Evaluation Matters

Volume 22, Number 01 June 2023 Steven M. Urdegar, Ph.D., Director

i-Ready: An Analysis of Usage and Impact *2021-22*

1. What is the purpose of this report?

i-Ready is an English Language Arts (ELA) and mathematics instructional program for Grades K–8 students featuring an individualized plan for instruction based on each student's performance on an online, computer-adaptive diagnostic assessment. Once students complete the diagnostic assessment, i-Ready builds a unique lesson plan with a differentiated starting point for every learner based on their overall and domain-level placement. i-Ready allows teachers to add lessons and/or adjust the lesson sequence provided to individuals or groups of students. Lessons provide explicit instruction and extensive practice, offer supportive feedback, with the goal of building conceptual understanding for learners of all levels. Curriculum Associates recommends that all students using i-Ready maintain an average usage of 30–49 minutes per week and pass 70% of lessons during a school year. The purpose of this study is to separately ascertain for ELA and mathematics the extent to which increases in usage are associated with increases in achievement (Curriculum Associates, n.d.). The i-Ready Diagnostic test is administered three times per year, in the Fall (AP1), Winter (AP2), and Spring (AP3).

2. Which populations were targeted in this report?

The sample for each subject area included all M-DCPS students in Grades K through 8 with valid i-Ready usage and diagnostic data, as well as outcome data during the 2021-22 school year. Students in schools that did not participate in the i-Ready program, or whose grade level on the usage and end-of-year demographic files differed, and those who used i-Ready for less than one minute during the 2021-22 school year or did not complete at least one exercise during that time were excluded. The sample included 155,727 students in ELA and 140,944 students in mathematics.

3. How were the data for this report collected and analyzed?

Design

A predictive correlational design (Tuckman, Bruce W. (1972). Conducting Educational Research, Correlational Study, p. 124.) was used to gauge the impact of the i-Ready program on students' achievement. The i-Ready usage file contained two measures of usage: time and completed tasks. A preliminary analysis examining the partial correlation between each measure of usage and the posttest, controlling for the pretest, indicated that the number of completed tasks was the measure most consistently associated with achievement growth. The students included in the analysis were limited to those students in the sample who had valid pre-and post- test scores at the same grade.

Instrumentation

The results of three different achievement outcome measures were used in this analysis: (a) the Stanford Achievement Test, Tenth Edition (SAT-10); (b) the Florida Standards Assessments (FSA); and (c) the FSA Algebra 1 End of Course (EOC) test. The i-Ready Diagnostic AP1 in either ELA or mathematics served as the pretest at each grade.

The SAT-10 is a standardized norm-referenced test designed to measure students' performance in comparison to a national normative sample. Students' performance is measured in scale scores that are equal units of achievement that vertically align across grades, are amenable to mathematical manipulation, and are specifically designed to compare individuals and groups. The SAT-10 has been administered locally in Reading and Mathematics to all students in Grades K-2 during the spring of each school year, through Spring 2022.

The FSA, administered in English Language Arts (ELA) and Mathematics, is the primary accountability measure used by the state of Florida. It is a criterion referenced test designed to measure students' mastery of the state's Florida Standards. Student performance is measured in scaled scores and reported in achievement levels that range from 1 (low) to 5 (high). The FSA has been administered statewide in ELA to Grades 3 through 10 and in mathematics to Grades 3-8 as the most recent statewide standardized achievement test, during the spring of each school year, through Spring 2022. An achievement level of 3 or higher constitutes the statewide standard for proficiency. The FSA Algebra 1 and Geometry EOC exams are computer-based subject area tests that measure students' mastery of the Florida Standards in Algebra 1 and Geometry, respectively or equivalent courses. Results are measured in scaled scores and reported in achievement levels that range from 1 (low) to 5 (high). An achievement level of 3 or higher constitutes the statewide standard for proficiency.

