Does Adoption of ACT Aspire Periodic Assessments Support Student Growth?

Jeff Allen, PhD

ABOUT THE AUTHOR

Jeff Allen, PhD

Jeff Allen is a statistician and director in Validity and Efficacy Research. He specializes in longitudinal research of educational outcomes, student growth models, and validation of college readiness measures.

SUMMARY

This study examines the effects of adoption of ACT® Aspire® Periodic Assessments on student academic growth, as measured by the ACT Aspire Summative Assessments. A difference-in-difference analysis shows that adoption of the ACT Aspire Interim Assessments leads to improvements in academic growth. Averaging results across all subject areas and grade levels, adoption typically led to an increase in student growth of 1.3 student growth percentile units, which is comparable to moving from the 50th percentile of school growth to the 56th percentile of school growth. The effect of adoption was strongest for English (+2.9 growth percentile units), followed by science (+1.4 growth percentile units) and math (+1.1 growth percentile units). For reading, effects of adoption were inconsistent across grade levels.

SO WHAT?

Generally, positive effects of adoption were larger for lower grade levels. There was also evidence of Periodic Assessment dosage effects, as student growth increased with more Interim and Classroom Assessments taken. Higher performance on the Interim Assessments was predictive of higher academic growth, as measured by the Summative Assessments.

NOW WHAT?

While the study showed positive effects of ACT Aspire Periodic Assessment, it did not address how the assessment data was used, or how variation in assessment use related to differences in improvement. One idea for additional research would be to survey schools that have used Aspire's Periodic Assessments to understand variation in how the assessment data are used and whether different usage types are related to student growth.

Table of Contents

Does Adoption of ACT Aspire Periodic Assessments Support Student Growth?

Jeff Allen, PhD

Introduction

The ACT® Aspire® Periodic Assessments include Interim and Classroom Assessments (ACT, 2018). The Periodic Assessments can be taken at any time during the academic year, and there are four Interim test forms and 10 Classroom test forms for each subject area (English, math, reading, and science) and grade level (grades 3-10 for Interim, grades 3-8 for Classroom). The Interim and Classroom Assessments are fixed-format, computer-based, and multiple choice. The Interim tests are untimed, and teachers typically allow 45 minutes or less,[1] while the Classroom tests take 10 to 15 minutes. The Interim tests can be thought of as abbreviated versions of the Summative tests, covering the same knowledge and skills and using the same reporting categories as the Summative test. Within grade level and subject area, the content of the Interim tests is not sequenced. Thus, any of the four test forms can be administered at any point during the academic year. Conversely, each Classroom test is mapped to one or two content standards, and teachers can administer the tests in conjunction with lessons or instructional units. Both types of assessments offer immediate reporting. Interim provides reports for students or parents, teachers (or other user-defined groups of students), schools, and districts. Classroom provides reports for students (or parents) and teachers (or other user-defined groups). Reports for both assessments include item response analysis.

In general, interim assessments are used to (a) generate data to inform instruction, (b) gauge how well students are progressing towards meeting academic standards, (c) help students prepare for summative assessments, and (d) evaluate educational programs (Burch, 2010; Li, Marion, Perie, & Gong, 2010). Classroom assessment generally refers to assessment practices that are intertwined with instruction, designed to allow students to demonstrate their learning with a clear purpose of supporting teaching and learning. Typically, classroom assessment occurs in short cycles coinciding with learning objectives.

The stated purpose of ACT Aspire's Periodic Assessments is "to help students prepare for the ACT Aspire Summative assessment" (ACT, 2018, p. 1.2). One way to determine if it is fulfilling this purpose is to examine the effect of using the Aspire Periodic Assessments on academic growth, as measured by the ACT Aspire Summative Assessment. Use of the Aspire Periodic Assessments could lead to improved student growth due to (a) feedback to teachers and instructional coaches on curricular areas that should be strengthened or retaught, (b) individualized diagnosis of knowledge, skills, and abilities (KSAs) in need of improvement, (c) greater recognition of the KSAs tested by Aspire Summative, and (d) practice with items measuring the KSAs tested by Aspire Summative (e.g., summative test prep).

This study examines effects of adopting ACT Aspire Periodic Assessments on student academic growth. Specifically, the study examined the effects of schoolwide adoption of Aspire Interim Assessments on student growth measured by the Aspire Summative Assessments. The study provides initial evidence of how use of Aspire's Periodic Assessments can lead to improvements in academic growth.

Methods

Study Design

Quasi-experimental research designs are possible when outcome data are available for the time period before and after an intervention (or a treatment, or a policy change) for groups that receive the intervention, as well as for groups that do not receive the intervention. One such design is referred to as the *untreated control group design with pretest and posttest*, which can be examined using a *difference in difference* (DiD, Meyer, 1995) analysis. The DiD analysis can be used to estimate the causal effect of the intervention.

