benchmark in Grade 2 and, by extension, the appropriateness of the quarterly benchmarks on which the interim benchmarks were based. The usefulness of these benchmarks as predictors did not vary based on racial group or support services received.

Table 11. Students Who Met Reading Benchmarks at Fall or Winter Assessments and Met Annual Grade 2 Reading Benchmark, by Demographic Group

Student Group (Students with fall 2002 & spring 2003 scores)	Students who met fall benchmark	Students who met fall and annual benchmarks		Students who met winter benchmark	Students who met winter and annual benchmarks	
	n	n	% of all who met fall benchmark	n	n	% of all who met winter benchmark
All Students						
	3919	3505	89	2363	2100	89
Race/Ethnicity						
American Indian	11	10	91	*	*	
Asian American	655	603	92	381	335	88
African American	598	497	83	507	441	87
White	2262	2070	92	1038	954	92
Hispanic	393	325	83	433	367	85
Special Education						
IEP	135	112	83	90	73	81
Non-IEP	3784	3393	90	2273	2027	89
ESOL Services						
ESOL	96	73	76	165	128	78
Non-ESOL	3823	3432	90	2198	1972	90
FARMS Services						
FARMS	540	446	83	597	493	83
Non-FARMS	3379	3059	91	1766	1607	91

* Cell size too small to report.

Internal Predictive Validity: Yearly Benchmarks

This section examines whether the annual benchmarks are appropriate predictors for the attainment of benchmark performance in subsequent grades and whether there are differences in the usefulness of these benchmarks as predictors, based on racial groups or support services received.

Kindergarten and Grade 1. For students in Grade 1 for 2002–2003, there was information from kindergarten on the annual benchmark of reading a text at level 3 or above, with a running record of 90% or higher.

As with the interim benchmarks, for an end-of-year benchmark to be an appropriate predictor, a majority of students who meet it also should go on to meet the end-of-year benchmark for the following grade, and likewise, a majority of students who fail to meet it would fail to meet the end-of-year benchmark for the following grade. As seen in Table 12, more than 75% of Grade 1 students who met the annual kindergarten benchmark went on to meet the annual Grade 1 benchmark and more than 60% of students who did not meet the kindergarten benchmark also

failed to meet the Grade 1 annual benchmark. The phi coefficient for these two benchmarks equals .43 and is statistically significant ($P < .0001$).

Table 12. Cross Tabulation of Grade 1 Students' Performance on Annual Kindergarten Benchmark and Annual Grade 1, Benchmark, Fall 2002 and Spring 2003

All students who took fall 2002 and spring 2003 assessments	Did not meet Grade 1 benchmark		Met Grade 1 benchmark	
	n	% of total	N	% of total
Did not meet Kindergarten benchmark (N=3133)	2013	64	1120	36
Met Kindergarten benchmark (N=4865)	1055	22	3810	78

Phi coefficients were calculated for each of the five racial/ethnic groups and for each of the groups with and without special education, ESOL, and FARMS services. The phi coefficient for each reported group was statistically significant (see Table 30, Appendix C). The number and percentage within each group that met the kindergarten benchmark and the Grade l benchmark are outlined in Table 13. These findings support the usefulness of the kindergarten benchmark for predicting Grade 1 benchmark performance; there are no differences in its usefulness based on racial group or support services received.

Table 13. Number and Percentage of Grade 1 Students Who Met Annual Reading Kindergarten Benchmark and Met Annual Grade 1 Reading Benchmark, by Demographic Group

Student Group (Students with fall 2002 & spring 2003 scores)	Students who met kindergarten benchmark	Students who met kindergarten and Grade 1 annual benchmarks	
	n	n	% of all who met K benchmark
All Students			
	4865	3810	78
Race/Ethnicity			
American Indian	16	12	75
Asian American	853	680	80
African American	841	615	73
White	2381	1989	83
Hispanic	774	514	66
Special Education			
IEP	276	174	63
Non-IEP	4589	3636	79
ESOL Services			
ESOL	186	95	51
Non-ESOL	4679	3715	79
FARMS Services			
FARMS	974	634	65
Non-FARMS	3891	3176	82

Kindergarten and Grade 2. For students in Grade 2 for 2002–2003, there was information on their performance on the annual kindergarten benchmark and the annual Grade 1 benchmark. As seen in Table 14, 85% of students who met the annual kindergarten benchmark went on to meet the annual Grade 2 benchmark, but only 50% of students who did not meet the kindergarten benchmark also failed to meet the Grade 2 annual benchmark. The phi coefficient for these two benchmarks equals .36 and is statistically significant ($P < .0001$). For the annual Grade 1 benchmark, 84% of Grade 2 students who met it also met the annual Grade 2 benchmark and 63% of the students who failed the Grade 1 benchmark also failed the Grade 2 benchmark. The phi coefficient for these two benchmarks equals .49 and is statistically significant ($P < .0001$); it is higher than that for the kindergarten benchmark because the percentage who failed to meet both benchmarks is higher for the Grade 1 benchmark (i.e., 63% vs. 50%).

Table 14. Cross-Tabulation of Grade 2 Students' Performance on Annual Kindergarten or Grade 1 Benchmarks, by Annual Grade 2 Benchmark

	Did not meet end of Grade 2 benchmark		Met end of Grade 2 benchmark	
	n	**% of total**	**n**	**% of total**
All students who took spring 2001, fall 2002 and spring 2003 assessments				
Did not meet kindergarten benchmark (n=4323)	2171	50	2152	50
Met kindergarten benchmark (n=3043)	471	15	2572	85
All students who took spring 2002, fall 2002 and spring 2003 assessments				
Did not meet end of Grade 1 benchmark (n=3611)	2276	63	1335	37
Met end of Grade 1 benchmark (n=5081)	801	16	4280	84

Phi coefficients were calculated for each of the five racial/ethnic groups and for each of the groups with and without special education, ESOL, and FARMS services, for both the kindergarten and Grade 1 benchmarks. The phi coefficient for each group for each benchmark was statistically significant (see Table 31 in Appendix C). The number and percentage within each group that met the kindergarten benchmark and the Grade 2 benchmark, and also that met the Grade 1 benchmark and the Grade 2 benchmark are in Table 15. These findings support the usefulness of both the kindergarten benchmark and Grade 1 benchmark for predicting Grade 2 benchmark performance; there are no differences in their usefulness based on racial groups or support services received.

Table 15. Number and Percentage of Grade 2 Students Who Met Annual Reading Benchmarks by the End of Kindergarten (K) or Grade 1 and Went on to Meet the Annual Grade 2 Reading Benchmark, By Demographic Group

Student Group (Students with fall 2002 & spring 2003 scores)	Students who met kindergarten benchmark	Students who met kindergarten and Grade 2 benchmarks		Students who met Grade 1 benchmark	Students who met Grade 1 and Grade 2 benchmarks	
	n	n	% of all who met K benchmark	n	N	% of all who met first grade benchmark
All Students						
	3043	2572	85	5081	4280	84
Race/Ethnicity						
American Indian	5	5	100	15	12	80
Asian American	520	452	87	809	703	87
African American	470	367	78	847	667	79
White	1652	1460	88	2797	2447	87
Hispanic	396	288	73	613	451	74
Special Education						
IEP	138	87	63	196	140	71
Non-IEP	2905	2485	85	4885	4140	85
ESOL Services						
ESOL	79	47	60	163	101	62
Non-ESOL	2964	2525	85	4918	4179	85
FARMS Services						
FARMS	544	395	73	852	617	72
Non-FARMS	2499	2177	87	4229	3663	87

External Predictive Validity: Comprehensive Test of Basic Skills (CTBS) Scores

This section examines whether student performance on the MCPSAP, a locally developed assessment, predicts performance on the CTBS, a nationally standardized test. Grade 2 students took the CTBS in spring 2003.

Method. There is not a benchmark score for CTBS scores; therefore, this analysis uses student scale scores from CTBS. Specifically, a correlation coefficient (R) between each of two CTBS scale scores, reading and language, and each reading measure from MCPSAP-PR were calculated. A statistically significant and positive relationship between the measures means that students with higher performance on MCPSAP-PR have higher scale scores on the CTBS. Such a result supports the conclusion that student performance on the MCPSAP predicts performance on the CTBS.

The available MCPSAP-PR reading measures were the student's text-reading level with at least 90% running record for kindergarten from spring 2001, plus text-reading level with at least 90%

running record and adequate comprehension for Grade 1 from spring 2002, and for Grade 2 from winter 2003.

Results. For all the MCPSAP-PR measures, the correlation coefficients with CTBS scores were statistically significant and positive, as seen in Table 16. These results suggest that reading performance on MCPSAP-PR is predictive of performance on the reading and language subtests of the CTBS.

Table 16. Correlations Between Reading Performance on MCPSAP-PR and CTBS Scale Scores

Text-reading level from MCPSAP-PR	CTBS scale scores from spring 2003					
	Language			Reading		
	n	r	p	n	r	p
Spring 2001, end of kindergarten	5187	.42	.01	5188	.45	.01
Spring 2002, end of Grade 1	5265	.49	.01	5265	.48	.01
Winter 2003 in Grade 2	5030	.62	.01	5028	.58	.01

External Predictive Validity: Maryland State Assessment (MSA) Scores

This section examines whether the annual benchmarks for the MCPSAP-PR, a locally developed assessment, predicts performance on the MSA, a statewide, standardized test, and further, whether there are differences in the usefulness of these benchmarks as predictors, based on racial groups or support services received.

Method. Students in Grade 3 in 2002–2003 took the MSA in March 2003; their benchmark performance from Grade 2 and Grade 1 were available for analysis. The benchmark scale score for MSA reading in Grade 3 was 404. Phi coefficients were calculated for the relationship between the Grade 1 benchmark and MSA benchmark and also the Grade 2 and the MSA benchmarks.

Results. Table 17 includes students who took reading assessments in Grades 1 or 2 and then in Grade 3 took the MSA. Among the students who met the annual Grade 1 benchmark in spring 2001, 90% went on to meet the MSA benchmark. However, among those who failed to meet the benchmark in Grade 1, 54% also failed to meet the MSA benchmark. The phi coefficient for these two benchmarks equals .48 and is statistically significant ($P < .0001$). For the annual Grade 2 benchmark, 93% of Grade 3 students who met it went on to meet the MSA benchmark and 58% of the students who failed the Grade 2 benchmark also failed to meet the MSA benchmark. The phi coefficient for these two benchmarks equals .54 and is statistically significant ($P < .0001$).