The i-Ready Diagnostic in ELA and mathematics is a standardized norm-referenced test designed to measure students' performance in comparison to a national normative sample. Students' performance is measured in scale scores that are equal units of achievement that vertically align across grades, are amenable to mathematical manipulation, and are specifically designed to compare individuals and groups. In addition, i-Ready Diagnostic results are presented by achievement levels, with red, yellow, and green bands representing low – high performance overall and by strand. The i-Ready Diagnostic is administered locally to all students in Grades K-8 during the fall, winter, and spring of each school year.

Data Analysis

Usage, in terms of the number of completed tasks, was sorted within grade and classified in five bands called quintiles, based on percentile: 1 (00.01 to 20.00), 2 (20.01 - 40.00), 3 (40.01 – 60.00), 4 (60.01 – 80.00) and 5 (80.01 - 100.00). The quintiles were defined to provide for inferential comparisons between targeted percentiles of usage located at the midpoint of each band within the distribution, i.e., Quintile 1 (10^{th}), Quintile 2 (30^{th}), Quintile 3 (50^{th}), Quintile 4 (70^{th}), and Quintile 5 (90^{th}).

Separate regression analyses for ELA and mathematics were conducted at each grade, and used to apply the predictive correlational design and compare the difference in the posttest scores for students at each of the four usage levels (i.e., quintiles, 2,3,4, and 5) with the posttest scores for students at the lowest usage level (i.e., quintile 1), controlling for the influence of the pretest and demographic characteristics (i.e., gender, ethnicity, free/reduced lunch eligibility, English language learner status, special education status, and age relative to grade). Average time per completed task and its interaction with pretest was also included as a control variable to improve the accuracy of prediction.

4. What was the impact of completing i-Ready tasks on subsequent FSA results?

English Language Arts

Usage of the i-Ready platform for instruction in English Language Arts, in terms of the number of completed tasks, was sorted within grade and classified in five bands called quintiles, based on percentile: 1 (1 to 20.00), 2 (20.01 -40.00), 3 (40.01 – 60.00), 4 (60.01 – 80.00) and 5 (80.01 to 100.00). These bands

were defined to provide for inferential comparisons between targeted percentiles of usage located at the midpoint of each band within the distribution, i.e., Quintile 1 (10th), Quintile 2 (30th), Quintile 3 (50th), Quintile 4 (70th), and Quintile 5 (90th). Table 1 lists, for each grade, the minimum, mean, and maximum number of completed tasks within each quintile.

Table 1. Distribution of Completed English Language Arts Tasks within each Quintile by Grade

	Quintile 1			Quintile 2			Quintile 3			Quintile 4			Quintile 5		
Grade	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max
K	1	**20**	39	40	**53**	66	67	**77**	88	89	**98**	110	111	**135**	365
1	1	**23**	42	42	**56**	67	67	**77**	87	87	**97**	109	109	**139**	581
2	1	**29**	46	46	**57**	67	67	**76**	85	85	**96**	108	108	**137**	388
3	1	**32**	47	47	**58**	68	68	**77**	87	87	**98**	112	112	**149**	561
4	1	**26**	39	39	**49**	58	58	**67**	76	76	**86**	97	97	**128**	409
5	1	**23**	36	36	**46**	54	54	**62**	71	71	**80**	92	92	**122**	435
6	1	**9**	16	16	**23**	29	29	**36**	44	44	**53**	65	65	**90**	367
7	1	**6**	11	11	**17**	23	23	**29**	36	36	**44**	54	54	**78**	325
8	1	**5**	10	10	**15**	20	20	**25**	30	30	**36**	44	44	**66**	319

The typical student in Grades K to 3 completed between 67 and 87 exercises for an average of about 77. Typical practice declined in Grades 4 and 5, ranging from around 55 to 75 for an average of 64 before declining further in Grades 6 to 8. It may be expected that fewer tasks would be completed as the length/complexity of a task increased across the grade levels. The average time per task was around 20 minutes in Grades K-2; around 30 minutes in Grades 3-5, and 52-62 minutes in Grades 7 and 8, supporting that expectation. However, the overall usage time in ELA for the year varied, and not always around the variation in the time required to complete individual tasks. Students in Grade 8 completed the fewest tasks and had the lowest average overall usage time for the year, 967 minutes, followed by students in Grade 7 (average usage time 1064 minutes); students in Grade 6 also completed relatively few assignments, and had a relatively low overall usage time (1220 minutes/student). In Grades K-5, the overall usage time was more consistent, with 1252 minutes/student in Kindergarten; between 1454 and 1645 minutes/student in Grades 1, 2, 4, and 5; and a peak of 1843 minutes/student in Grade 3.