In this study, the effects of Aspire Interim adoption are examined using a DiD analysis. All schools included in the study administered the Aspire Summative Assessments in at least three consecutive academic years (e.g., Year 1, Year 2, and Year 3), providing two years of yearly student growth measures (e.g., Year 1 to Year 2, Year 2 to Year 3). Adoption schools used Aspire Interim Assessments during the second growth period (Year 2 to Year 3) but not during the first growth period (Year 1 to Year 2). Comparison schools did not use the Aspire Interim Assessments during either growth period. Improvement in academic growth can be measured for both adoption schools and comparison schools by comparing student growth for the two growth periods (e.g., improvement in growth = average growth percentile for Year 2 to Year 3 – average growth percentile for Year 1 to Year 2). The DiD is calculated as the difference in improvement for adoption schools versus comparison schools and estimates the effect of Aspire Interim adoption. Figure 1 illustrates the DiD approach for hypothetical data. In this example, the adoption schools had an improvement of 2.1 in average student growth percentile from pre-adoption to post-adoption. During the same period, the comparison schools had an improvement of 0.8 in average student growth percentile. Therefore, the DiD estimate is 1.3 (2.1-0.8) and represents the estimated effect of adopting the Interim Assessments.

Figure 1. Hypothetical Difference-in-Difference Analysis
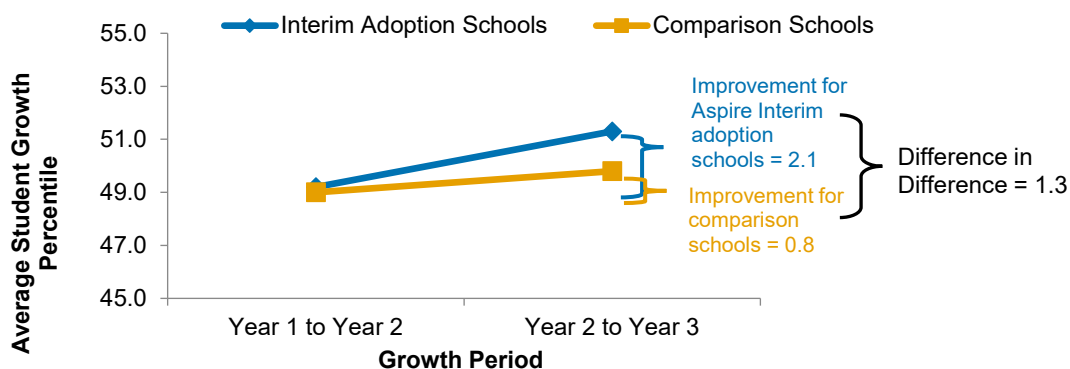
Sample

Inclusion criteria were assessed for each combination of school, grade level pair (3-4 through 9-10),[2] and subject area (English, math, reading, and science). We refer to each combination as a study unit. Inclusion criteria for each unit included:

- The Aspire Summative Assessments must have been administered in the spring in at least three consecutive years to the majority (≥ 50%) of the student body. This allows us to measure student growth for two consecutive cohorts, and thus measure *improvement* in student growth. The unit must have been matched to one of two school databases.[3] For public schools, data from the National Center for Education Statistics Common Core of Data (Glander, 2016) was used to obtain enrollment count for each grade level and proportion of students eligible for free or reduced lunch. For non-public schools, data from Market Data Retrieval[4] was used to estimate enrollment count for each grade level.

- The unit must be eligible for assignment of treatment status (e.g., classification as an adoption school or a comparison school). Adoption units must have administered Interim Assessments to no more than 5% of the student's assessed with Aspire Summative during the first growth period (e.g., Year 1 to Year 2) and then administered Interim Assessments to at least 75% of the student's assessed with Aspire Summative in the next growth period (e.g., Year 2 to Year 3). Comparison units must have administered Interim Assessments to less than 5% of the students assessed with Aspire Summative during both growth periods. Comparison units must not have administered Classroom Assessments (0% among students assessed with Aspire Summative) during either growth period. Units could be classified as adoption or comparison units for more than two growth periods. In that case, the earlier growth periods are used for analysis. For example, suppose a unit administered Aspire Summative for growth periods 2013-2014, 2014-2015, 2015-2016, and 2016-2017 and administered the Interim Assessments for the 2014-2015 and 2016-2017 growth periods. In this case, the first two growth periods (2013-2014 and 2014-2015) would be included in the analysis, with 2014-2015 considered the growth period of adoption.

- The unit must be located in a state that includes both adoption and comparison units.