Table 17. Cross-Tabulation of Grade 3 Students' Performance on Annual Grade 1 or Grade 2 Benchmarks, by MSA Benchmark

	Did not meet MSA benchmark		Met MSA benchmark	
	n	**% of total**	**n**	**% of total**
All students who took spring 2001assessments				
Did not meet Grade 1 benchmark (n=3474)	1893	54	1581	46
Met Grade 1 benchmark (n=4086)	426	10	3660	90
All students who took spring 2002 assessments				
Did not meet end of Grade 2 benchmark (n=4282)	2500	58	1782	42
Met end of Grade 2 benchmark (n=4954)	383	7	4571	93

Phi coefficients were calculated for each of the five racial/ethnic groups and for each of the groups with and without special education, ESOL, and FARMS services, for both the Grade 1 and Grade 2 benchmarks. The phi coefficient for each group for each benchmark was statistically significant, except for American Indians in both Grade 1 and Grade 2 (see Table 32 in Appendix C). The number and percentage within each group that met the Grade 1 benchmark and went on to meet the MSA benchmark, and also that met the Grade 2 benchmark and went on to meet the MSA benchmark are in Table 18. These findings support the usefulness of both the Grade 1 benchmark and Grade 2 benchmark for predicting MSA benchmark performance; there are no differences in their usefulness based on racial groups or support services received.

Table 18. Number and Percentage of Grade 3 Students Who Met Annual Reading Benchmark by the End of Grade 1 or End of Grade 2 and Met Standard for Proficient on MSA, by Demographic Group

Student group	Students who met Grade 1 benchmark	Students who met Grade 1 and MSA benchmarks		Students who met Grade 2 benchmark	Students who met Grade 2 and MSA benchmarks	
	n	n	% of all who met Grade 1 benchmark	n	n	% of all who met Grade 2 benchmark
All Students						
	4086	3660	90	4954	4571	92
Race/Ethnicity						
American Indian	10	9	90	17	16	94
Asian American	682	632	93	841	799	95
African American	643	488	76	730	608	83
White	2322	2211	95	2880	2761	96
Hispanic	429	320	75	486	387	80
Special Education						
IEP	141	108	77	188	159	85
Non-IEP	3945	3552	90	4766	4412	93
ESOL Services						
ESOL	225	163	72	136	94	69
Non-ESOL	3861	3497	91	4818	4477	93
FARMS Services						
FARMS	652	466	72	696	546	78
Non-FARMS	3434	3194	93	4258	4025	95

3. What were the students' academic outcomes in reading for Grades 1 and 2?

Methodology

For the MCPSAP-PR, teachers assessed each student individually during the Fall testing window (Sept 17, 2002–Oct 11, 2002) and the spring testing window (May 12, 2003–June 6, 2003). During the winter testing (January 6, 2003–January 24, 2003), teachers assessed every student at schools with full-day kindergarten. In other schools, teachers were directed to assess students who were reading instructionally below the quarterly target, specifically those at level K or below for second graders and at level 11 or below for first graders. Following the assessment, school-based personnel entered the scores into the MCPS reading assessment database, which was used for the analyses in this report

MCPSAP–PR consists of two components—foundational reading skills and reading proficiency. The foundational reading skills are the building blocks that precede text reading; these skills are

the focus of teaching and assessing in kindergarten. In Grades 1 and 2, the focus shifts to text reading. The reading proficiency component assesses reading fluency and comprehension.

Foundational Reading Skills. All students new to MCPS and all Grade 1 and Grade 2 students at high risk for not achieving reading benchmarks are to be assessed in the foundational skills until the student has mastered them. (Teachers receive guidelines to identify high-risk students.)

The foundational reading skills component assesses six skills necessary for a student to become a strong reader: letter knowledge, print concepts, oral language, phonemic awareness, phonics, and reading vocabulary. Details on each of these tests are outlined in Report on Kindergarten Student Progress in Reading for 2002–03 (Curry-Corcoran, 2003). School staff enter the raw score for each foundational skill test. Benchmark performance targets have been established for each test.

Reading Proficiency. The reading proficiency component assesses four areas—accuracy of reading, reading behaviors, oral reading fluency, and comprehension of the text. School staff enter the book title and appropriate scores for that title in the database.

Accuracy of reading is assessed with a running record. In giving this assessment individually to a student, the teacher first selects a book at an appropriate level of difficulty and then asks the student to read it aloud. Errors in word recognition are recorded as the student reads. If word recognition accuracy is less than 90%, the teacher should select a lower-level text and repeat the procedure until the 90% criterion is met.

While the student is reading the text aloud, the teacher assesses reading behaviors—such as concepts about print, word recognition, and phonemic awareness; and oral reading fluency—the student's use of expression and ability to divide the text into meaningful chunks.

Comprehension is measured in the manner most appropriate for the given reading level. For texts at levels 3 to 9, the required comprehension task is oral retell, which allows the student to demonstrate story structure understanding of narration text. After a second reading of the book (following a reading to complete the running record), the teacher directs the student to start at the beginning and tell what happened in this story. Prompts are available to allow the student to convey everything he/she remembers and understands from the story.

For text levels 10 to 16, the required comprehension tasks include oral comprehension questions and a written response. The former allows the student to demonstrate understanding of explicit and implicit details from the book. After the student reads the text, the teacher asks five questions that are either explicit (the information is right there) or implicit (the student needs to think and search). Following these oral questions, the student completes a written response by answering two questions.

For text levels J to P, the required comprehension task is a written response. Subsequent to reading the text, the student answers five questions. The student is expected to demonstrate understanding of the text through four stances—global understanding, developing interpretation, personal response, and critical analysis. The oral comprehension questions are optional; when

the written response appears weak (they are not scored immediately), teachers are to administer these questions. Each student's written response is scored by at least two teachers. For the written responses, there are four possible score points, as follows—

- 3, representing essential understanding
- 2, representing partial understanding
- 1, representing minimal understanding
- 0, representing no understanding or no response.

Scoring is holistic, meaning that the student's responses to all of the questions are taken into account. If the teachers disagree on the score, a third teacher scores the paper. If no two of the three scorers agree, the paper is scored through arbitration.

To show proficiency at a reading level, a student must have a running record of 90% or higher and adequate comprehension. Level 2, the lowest text level on the MCPSAP-PR, is an exception and requires neither an adequate running record nor a comprehension measure. Measures of adequate comprehension vary by the reading level, as follows:

- Levels 3–9, 2 or 3 out of 3 on oral retell
- Levels 10–16, 4 or 5 out of 5 on oral comprehension
- Levels J and above, 2 or 3 out of 3 on written response.

MCPS has set targets for student achievement by the end of the year for kindergarten through Grade 2. Using data from the 2001–2002 school year and the input of reading specialists familiar with the texts used in the MCPS Assessment Program, the proficiency benchmarks for 2002–2003 for Grades 1 and 2 were set as follows:

- The Grade 1 target is reading a text at level 16 or higher with an accuracy rate of 90% or higher, along with a score of 80 to 100% on oral comprehension.
- The Grade 2 target is reading a text at Level M or higher with an accuracy rate of 90% or higher, along with a score of 2 or 3 on the written response, representing partial or essential understanding.

Analysis. Descriptive statistics were used to summarize reading performance in Grades 1 and 2 students. Tests of proportions were used to examine differences in percentages. A Kruskal-Wallis test was used to determine whether variations in reading level were related to race/ethnicity. Mann-Whitney tests were used to test differences in reading level between pairs of student groups (the associated Z scores are reported).

Results for students in Grade 1

The following results are based on Grade 1 students with test scores for the windows reported. Demographic information about these students is in Table 19.

Table 19. Characteristics of Grade 1 Students With Reading Scores for 2002–2003

Student group	All Students with spring 2003 scores		All Students with fall 2002 & spring 2003 scores	
	n	%	n	%
All Students				
	9914	100	9434	100
Race/Ethnicity				
American Indian	32	< 1	31	< 1
Asian American	1522	15	1435	15
African American	2019	20	1885	20
White	4286	43	4150	44
Hispanic	2055	21	1933	20
Special Education				
IEP	812	8	792	8
Non-IEP	9102	92	8642	92
ESOL Services				
ESOL	1176	12	1031	11
Non-ESOL	8738	88	8403	89
FARMS Services				
FARMS	2256	23	2208	23
Non-FARMS	7658	77	7226	77
Combination				
Non-ESOL & Non-FARMS	7039	71	6733	71
ESOL & FARMS	557	6	538	6

Benchmark. Performance In 2002–2003, more than 60% of Grade 1 students met the annual benchmark target. Of the 9434 Grade 1 students with both fall and spring test scores, 62.0% met the annual benchmark and of the 9914 students with spring scores, 61.2% met the benchmark. Students who did not meet the benchmark varied considerably in their reading performance (see Table 20). The grade levels corresponding to the various text levels are based on information in the instructional guides for Grades 1 and 2 (see page 58 in Office of Instruction and Program Development, 2002b).