A dose response analysis was used to assess the impact of the number of i-Ready instructional tasks completed on the students' achievement posttest scores. The difference between student outcomes at quintiles 2 through 5 were compared with student outcomes at the lowest level of usage (i.e., quintile 1, as a reference group). A positive dose response would result if increased usage were predictive of increased achievement when initial ability and demographic characteristics were taken into account. Interactions between each of the quintiles (i.e., quintiles 2, 3, 4, and 5) and the pretest were defined to account for the possibility that the effect of usage varied with students' initial ability. All predictors except pretest and time per task are dichotomous. The two continuous predictors are grand-mean-centered (i.e., expressed as deviations from the sample mean), such that the intercept of the regression equation gives the value assumed by the posttest when all dichotomous predictors are zero and all continuous predictors are their sample mean values. The unstandardized B coefficient for each quintile gives the impact of a one-point change in that predictor on the posttest when both the predictor and the posttest are in original

units. As such, dichotomously coded predictor weights give the difference in the posttest between the listed group coded "1" and the reference group coded "0." Table 2 provides descriptive statistics of the pre- and post-test for each grade to facilitate the analytical interpretation that follows.

The number of students with complete data ranged from a low of 13,886 in Grade K to a high of 18,116 in Grade 5 for an average of 15,991. As the outcomes in Grades K-2 and 3-8 are aligned to a vertical scale that progresses throughout the grades, by comparing the mean at the center of each grade it is possible estimate a year of growth, which can then be expressed in terms of the standard deviation. In Grades K-2 on the SAT-10, one year of growth is 1 standard deviation; while in Grades 3 – 8 on the FSA, one year of growth is approximately one third of a standard deviation. The standard deviations shown are sample standard deviations, which only approximate those values; this metric is not applicable between Grades 2 and 3, with the transition from SAT-10 to FSA.

Table 2. Distribution of the English Language Arts Pre- and Post-test by Grade

		Pretest				Posttest			
Grade	N	M	S.D.	Min.	Max.	M	S.D.	Min.	Max.
K	13,886	344.8	35.6	139	574	497.6	59.3	318	596
1	14,110	401.9	46.2	229	582	549.8	61.3	351	661
2	15,451	454.0	58.0	247	648	595.1	50.6	419	721
3	17,004	493.1	59.5	192	660	299.7	23.4	240	360
4	16,075	532.8	58.5	284	696	314.9	24.3	251	372
5	18,116	552.2	60.0	283	701	323.5	25.5	257	385
6	15,856	569.6	60.9	299	707	324.7	25.6	259	391
7	16,636	582.0	62.1	143	800	330.7	25.3	267	397
8	16,782	593.7	63.7	243	745	335.2	26.4	274	403

Table 3 presents the results of the regression analyses in terms of the main-effects regression coefficients for each quintile and any significant quintile x pretest (scaled to a one standard deviation change) interaction, followed by the R^2 value which gives the proportion of variance in the posttest explained by the model. Statistically significant components are shown in bold font.

Table 3 shows that only Grades 2 and 4 are free of interactions. In those grades the effects of each quintile of usage (compared to Quintile 1) are unaffected by students' initial ability, so that the significant main effect of usage is consistent across all ability levels. For example, in Grade 2, by applying the completion statistics in Table 1 to the coefficients in Table 3, an increase in average completed tasks from 29 (Quintile 1) to 57 (Quintile 2) predicts a non-significant increase of 0.31 scale score points on the posttest (SAT-10). An additional increase to an average of 77 tasks completed by students in Quintile 3 predicts a significant increase of 2.76 scale score points on the posttest. A similar increase of 2.73 points is seen for students in Quintile 4, who completed an average of 96 tasks, so that while students still benefited (over those in Quintile 1), the benefit was not any higher than that seen for students in Quintile 2, who completed only 76 tasks, so the increase from 76 to 96 tasks did not yield any additional benefit. However, students in

Quintile 5, who completed an average of 137 tasks did yield a significant additional benefit, with an increase of 5.05 scale score points.

Table 3 Results of the English Language Arts Dose Response Regression Analysis

	Quintile 2		Quintile 3		Quintile 4		Quintile 5		
Grade	Effect	Interaction	Effect	Interaction	Effect	Interaction	Effect	Interaction	R^2
K	1.93		**7.15**	**4.00**	**14.95**	**1.70**	**24.03**		.35
1	0.77		1.69	**2.01**	**5.80**	**2.15**	**9.77**		.57
2	0.31		**2.76**		**2.73**		**5.05**		.67
3	0.25	**1.71**	0.23	**2.60**	0.67	**2.49**	0.62	**2.86**	.70
4	**0.87**		**1.54**		**2.50**		**4.08**		.74
5	**1.85**	**0.75**	**2.87**	**1.80**	**3.05**	**2.40**	**3.21**	**2.07**	.75
6	**0.82**		**2.59**	**0.95**	**2.83**	**0.87**	**3.21**	**1.47**	.73
7	**0.95**		**2.05**		**2.73**	**0.55**	**2.55**		.72
8	-0.55	**0.81**	0.32	**1.17**	**0.89**	**2.24**	-0.30		.73

Note: Significant effects and interactions ($p < .05$) are shown in boldface type. R^2 gives the percentage in posttest variance explained by the model. An R^2 value of .50 or greater denotes a suitable model fit.

In Grade 4, benefits are accrued at each level of usage relative to the Quintile 1, where students completed an average of 26 tasks. For students in Quintile 2, who completed an average of 49 tasks, the effect was an increase of 0.87 scaled score points on the FSA; students in Quintile 3, who completed an average of 67 tasks with an increase of 1.54 points, Quintile 4, with an average of 87 tasks and an increase of 2.50, and Quintile 5, completing an average of 128 tasks for an increase of 4.08 points.

When significant interaction effects are present, it means that even when significant findings are present for usage, the impact differs for students of different abilities; this is true in Grades K, 1, 3, 5, 6, 7, and 8. In all cases, the significant interaction effects were positive , which indicates that students with higher initial ability levels (as measured by their pretest scores) benefited more from increased task completion, while students with lower pretest scores benefited less from task completion. The larger the magnitude of the interaction effect, the less consistent the benefit, in some cases negating any main effect benefit of dosage.

For example, in Grade 3, the main effects of usage are all non-significant across quintiles, while the interaction effects are significant. As such, for students at the middle of the distribution of usage – in Quintile 3, who completed an average of 77 tasks, students who scored one standard deviation above the mean on the pretest increased by 2.6 points on the FSA, while those who scored one standard deviation below the mean on the pretest decreased by an equivalent amount, 2.6 points. This discrepancy increases for students farther from the mean, two or more standard deviations from the mean pretest score.

Returning to Table 3, one sees that in Grade 1, the interaction effect largely erases the benefits for below average students at Quintile 3 of usage. At Quintile 4, overall students increased by 5.80 points, but parsed out by initial ability level, the benefit was only 3.65 points for the lowest performing students (16^{th} percentile on the pretest), and 8.05 points for the highest performing students (84^{th} percentile on the pretest). It may be noted that for the students in Quintile 5, who completed an average of 139 tasks, an increase of 9.77 points was seen, without an interaction effect indicating that students of all ability levels benefitted from this level of usage.

In Grades 6 and 7, the interactions at Quintiles 3, 4 and 5 are small enough to maintain a benefit for all students. In Grade 8, the results are less consistent, with insignificant and/or small main effects. Finally, the interactions in Grade K are also small enough to maintain a benefit for all students. However, it should be noted that the R^2 for that grade at .35 indicates that the model does not provide a good fit to the data.