Table 20. Text Level for Spring 2003 of Grade 1 Students Who Did Not Meet The Annual Benchmark

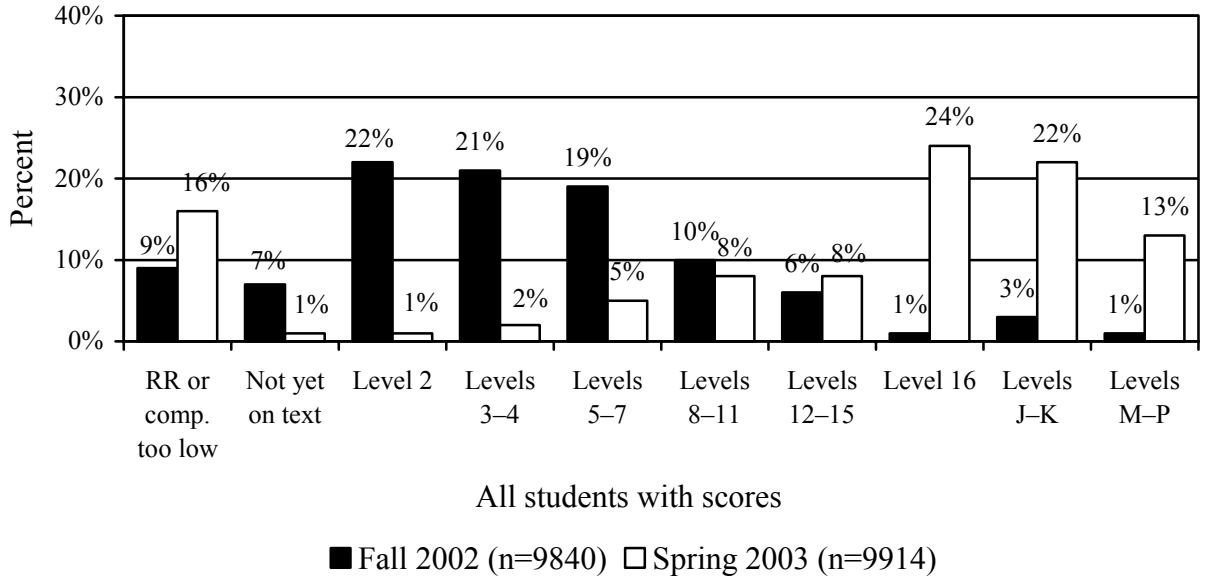
All students with spring scores	Number of students at this level	% of all students	
		Not at benchmark (N=3843)	With spring scores (N=9914)
	n	%	%
Text level indeterminate (Running record < 90%)	119	3	1
Kindergarten (not yet on text to level 4)*	403	10	4
First quarter of Grade 1 (levels 5–7)*	522	14	5
Second quarter of Grade 1 (levels 8–11)*	959	25	10
Second half of Grade 1 (levels 12–15)*	851	22	9
Grade 1 benchmark (level 16) without adequate oral comprehension	361	9	4
Above Grade 1 benchmark (levels J–P) without adequate written comprehension	628	16	6

* Includes students with and without adequate comprehension.

Of the 3843 students that did not meet the Grade 1 benchmark, 25 % were able to read a text at the benchmark level (16) or higher (levels J–P) with an adequate running record, but failed to meet the target, due to scores of 0 or 1 on written comprehension. To categorize all of the remaining students, those without adequate comprehension were assigned to a text level if they read it accurately (i.e., had a running record above 90%); thus, the categories in Table 20 are suggestive, not precise. There were students who read a book at the kindergarten level and thus were one grade level behind; they represent 4% of those with spring scores. Another 15% of all students read at the target level set for the first half of Grade 1.

Reading Levels. For fall 2002, assessment scores for 9840 Grade 1 students were available. As seen in Figure 1, the reading levels for these students were spread across a wide range of values from not on text, for students not yet reading, to level M, which is the benchmark text for the end of Grade 2. (Texts above level M were not available for the fall 2002 assessments). Among the 8920 students with a designated reading level, the median value was level 4.

Figure 1. Reading Levels of Grade 1 Students, Fall 2002 and Spring 2003



In Fall 2002, the reading level was not identifiable for 920 students because they had a running record or comprehension level that was too low. Of this group, the majority (495 students or 5% of all students) read at Level 16 or below with inadequate oral comprehension. There were also 182 students (2% of all students) with running records below 90%. According to test procedures, teachers should have retested these students with inadequate oral comprehension and with inadequate running records using a lower text level; such students represented 7% of all those tested in fall 2002. The rest of this group without a reading level included 243 students (2% of all students) who read a text at level J or higher with a running record of 90% or higher, but had scores of 0 or 1 on the written response. Teachers scored oral retell, oral comprehension questions, and running records during the assessment, but written responses were scored after the assessment because they required two scores. Therefore, teachers had to score written responses with others; by the time this scoring was completed, teachers typically were unable to repeat the assessment so as to retest the student at a lower level and were not required to do so.

Assessment scores for 9914 Grade 1 students were available for spring 2003; their reading levels are presented in Figure 1. The median reading level was 16. The majority of students (59%) read at this level or higher, including levels M–P, which are considered a Grade 2–3 level. For spring 2003, there were 1,589 students without a reading level. They included two groups that were comparable in size to fall 2002—those with inadequate running records (118 students or 1% of all students) and those who read at level 16 or below with inadequate oral comprehension (678 students or 7% of all students). But, there was a larger group that read a text at Level J or above with a running record of 90% or higher, but with scores of 0 or 1 on the written response (791 students or 8% of all students). The increase in the latter may reflect that at the end of the year more Grade 1 students read at levels that require written comprehension.

Gaps in Benchmark Performance. The percentages of Grade 1 students achieving the end-of-year benchmark are presented in Table 21, according to race/ethnicity and support

services received. To analyze patterns of growth over the year, Table 21 also includes the percentages of students who met the fall benchmark in the fall. This benchmark was reading a book at level 3 or higher with a running record of 90% or higher and with adequate comprehension. Comprehension was included to make the benchmark more comparable to the Grade 1 end-of-year benchmark that includes comprehension.

Table 21. Percentages of Grade 1 Students Achieving Reading Benchmarks, by Demographic Groups

Student group	All students with both fall 2002 & spring 2003 scores	Students who met the fall benchmark with comprehension in fall 2002*		Students who met the annual benchmark by spring 2003		Increase (or decrease) in percentage from fall 2002 to spring 2003
	n	n	%	n	%	%
All Students						
	9434	6040	61	5852	62	1
Race/Ethnicity						
American Indian	31	16	48	15	48	0
Asian American	1435	1004	66	990	69	3
African American	1885	1084	55	1014	54	(1)
White	4150	2956	69	2997	72	3
Hispanic	1933	980	48	836	43	(5)
Special Education						
IEP	792	318	39	269	34	(5)
None	8642	5722	63	5583	65	2
ESOL Services						
ESOL	1031	305	29	292	28	(1)
Non-ESOL	8403	5735	65	5560	66	1
FARMS Services						
FARMS	2208	1115	48	919	42	(6)
Non-FARMS	7226	4925	65	4933	68	3
Combination						
Non-ESOL & Non-FARMS	6733	4785	68	4768	71	3
ESOL & FARMS	538	165	30	127	24	(6)

* Fall benchmark set to level 3 or above with a running record of 90% or above and adequate comprehension.

Among all Grade 1 students, the percentage that met the fall benchmark with comprehension was nearly identical to the percentage that met the end-of-year benchmark by spring. As shown in Table 22, this pattern of steady growth from fall to spring was consistent across all the subgroups, suggesting that the gap between subgroups in the fall changed very little over the year. The biggest changes were decreases of 5% among Hispanic students and special education students and decreases of 6% among students receiving FARMS and among students receiving both FARMS and ESOL.

There were differences between groups in the percentages that met the annual benchmark by the end of the year. The following differences were all statistically significant (see Appendix D). Among the racial/ethnic groups, Asian American students had the highest percentage of students who achieved the benchmark followed by white students, African-American students and Hispanic students. The impact of poverty and second language learning can be seen again; for both FARMS and ESOL, fewer students receiving the service reached the benchmark compared with students not receiving the service. Finally, fewer students with an IEP achieved the benchmark, compared with students without an IEP.

Gaps in Reading Levels. The median reading level for fall and spring are presented in Table 22, according to race/ethnicity and support services received. Because the difference in difficulty between each reading level and the one above it varies, the median is used.

Table 22. Reading Levels of Grade 1 Students, by Demographic Groups

Student group (Students with both fall & spring scores)	**Reading level in Fall 2002***		**Reading level in Spring 2003***	
	n	**Median**	**n**	**Median**
All Students				
	8542	4	7925	16
Race/Ethnicity				
American Indian	26	3	26	16
Asian American	1224	6	1143	20
African American	1714	3	1542	16
White	3751	5	3608	18
Hispanic	1827	3	1606	16
Special Education				
IEP	730	2	644	13
Non-IEP	7812	4	7281	16
ESOL Services				
ESOL	984	2	862	11
Non-ESOL	7558	4	7063	16
FARMS Services				
FARMS	2033	3	1808	14
Non-FARMS	6509	4	6117	16

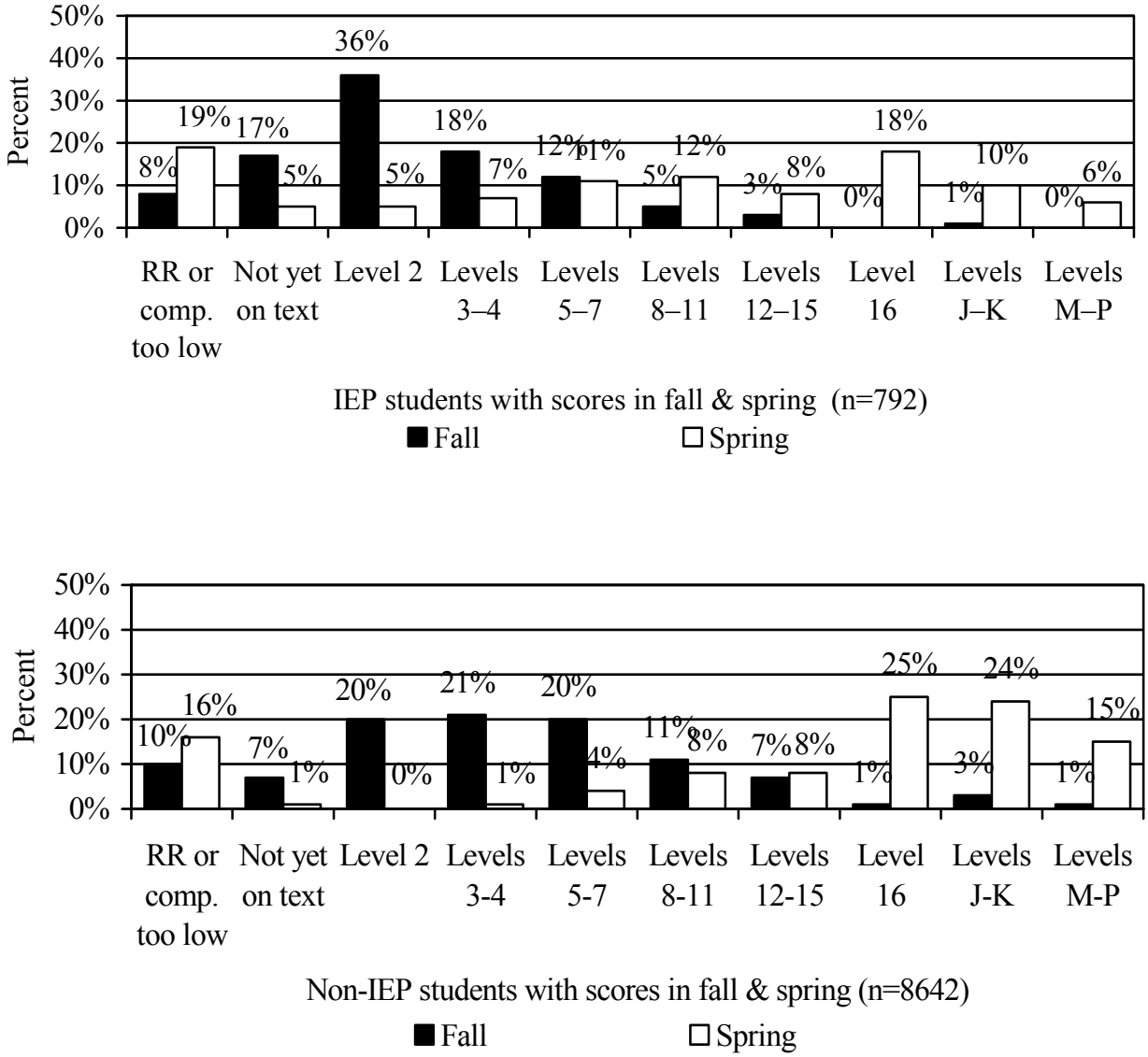
*The total number differs between these columns because some students had inadequate comprehension and thus no identifiable reading level in fall or spring.