In sum, i-Ready appears to provide a significant benefit in English Language Arts to students who use the i-Ready platform in Grades 2, 4, 6, and 7, regardless of ability. In Grades 1, 3, and 5 the benefit varies with students' ability, such that not all students benefit to the same degree. While the benefit in Grade K is substantial, poor model fit may affect the reliability of the findings.

Mathematics

Table 4 lists, for each grade, the minimum, mean, and maximum number of mathematics tasks completed on the i-Ready platform within each quintile.

Table 4. Distribution of Completed Mathematics Tasks within each Quintile by Grade

	Quintile 1			Quintile 2			Quintile 3			Quintile 4			Quintile 5		
Grade	Min.	M	Max.	Min.	M	Max.	Min.	M	Max.	Min.	M	Max.	Min.	M	Max.
K	1	14	26	27	36	47	48	57	67	68	76	86	87	107	246
1	1	20	36	37	47	57	58	66	74	75	83	93	94	116	307
2	1	25	38	39	47	55	56	62	69	70	78	87	88	112	342
3	1	23	35	36	43	51	52	59	67	68	76	87	88	117	398
4	1	21	32	33	40	47	48	54	62	63	71	81	82	109	362
5	1	15	24	25	31	38	39	44	51	52	59	68	69	92	362
6	1	7	12	13	17	23	24	29	35	36	43	52	53	73	331
7	1	4	9	10	13	17	18	23	29	30	36	44	45	64	243
8	1	3	5	6	9	13	14	18	22	23	29	36	37	56	450
8[a]	1	2	4	5	8	12	13	18	22	23	28	35	36	46	105

[a]Algebra 1 End of Course Exam.

The typical student in Grades K to 4 completed between 48 and 74 exercises for an average of about 63. Typical practice declined in Grades 5 and 6, ranging from around 24 to 51 for an average of 35 before declining further in Grades 7 to 8. It should be noted that the time per task in mathematics increased incrementally across grade levels, from about 20 minutes per task in K-2, about 30 minutes in Grades 4 and 5, and closer to 60 minutes per task in Grades 6 – 8, supporting the expectation that tasks increased in length/complexity across grade levels. Overall usage time for the year was more consistent in mathematics, ranging from a low of 1131 minutes/student in Kindergarten to a high of 1646 minutes in Grade 4.

Table 5 provides descriptive statistics of the pre- and post-test for each grade to facilitate the analytical interpretation that follows. The number of students with complete data ranged from a low of 2,177 in Grade 8 (Algebra 1) to a high of 18,058 in Grade 5 (FSA) for an average of 14,992. The outcomes within Grades K-2 and within 3-8 are aligned to vertical scales that progress across grades. By comparing the mean at the center of each grade it is possible estimate a year of growth which can then be standardized in terms of the standard deviation. In Grades K-2 one year of growth is 1 standard deviation, while in Grades 3 – 8 one year of growth is a third of a standard deviation. The standard deviations shown are

sample standard deviations which approximate those values. Table 6 present the results of the regression analyses in terms of the regression coefficients for each quintile and any significant quintile x pretest (scaled to a one standard deviation change) interaction, followed by the R^2 value which gives the proportion of variance in the posttest explained by the model.

Table 5. Distribution of the Mathematics Pre- and Post-test by Grade

		Pretest				Posttest			
Grade	N	M	S. D	Min.	Max.	M	S.D.	Min.	Max.
K	13,763	339.1	24.7	188	486	501.7	52.1	283	607
1	14,001	374.7	26.8	234	492	548.9	50.0	364	659
2	15,333	400.0	28.3	268	546	584.5	50.1	400	699
3	16,868	425.0	27.4	289	560	300.2	23.3	240	360
4	16,034	447.6	27.5	293	548	316.5	24.8	251	376
5	18,058	463.3	29.5	297	573	320.0	26.7	256	388
6	15,702	476.6	31.6	292	610	321.5	26.7	260	390
7	13,956	481.6	32.7	297	616	326.1	23.2	269	391
8	9,040	476.9	34.0	299	615	324.5	23.5	273	393
8[a]	2,177	519.9	24.1	334	613	514.5	22.0	425	575

[a]Algebra 1 End of Course Exam.