There were differences in reading level at both fall and spring among the racial/ethnic groups. At each assessment window, the reading levels of Asian American students were highest, followed by white students, African American students, and Hispanic students. These differences reflect a statistically significant relationship between racial/ethnic groups and fall reading levels (χ^2=652.1, $p < .0001$) and also spring reading levels (χ^2=789.9, $p < .0001$).

In fall 2002, the median reading level for IEP students was 2, which is the lowest reading level in the MCPSAP-PR. The reading levels of special education students were clustered around 2; the

majority (71%) of these students read at level 4 or below (see Figure 2). In contrast, the reading levels of non-IEP students were more evenly distributed across level 2 (20%), level 3–4 (21%) and level 5–7 (20%). By spring 2003, the reading levels of IEP students were more widely dispersed; only 26% of students read at levels 12–16 (a range that includes the median level of 13). But the majority of non-IEP students (64%) were reading at level 16 and above in spring. Thus, there was a statistically significant gap in both fall ($z=15.6$, $p < .0001$) and spring reading levels ($z=18.7$, $p < .0001$) of IEP students compared with non-IEP students.

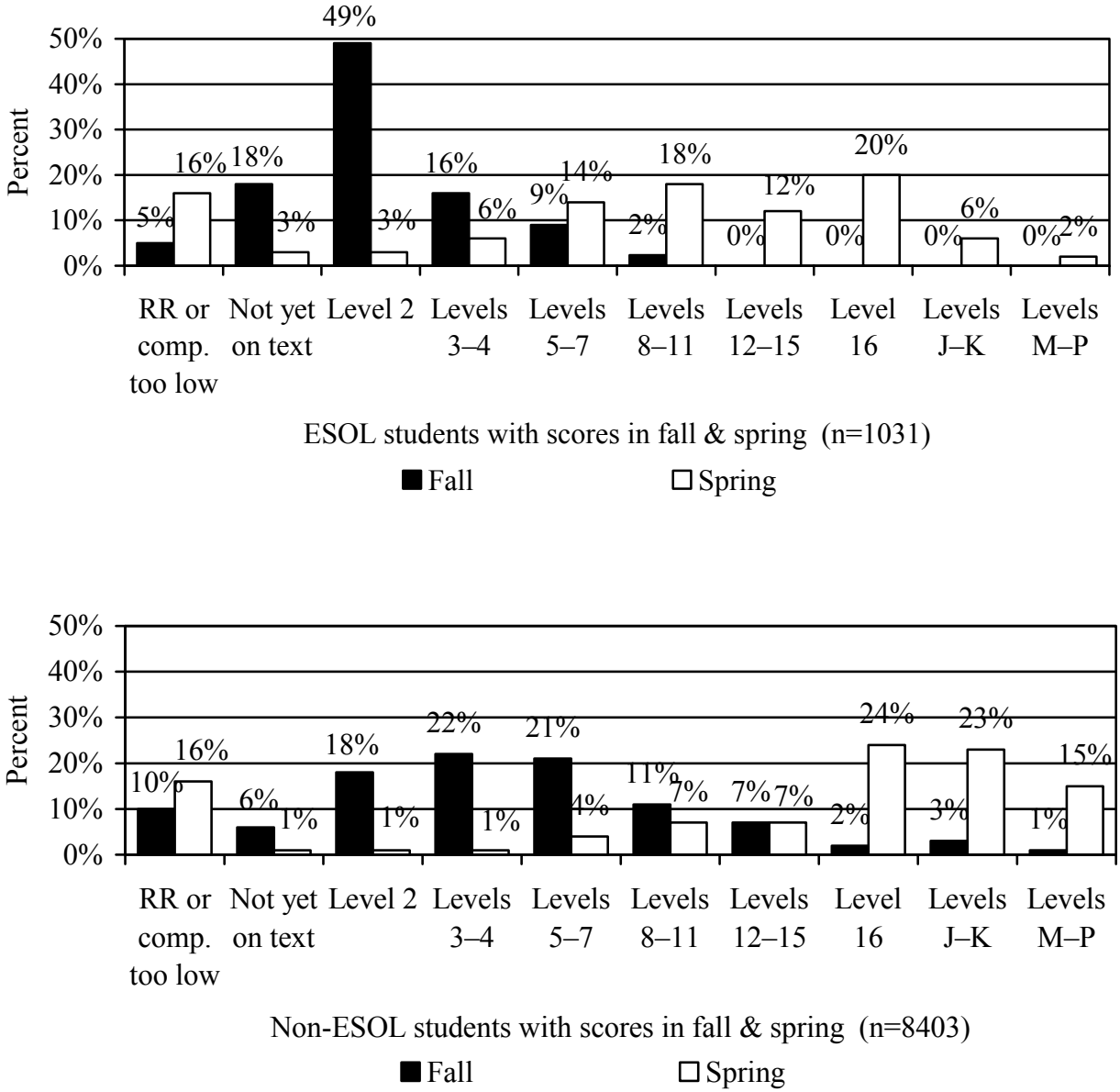
Figure 2. Reading Levels of Grade 1 IEP and Non-IEP Students for Fall 2002 and Spring 2003



In fall 2002, nearly half of ESOL students (49%) read a level 2 book, which was the median reading level for this group, while only 22% of non-ESOL students read at level 3–4, which

included their median level of 4 (see Figure 3). In spring 2003, the reading levels of non-ESOL students were more concentrated than ESOL students; 62% of the former read at levels 16–P while 58% of ESOL students read at levels 8–P. The pattern of differences between ESOL and non-ESOL students also included statistically significant gaps; ESOL Students had lower reading levels in both fall ($z=27.0$, $p < .0001$) and spring ($z=27.2$, $p < .0001$).

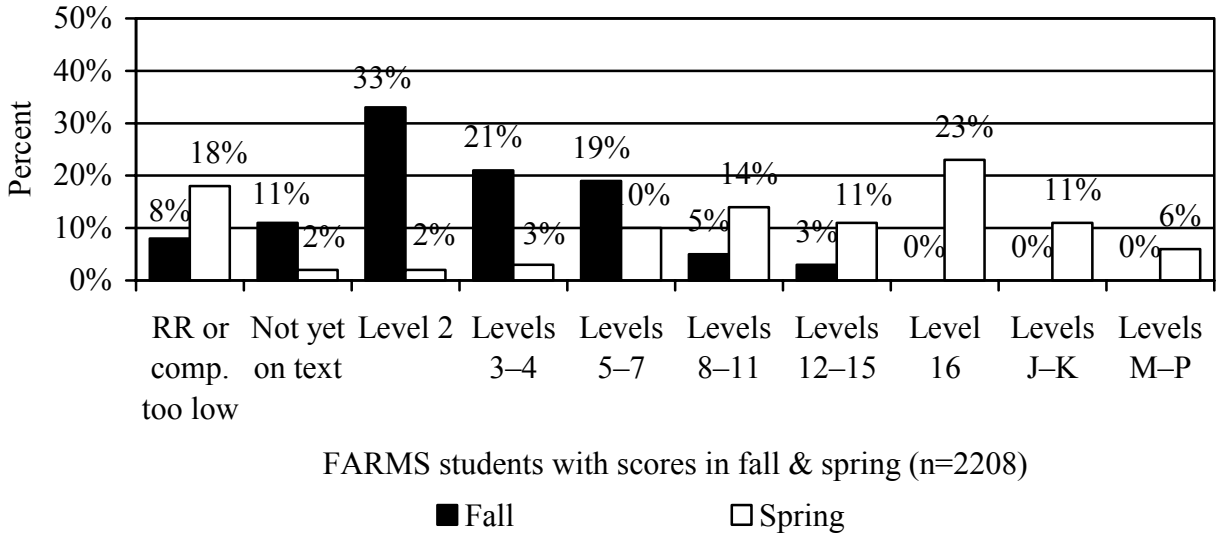
Figure 3. Reading Levels of Grade 1 ESOL and Non-ESOL Students for Fall 2002 and Spring 2003