Table 6 Results of the Mathematics Dose Response Regression Analysis

	Quintile 2		Quintile 3		Quintile 4		Quintile 5		
Grade	Effect	Interaction	Effect	Interaction	Effect	Interaction	Effect	Interaction	R^2
K	**5.62**		**11.16**		**14.70**	**-4.27**	**21.85**	**-7.92**	.41
1	**3.01**		**5.50**		**8.96**		**12.79**		.49
2	**3.13**		**5.85**		**7.22**		**9.11**		.57
3	**1.94**		**4.11**		**6.39**		**9.38**		.66
4	**2.41**		**4.92**		**6.82**		**9.02**		.70
5	**4.24**		**6.42**		**8.58**	**0.80**	**11.01**		.71
6	**1.66**		**3.46**		**5.35**		**7.20**	**-0.75**	.72
7	0.67		**1.41**		**3.05**	**0.70**	**4.83**		.65
8	**1.90**		**3.10**		**5.31**		**8.59**		.44
8[a]	**-2.37**	**-2.84**	0.25		0.80		2.21		.40

[a]Algebra 1 End of Course Exam.

Table 6 shows Grades 1 – 4, 6, and 8 (FSA) to be free of interactions. In those grades, the effects of each quintile of usage with respect to Quintile 1 are unaffected by students' initial ability as measured by the pretest. For example, by applying the completion statistics in Table 4 to the coefficients in Table 6, in Grade 2, an increase in usage of 22 tasks from 25 (Quintile 1) to 47 (Quintile 2) predicts a significant

increase of 3.13 scale score points on the posttest. An additional 15 task increase to 62 predicts a significant 5.85 increase of scale score points on the posttest relative to the Quintile 1 mean of 25. A further 16 task increase to 78 produces a 7.22 scale score increase and an additional 41-point increase predicts an increase of 9.11 scale score points. In Grade 2, one year (180 school- days or 36 school-weeks) of growth is one standard deviation (50.1 scale score points, see Table 5). As such, each scale score increase is equivalent to 0.72 weeks of growth (multiplier). The significant interaction effects found in Grades K, 5, and 7, and 8 indicate that the effect at each quintile varies with students' pretest scores. That said, the interactions found have little effect throughout the range of the pretest. In sum, i-Ready appears to provide a significant benefit to student users in Grades K-8 (FSA), regardless of ability. In Grades 8 (Algebra 1 EOC), the effect was non-significant except for the highest usage level where a small benefit was found.

5. What are the primary conclusions of this report?

In English Language Arts, i-Ready appears to provide a significant benefit to student users in Grades 2,4, 6, and 7, regardless of ability. In Grades 1, 3, and 5 the benefit varies with students' ability such that not all students benefit to the same degree. While the benefit in Grade K is substantial, poor model fit calls the reliability of the results into question.

In mathematics, i-Ready appears to provide a significant benefit to student users in Grades K-8 (FSA), regardless of ability. In Grade 8 (Algebra 1 EOC), the effect was non-significant except for the highest usage level where a small benefit was found.

Thus, i-Ready appears to provide a significant benefit to all students who complete a minimal number of exercises in most grades and subjects: ELA (Quintiles 2 through 5 - Grades K and 2, 4 through 7; and Quintile 3 through 5 - Grade 1) and mathematics (Quintiles 2 through 5 - Grades K through 8). The benefit was not seen in ELA, Grades 3 and 8; or in mathematics for students taking the Algebra 1 EOC. Significant interactions of usage and pretest, mostly in ELA, indicate that the programs' benefit varies with students' ability. Nevertheless, in nearly all cases the benefit obtained from additional time persisted despite interactions. Since the primary accountability instrument was changed in the fall of the 2022-23 school year to the Florida Assessment of Student Thinking, this analysis should be repeated to ensure to control for the effects of instrument variation.