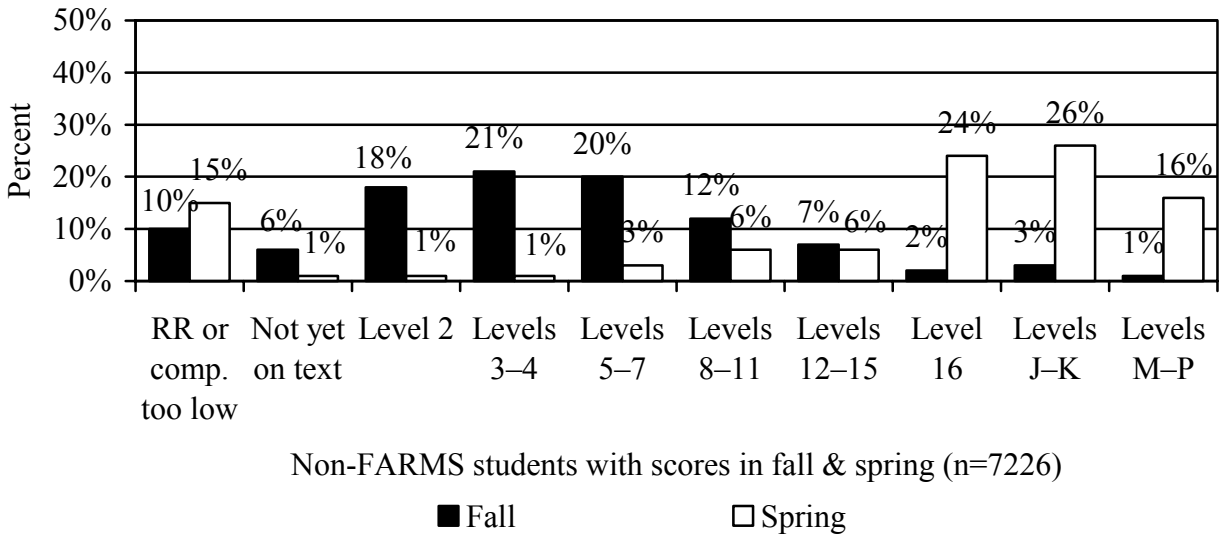


As with special education and ESOL, the reading levels of students receiving FARMS were more concentrated in the fall (see Figure 4). Specifically, in Fall 2002, 65% of FARMS students read at level 4 or lower compared with 45% for non-FARMS students. In spring 2003, the non-

FARMS students were more concentrated; 66% of them read at level 16 or above, compared with 40% for FARMS students. These differences reflect statistically significant gaps between non-FARMS and FARMS students in both fall reading level ($z=20.5$, $p < .0001$) and spring reading level ($z=25.5$, $p < .0001$).

Figure 4. Reading Levels of Grade 1 FARMS and Non-FARMS Students for Fall 2002 and Spring 2003





Results for students in Grade 2

The following results are based on Grade 2 students with test scores for the windows reported. Demographic information about these students is in Table 23.

Table 23. Characteristics of Grade 2 Students With Reading Scores for 2002–2003

Student Group	All students with spring scores		All students with fall & spring scores	
	N	%	n	%
All Students				
	9875	100	9595	100
Race/Ethnicity				
American Indian	31	<1	31	<1
Asian American	1412	14	1365	14
African American	2048	21	1969	21
White	4463	45	4369	45
Hispanic	1921	20	1861	19
Special Education				
IEP	912	9	884	9
Non-IEP	8963	91	8711	91
ESOL Services				
ESOL	1207	12	1117	12
Non-ESOL	8668	88	8478	88
FARMS Services				
FARMS	2753	28	2634	27
Non-FARMS	7122	72	6961	73
Combination				
Non-ESOL & Non-FARMS	6685	68	6553	68
ESOL & FARMS	770	8	709	7

Benchmark Performance. During 2002–2003, more than 63% of Grade 2 students met the annual benchmark target. Of the 9,595 Grade 2 students with both fall and spring test scores, 63.6% met the annual benchmark, and of the 9,875 students with spring scores, 63.1% met the benchmark. Students who did not meet the benchmark varied considerably in their reading performance (see Table 24). The grade levels corresponding to the various text levels are based on information in the instructional guides for Grades 1 and 2 (Office of Instruction and Program Development, p. 8, 2002b).

Table 24. Text Level for Spring 2003 of Grade 2 Students Who Did Not Meet the End-of-Year Benchmark

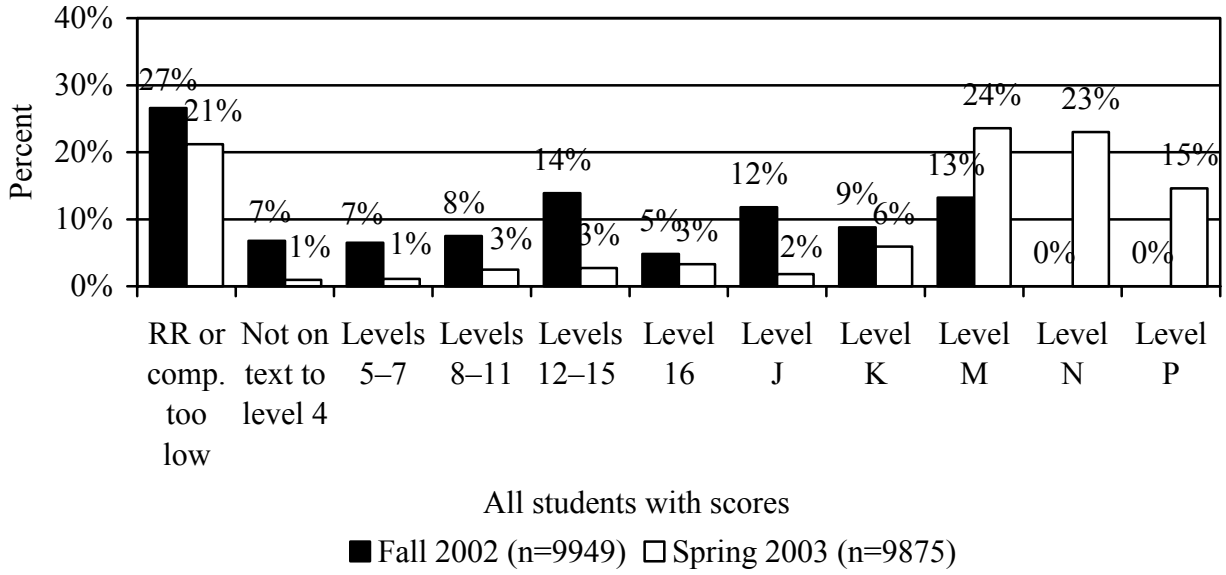
All students with spring scores	Number of students at this level	% of all students	
		Not at benchmark (n=3647)	With spring scores (n=9875)
	N	%	%
Text-level indeterminate (running record < 90%)	72	2%	1%
Kindergarten (not yet on text to level 4)*	100	3%	1%
First half of Grade 1 (levels 5–11)*	413	11%	4%
Second half of Grade 1 (levels 12–16)*	727	20%	7%
First half of Grade 2 (levels J–K)*	1107	30%	11%
Grade 2 benchmark (level M) without adequate written comprehension	771	21%	8%
Above Grade 2 benchmark (levels N & P) without adequate written comprehension	457	13%	5%

* Includes students with and without adequate comprehension.

Of the 3,647 students who did not meet the Grade 2 benchmark, 34% were able to read a text at the benchmark level (M) or higher (N or P) with an adequate running record, but failed to meet the benchmark because their written response showed minimal or no understanding. To categorize all of the remaining students, those without adequate comprehension were assigned to a text level if they read it accurately (i.e., had running record above 90%); thus, the categories in Table 24 are suggestive, not precise. There were still students who read a book at the kindergarten or first grade level and thus were at least one grade level behind; they represent 34% of those who did not meet the benchmark. Additionally, 30% of students who did not meet the benchmark read at the target level set for the first half of Grade 2.

Reading Levels. For fall 2002, assessment scores for 9,949 Grade 2 students were available. The reading levels for this window were spread across a wide range of values, from not on text, for students not yet reading, to level M, which is the benchmark text for the end of Grade 2 (see Figure 5). (Texts above level M were not available for the Fall 2002 assessments). Among students with a designated reading level, the median value was level 16, which is the benchmark level for Grade 2 at the fall assessment window.

Figure 5. Reading Levels for Grade 2 Students for Fall 2002 and Spring 2003



In fall 2002, the reading level was not identifiable for 2,633 students, representing 27% of the students. This group included 705 students (7% of all students) who read at level 16 or below with inadequate oral comprehension, and 200 students (2% of all students) with running records below 90%. Thus, because teachers did not follow the appropriate test procedures, 905 students (9% of all those tested in the fall) did not have an identifiable reading level. The majority of the group without a reading level (1,728 or 17% of all students) read a text at level J or higher with a running record of 90% or higher, but had scores of 0 or 1 on the written response.

Assessment scores for 9,875 Grade 2 students were available for spring 2003. Very few students read at levels J and below (13% of the total, see Figure 5). The median reading level was M and the majority of students (62%) read at level M or higher, including levels N and P, which are considered at a Grade 3 level. There was still a sizable group of students without a reading level, but it was smaller compared with fall 2002. For spring 2003, this group included 1,820 students (18% of the total) who read at level J and above with inadequate written comprehension, 199 students (2% of the total) who read at level 16 or below with inadequate oral comprehension, and 72 students (1% of the total) with inadequate running records. Thus, only 3% of all students tested in spring 2003 were missing a test level because they were not retested, according to test procedures.

Gaps in Benchmark Performance. The percentages of Grade 2 students achieving the end-of-year benchmark are presented in Table 25 according to race/ethnicity and support services received. To analyze patterns of growth over the year, Table 25 also includes the percentages of students who met the fall benchmark in the fall. This benchmark was reading a book at level 16 or higher with a running record of 90% or higher, and with adequate comprehension.

Among all Grade 2 students, the percentage that met the end of the year benchmark by spring was 23% higher than the percentage that met the fall benchmark. This rate of growth from fall to spring was consistent across all racial/ethnic groups and most of the subgroups related to support services received. The rate of growth was lower for IEP students; the difference between this group and students without an IEP was statistically significant ($z=8.05$, $p < .05$).

Table 25. Percentages of Grade 2 Students Achieving Fall and Annual Reading Benchmarks, by Demographic Group

Student group	All students with both fall 2002 & spring 2003 scores	Students who met the fall benchmark in fall 2002		Students who met the end of the year benchmark by spring 2003		Increase in percentage from fall 2002 to spring 2003
	n	n	%	n	%	%
All Students						
	9434	3719	39%	6104	64	25
Race/Ethnicity						
American Indian	31	11	36	18	58	22
Asian American	1435	655	48	993	73	25
African American	1885	598	30	1053	54	24
White	4150	2262	52	3243	74	22
Hispanic	1933	393	21	797	43	22
Special Education						
IEP	792	135	15	256	29	14
None	8642	3784	43	5848	67	24
ESOL Services						
Current ESOL	1031	102	9	329	29	20
Not current ESOL	8403	3617	43	5775	68	25
FARMS Services						
Current FARMS	2208	448	18	1038	42	24
Not current FARMS	7226	3271	46	5066	71	25
Combination						
Non-ESOL & Non-FARMS	6733	3324	51	4853	74	23
ESOL & FARMS	538	42	6	169	25	19

There were also statistically significant differences in the percentages that met the annual benchmark by the end of the year (see Appendix D). Among the racial/ethnic groups, Asian American students and white students had the highest percentages of students who achieved the benchmark, followed by African American students and then Hispanic students. As in Grade 1, poverty and second language learning again affected reading achievement; for FARMS and ESOL, fewer students receiving these services achieved the benchmark than students who did not receive the service. Also, fewer special education students achieved the benchmark, compared with students without an IEP.

Gaps in Reading Levels. The median reading levels for fall and spring are presented in Table 26, according to race/ethnicity and support services received. There were differences in reading

levels at both assessment windows among racial/ethnic groups. At fall 2002 and spring 2003, the reading levels of Asian American and white students were the highest, followed by African American students and then Hispanic students. These differences reflect statistically significant relationships between racial/ethnic groups and the fall reading levels (χ^2=905.3, $p < .0001$) and the spring reading levels (χ^2=708.4, $p < .0001$).

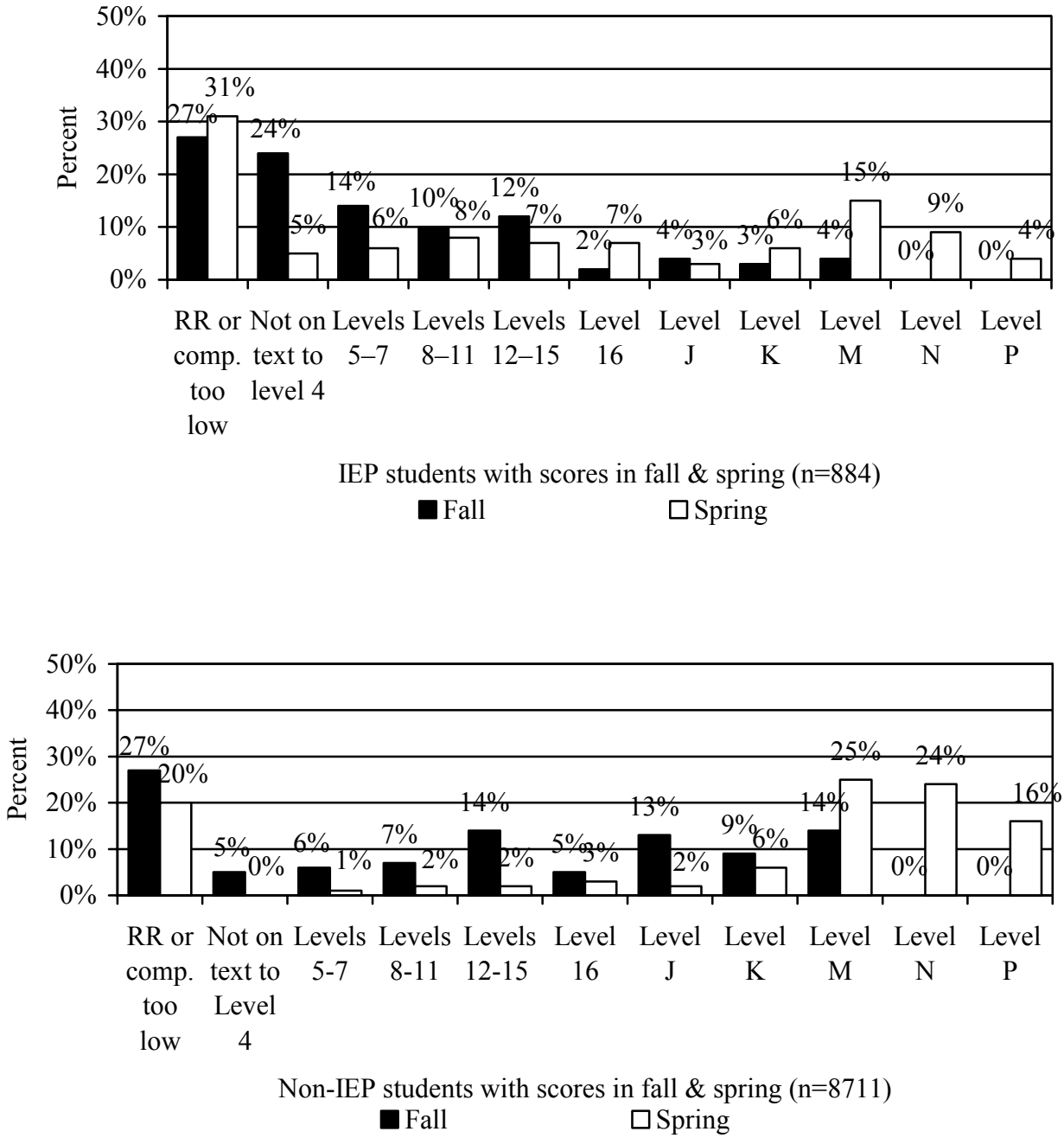
Table 26. Reading Levels of MCPS Grade 2 Students, by Demographic Group

Student group (Students with both fall & spring scores)	Reading level in fall 2002*		Reading level in spring 2003*	
	n	**Median**	**n**	**Median**
All Students				
	7046	16	7566	M
Race/Ethnicity				
American Indian	24	14	27	M
Asian American	980	18	1098	N
African American	1368	14	1443	M
White	3310	18	3653	N
Hispanic	1364	11	1345	M
Special Education				
IEP	643	7	610	K
Non-IEP	6403	16	6956	M
ESOL Services				
ESOL	841	7	808	K
Non-ESOL	6205	J	6758	N
FARMS Services				
FARMS	1868	11	1858	M
Non-FARMS	5178	J	5708	N

*The total number differs among these columns because some students had inadequate comprehension and thus no identifiable reading level in fall or spring.

In fall 2002, 24% of IEP students read at level 4 or below (see Figure 6). This concentration contributed to a median level of 7 for IEP students, compared with a median level of 16 for non-IEP students. In spring 2003, IEP students read at a wide variety of levels.

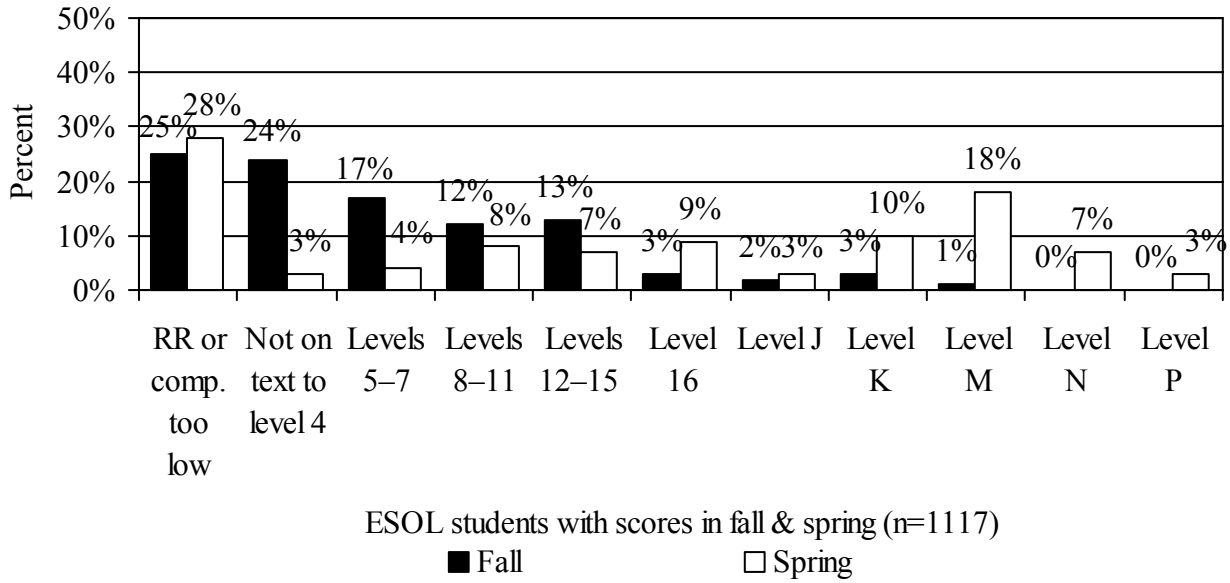
Figure 6. Reading Levels of Grade 2 IEP and Non-IEP Students for Fall 2002 and Spring 2003

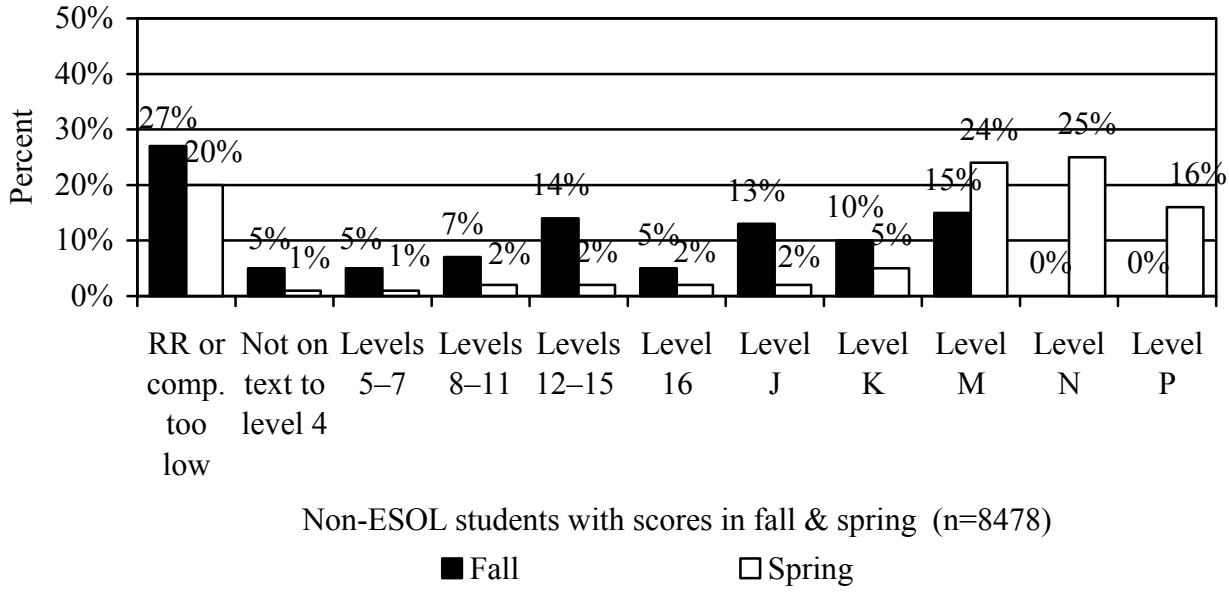


By contrast, the reading levels of non-IEP students in fall 2002 were spread across all the levels. (Levels N and P were not available for testing in the fall.) But in the spring, 65% of the non-IEP students read at level M or above. These differences are reflected in a statistically significant gap in both fall ($z=22.4$, $p < .0001$) and spring reading levels ($z=21.0$, $p < .0001$) between special education students and those without an IEP.

The reading levels of ESOL students in fall 2002 were concentrated in the lower levels; 66% of the students read below level of 16 while the reading levels of non-ESOL students were spread over a wider range (see Figure 7). In spring 2003, the reading levels of ESOL students were more widely dispersed, and non-ESOL students were concentrated such that 65% read at level M or above. Compared to non-ESOL students, ESOL students had lower reading levels in fall ($z=29.6$, $p < .0001$) and spring ($z=26.6$, $p < .0001$).

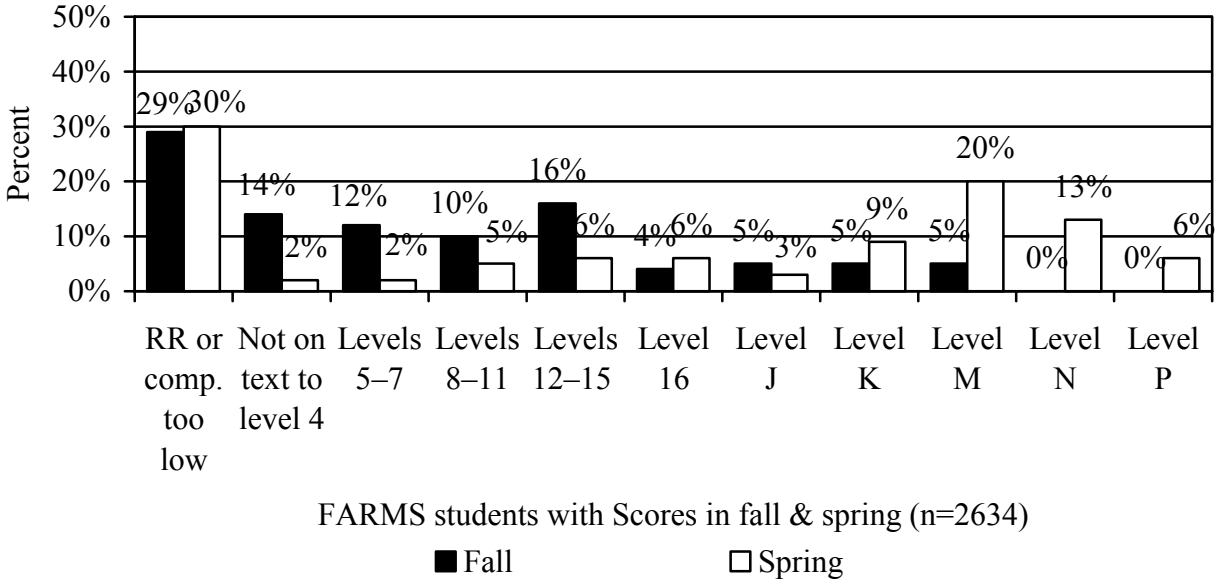
Figure 7. Reading Levels of Grade 2 ESOL and Non-ESOL Students for Fall 2002 and Spring 2003

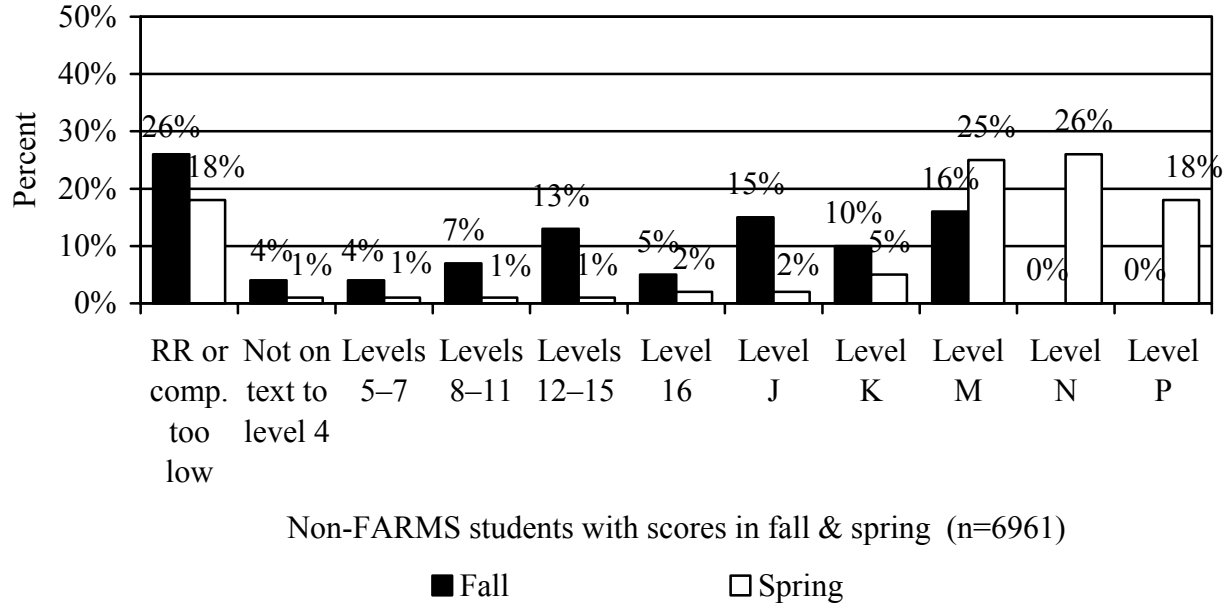




In fall 2002, the range of reading levels for FARMS and non-FARMS students was wide and students were not concentrated at any particular level (see Figure 8). For spring 2003, the reading levels of both groups were concentrated at level M and above; 39% of FARMS students and 69% of non-FARMS students read at these levels. Also, in spring 2003, more FARMS students, than non-FARMS students, had running records or comprehension levels that were too low. Compared to non-FARMS students, FARMS students started the year reading at a lower level ($z=29.5$, $p < .0001$) and ended the year at a lower level ($z=26.7$, $p < .0001$).

Figure 8. Reading Levels of Grade 2 FARMS and Non-FARMS Students for Fall 2002 and Spring 2003





Summary and Recommendations

More than 60% of students in Grades 1 and 2 were able to meet benchmark performance on the MCPSAP-PR. These assessments continue to serve as valuable formative assessments, not only providing teachers with information to guide instruction, but also predicting student performance in subsequent years and on CTBS and MSA.

The MCPSAP-PR has undergone continuous refinement to strengthen all components of the assessment. The addition of oral retelling measures for the 2002–2003 school year provided better information on student understanding at the earliest reading levels. Quarterly benchmarks also were introduced and proved effective for predicting end-of-year performance. Teachers are more comfortable with the assessment tools and use the results to plan differentiated instruction for students.

There is still one area of implementation that needs to be addressed. In spring 2003, the reading level could not be determined for 16% of Grade 1 students and 21% of Grade 2 students. Given the links between MCPSAP-PR benchmark performance and subsequent performance in later grades and on CTBS and MSA, it is important that teachers determine an accurate score for every student.

Students without a reading level fall into two groups. The first group includes students with running records that are too low (i.e., below 90%) or have inadequate oral comprehension. According to test procedures, teachers should retest these students using a lower test level. These students represented 8% of all students in Grade 1 at spring 2003 and 3% of all students in Grade 1 at spring 2003. An emphasis on correct implementation should help correct the problem for this group.

The second group of students without a reading level was students with inadequate written comprehension. These students represented 8% of all students in Grade 1 at spring 2003 and 18% of all students in Grade 1 at spring 2003. Unlike running records and oral comprehension measures, written responses are scored after the student completes the assessment because a second teacher has to score the paper. By the time this scoring is completed, teachers typically do not repeat the assessment so as to retest the student at a lower level.

A change to the administrative procedures for 2003–2004 may provide an opportunity to reduce the number of children with inadequate written comprehension. For 2003–2004, all students who read a text at a level that requires a written response must also answer oral comprehension questions, prior to the written task. It is recommended that teachers retest at a lower level any students who does not meet the target of 80% comprehension and that this point be clarified to teachers. In 2002–2003, the oral comprehension questions were optional at levels J and above. Among students who did answer these questions at level J and above in spring 2003, there were 369 Grade 1 students (4% of all students) and 274 Grade 2 students (3% of all students) who had inadequate scores on the oral comprehension questions but went on to take the written response.

References

MCPS. 2002b. Administrative Guide, MCPS Assessment Program. Primary Reading, 2002–2003 Rockville, MD: Montgomery County Public Schools, Office of Instruction and Program Development.

MCPS 2002b. Elementary Program Instructional Guide, Reading Writing Language Arts, Second Grade, Quarter 1, Appendix. Rockville, MD: Montgomery County Public Schools, Office of Instruction and Program Development.

Raber, Suzanne. Reading Initiative Study, Year 2 Assessment Report. Rockville, MD: Montgomery County Public Schools, Office of Shared Accountability.

MCPS. 2002. Reading Initiative Study, Year 3, 2000-2001 Report: Grade 2 Reading Performance on the Early Childhood Assessment Program, (2002) Rockville, MD: Montgomery County Public Schools, Office of Shared Accountability.

Appendix A: Observation Instruments

PRE-OBSERVATION QUESTIONS, READING

Teacher: _____ Interviewer: _____

School: _____ Grade: _____ Date: _____

1. What lesson/topic are you working on? What are the Essential Questions associated with the lesson I will be observing?

2. What do you anticipate doing in your reading/language arts class on the day I will be observing?

3. What do you hope students will learn as a result of the lesson you have planned?

4. Is there anything in particular that I should know about the group of students that I will be observing?

5. Please make a blank copy for me to keep of any handouts you plan to use on the day I will be observing.

Observation of Reading Instruction in Grade 1 – Grade 2
Winter 2002

School: **Date:**
Teacher: **Grade:** **Observer:**
Observation Start Time: **Observation End Time:**

 # Children Present for majority of time:_____

 # Adults Other Than Teacher Present in Classroom:_____ **Who (position(s)):**

Roles of other adults (describe):

If students leave room for support services, please note how many and for what service.

Use the grid below to record the sequence and time spent in each type of classroom organization and using each instructional strategy. Move to another line to record each time there is a transition in either the organization or the instructional strategies used or for each new small group. Write in time and check boxes as applicable. Write description of "other instructional strategies." You may check more than one category if both are occurring simultaneously within a time period. For example, small group and individual could be applicable if small groups of children are working at centers and the teacher is working one-on-one with a child. Or, another example, guided reading and centers can be occurring at the same time.

Use the reverse of this page to share other comments about this observation that you think are important to note.

Start Time/ End Time	Classroom Organization				Read Aloud /Modeled	Shared Reading	Guided Reading	Centers	Independent Wrk	Writing	Purpose or Focus (if unclear, ask teacher in post-observation)	Description of Activities
	Whole Group	Small Group	Individual	In Transition								

Start Time/ End Time	Classroom Organization											
	Whole Group	Small Group	Individual	In Transition	Read Aloud /Modeled	Shared Reading	Guided Reading	Centers	Independent Wrk	Writing	Purpose or Focus (if unclear, ask teacher in post-observation)	Description of Activities

The number of children in each small group/guided reading group was: _____group 1; _____group 2; _____group 3

For centers, did all children go the same ones? (Circle one) Yes No
If not, (check one) did children choose _____
OR did teacher assign them (verbally or via management system on the wall) _____ .

Observation of Reading Instruction in Grades 1 and 2
Winter 2002

Observer: **School:**
Date: **Teacher:**

Use one column for each whole group activity and for each small group. Please label each column. Please mark how frequently each observation activity occurs, unless noted as Yes/No.

Observation			
Checks for student understanding			
Teacher uses instructional activity (i.e. game) to check for student understanding			
Teacher uses every student response techniques (ESR) to check (e.g., answers on dry erase board)			
Teacher asks questions that check for understanding: and require single word response (e.g. yes/no)			
Teacher asks questions that check for understanding: and require multiple word response			
Teacher asks students to clarify thinking or to justify response			
Teacher repeats instruction when necessary for student understanding			
Teacher elicits questions from students			
Teacher uses exit cards to check for student understanding (i.e. brief response that is collected)			
Differentiates instruction			
Teacher gives alternative assignments, instruction, or activities to accommodate a variety of student abilities (Yes/No)			
Teacher gives alternative instructions for same assignment (Yes/No)			
Assesses			
Teacher takes running records			
Teacher takes anecdotal records (ie takes notes on students)			
Other notes on how did the teacher assess student progress towards the purpose?			

Observation of Reading Instruction in Grades 1 and 2
Winter 2002

Observer: **School:**
Date: **Teacher:**

Use one column for each whole group activity and for each small group. Please label each column.
Please mark how frequently each observation activity occurs, unless noted as Yes/No.

Observation			
Checks for student understanding			
Teacher uses instructional activity (i.e. game) to check for student understanding			
Teacher uses every student response techniques (ESR) to check (e.g., answers on dry erase board)			
Teacher asks questions that check for understanding: and require single word response (e.g. yes/no)			
Teacher asks questions that check for understanding: and require multiple word response			
Teacher asks students to clarify thinking or to justify response			
Teacher repeats instruction when necessary for student understanding			
Teacher elicits questions from students			
Teacher uses exit cards to check for student understanding (i.e. brief response that is collected)			
Differentiates instruction			
Teacher gives alternative assignments, instruction, or activities to accommodate a variety of student abilities (Yes/No)			
Teacher gives alternative instructions for same assignment (Yes/No)			
Assesses			
Teacher takes running records			
Teacher takes anecdotal records (ie takes notes on students)			
Other notes on how did the teacher assess student progress towards the purpose?			

READING POST-OBSERVATION QUESTIONS

Teacher: _____ Interviewer: _____

School: _____ Grade: _____ Date: _____

1. How did you use student data from the Primary Reading Assessments to plan this lesson and/or guide instruction? If not for this lesson, then how have you used it for other lessons?

2. How did you group the students for today's lesson? If small groups were not a part of today's lesson, how did you group students the last time you had small groups?

3. Are there any special strategies that you use for students who are performing below the quarterly benchmark for reading? Please describe.

4. What training have you received this year to adjust instruction based on the data from the reading assessments?
   - Is it adequate? Please explain.

5. What resources and support materials are available to support your use of reading assessment data?
   - Are they adequate? Please explain.

6. What training have you received *this year* to implement and score the reading assessments?

7. Are the Primary Reading Assessments worth the time you take to administer them? Why or why not?

8. Are there any additional comments you'd like to share about how you use the reading assessments?

Appendix B. In-depth Schools Survey

MCPS Assessment Program-Primary Reading
MCPS –AP In-Depth Study Schools Survey

Participants: Please review the following survey and respond to the questions prior to June 11, 2003, meeting at CESC, cafeteria, from 4:00 to 6:30 p.m. BRING your completed survey to the meeting for sharing and discussion.

Circle One:

Kindergarten Grade 1 Grade 2 Other

1. Explain the impact that the assessments have had on your implementation of the R/W/LA instructional guides in Grade 1 and Grade 2. (Kindergarten- Skip this one.)

2. Name the two highest priority reading/language arts topics that are important staff development needs for you and your staff in order to improve your reading/writing/language arts classroom instruction.

   1. _____

   2. _____

1. What has been your experience with using the data from MCPS-AP-Primary reading this year? Use the scale below to rank your experiences. Record a number on line of each suggested experience.

IMS_____
Grade Level Discussion_____
Cross-Grade Level Discussion_____
Parent Conferencing_____
EMT_____
Principal/Teacher Conferencing_____
Differentiating Instruction/Flexible Grouping_____
Other_____

1	2	3	4	5
Useless		Moderately Useful		Highly Useful

2. What percentage of your students met the end of year benchmark? What did you do to keep centered on guiding students' progress? Explain.

May 16, 2003

Appendix C: Phi Coefficients

Table 27. Phi Coefficients for Annual Grade 1 Reading Benchmark With Reading Benchmark at Fall and at Winter and Coefficients for Annual Grade 2 Reading Benchmark With Reading Benchmark at Fall and at Winter, by Demographic Group

Student Group (students with fall 2002 & spring 2003 scores)	Fall benchmark Grade 1	Winter benchmark Grade 1	Fall benchmark Grade 2	Winter benchmark Grade 2
	Phi	Phi	Phi	Phi
All Students				
	.45	.49	.45	.56
Race/Ethnicity				
African American	.43	.48	.39	.52
American Indian	*	*	.49	.41
Asian American	.39	.40	.42	.52
Hispanic	.43	.51	.42	.56
White	.43	.48	.41	.55
Special Education				
IEP	.53	.58	.51	.53
Non-IEP	.44	.47	.42	.54
ESOL Services				
ESOL	.38	.43	.31	.50
Non-ESOL	.43	.47	.42	.54
FARMS Services				
FARMS	.47	.52	.42	.52
Non-FARMS	.42	.46	.40	.55

* Cell size too small to report.

Table 28. Phi Coefficients for Annual Reading Kindergarten Benchmark and Annual Grade 1 Reading Benchmark, by Demographic Group of Grade 1 Students

Student Group (students with fall 2002 & spring 2003 scores)	
	Phi
All Students	
	.43
Race/Ethnicity	
African American	.41
American Indian	.58
Asian American	.36
Hispanic	.41
White	.40
Special Education	
IEP	.50
Non-IEP	.41
ESOL Services	
ESOL	.27
Non-ESOL	.40
FARMS Services	
FARMS	.43
Non-FARMS	.40

Table 29. Phi Coefficients for Annual Grade 2 Reading Benchmark With Annual Kindergarten Reading Benchmark and With Annual Grade 1 Reading Benchmark, by Demographic Group of Grade 2 Students

Student group (students with fall 2002 & spring 2003 scores)	Kindergarten benchmark	Grade 1 benchmark
	Phi	Phi
All Students		
	.36	.49
Race/Ethnicity		
African American	.35	.48
American Indian	.58	.47
Asian American	.31	.41
Hispanic	.34	.45
White	.31	.45
Special Education		
IEP	.36	.52
Non-IEP	.34	.46
ESOL Services		
ESOL	.25	.33
Non-ESOL	.33	.46
FARMS Services		
FARMS	.35	.45
Non-FARMS	.32	.44

Table 30. Phi Coefficients for MSA Benchmark With Annual Grade 1 Reading Benchmark and Annual Grade 2 Reading Benchmark, by Demographic Group of Grade 3 Students

Student group (students with fall 2002 & spring 2003 scores)	Grade 1 benchmark	Grade 2 benchmark
	Phi	**Phi**
All Students		
	.48	.55
Race/Ethnicity		
African American	.46	.54
American Indian	.19	.33
Asian American	.37	.48
Hispanic	.43	.49
White	.41	.46
Special Education		
IEP	.45	.59
Non-IEP	.46	.52
ESOL Services		
ESOL	.38	.34
Non-ESOL	.46	.52
FARMS Services		
FARMS	.42	.49
Non-FARMS	.42	.49

Appendix D. Z Scores

Table 31. Z Scores for Differences Between Groups in the Percentages That Met the Annual Benchmark by the End of the Year

Student group (students with fall 2002 & spring 2003 scores)	Grade 1	Grade 2
Race/Ethnicity		
African American vs. Hispanic	6.7	6.66
Asian American vs. African American	9.1	11.6
Asian American vs. Hispanic	15.5	18.2
Asian American vs. White	2.3	<2 *
White vs. African American	13.7	15.9
White vs. Hispanic	21.9	23.7
Special Education		
IEP vs. Non-IEP	17.4	23.7
ESOL Services		
ESOL vs. Non-ESOL	25.4	26.5
FARMS Services		
FARMS vs. Non-FARMS	22.6	28.5

* Difference not significant.