Table 1 documents the number of schools and students that met the inclusion criteria. Overall, the study included 1,477 schools from nine states, with most schools located in Alabama, Arkansas, and Wisconsin. The samples of adoption schools are largest for math and reading. The samples are relatively small for grade 8-9, likely due to state summative testing requirements changing from grade 8 to 9.[5]

Table 1. Number of Schools and Students Included in the Analysis

Subject	Grade level pair	N schools		N students	
		Adoption	Comparison	Adoption	Comparison
English	3-4	48	230	6,031	31,845
	4-5	41	225	5,055	31,050
	5-6	44	154	7,256	25,744
	6-7	34	146	7,092	26,526
	7-8	40	159	7,793	31,513
	8-9	17	60	7,147	17,808
	9-10	44	135	13,171	38,107
Math	3-4	186	458	28,400	63,241
	4-5	181	428	27,991	60,717
	5-6	135	337	27,365	59,809
	6-7	122	300	27,919	62,390
	7-8	128	310	29,318	65,326
	8-9	16	36	6,604	8,973
	9-10	44	116	14,044	29,634
Reading	3-4	196	455	29,602	62,783
	4-5	183	427	29,412	61,821
	5-6	144	350	28,881	61,955
	6-7	124	302	29,303	63,460
	7-8	127	313	29,545	66,165
	8-9	18	50	9,287	10,700
	9-10	49	131	17,519	31,492
Science	3-4	75	239	10,351	33,277
	4-5	73	285	10,304	37,959
	5-6	62	209	10,066	33,194
	6-7	51	188	9,778	35,478
	7-8	58	194	12,784	38,111
	8-9	19	53	7,825	13,372
	9-10	42	123	12,867	33,198

Background variables (demographics and prior achievement) of the study samples are summarized in Table 2. Note that the numbers in Table 2 represent the average of the results across the 28 conditions (grade level and subject area combinations). The adoption and comparison groups differed on average unit sample size (171.5 for the comparison group, 199.0 for the adoption group), percent eligible for free or reduced lunch (48% for comparison group and 52% for the adoption group), and public school affiliation (93% for comparison group, 98% for adoption group). The adoption group also had a larger percentage of Black students and a larger percentage of students from Arkansas. The groups also differ on the academic years (growth periods) included in the analysis. Because the adoption and comparison groups vary on some background variables, the two groups do not have baseline equivalence. As described later, sample weighting and regression models with covariate adjustment are used to make the two groups more similar.

Table 2. Comparison of Adoption and Comparison Samples

Variable	Sample		Weighted Sample	
	Comparison	Adoption	Comparison	Adoption
Growth period (%)				
2014-2015	32%	20%	29%	29%
2015-2016	42%	36%	40%	40%
2016-2017	18%	30%	21%	22%
2017-2018	8%	14%	10%	10%
Prior year ACT Aspire score (mean)	419.5	419.0	419.4	419.1
Unit N (mean)	171.5	199.0	176.1	168.5
Unit % tested (%)	91%	91%	91%	91%
School FRL (%)	48%	52%	49%	50%
Public school (%)	93%	98%	95%	95%
Race/ethnicity (%)				
Asian	2%	2%	2%	2%
Black	26%	29%	27%	28%
Hispanic	8%	9%	8%	7%
Missing	3%	1%	3%	3%
Other	2%	2%	2%	2%
White	60%	58%	59%	59%
State (%)				
AL	74%	66%	72%	72%
AR	14%	25%	17%	18%
WI	5%	7%	6%	6%
Other	7%	2%	5%	4%

Measures

ACT Aspire reports student growth percentiles (SGPs) for students who test in consecutive years with the Summative Assessments. ACT Aspire SGPs represent the percentile rank of a student's current year score, among all students with the same prior year score. The SGPs can be averaged to form a summary measure of student growth. For example, a school with a mean SGP of 50 demonstrated average growth, relative to schools and students included in the SGP norm group. The ACT Aspire SGP tables used for this study are the 2018 version of the ACT Aspire SGPs for grades 3-10.[6] SGPs are used as the measure of student growth in academic achievement, the outcome for the study analysis.

For each student within each condition, the number of Interim Assessments taken was categorized as 0, 1, 2, 3, or 4 or more. Number of Classroom Assessments taken was categorized as 0, 1-2, 3-4, 5-6, or 7 or more. While there are four Interim test forms and 10 Classroom test forms for each condition, students could test more than four (10) times by retaking the same test form.

When students take an Aspire Interim test, a predicted end-of-year Summative test score is produced. When multiple Interim tests are taken, the prediction utilizes the last three Interim test scores using multiple linear regression (ACT, 2018, p. 12.2). Predicted end-of-year Summative score was used as a summary measure of performance on the Interim Assessments.

Propensity Score Weighting

Adoption and comparison schools have underlying differences in background variables (see Table 2) that could impact the DiD analysis. A propensity score weighting approach (Austin, 2011) was used to ensure that the adoption and comparison schools were similar on several covariates, including number of students tested, proportion of students tested, school affiliation (public or non-public), school proportion eligible for free or reduced lunch, average prior year ACT Aspire Summative score, state, student race/ethnicity, and year of growth period. A logistic regression model was used to predict group membership (adoption or comparison) using the covariates, with stepwise selection used to fit a reduced model with significant ($p<0.05$) predictors of group membership. The logistic regression model was fit for each condition. Table A1 of the appendix shows the variables that were predictive of group membership.

The logistic regression model produces a predicted probability of being an adoption school, and this predicted probability is known as the propensity score (ps). After assigning inverse probability of treatment weights to adoption schools (weight = $1/ps$) and comparison schools (weight = $1/(1-ps)$), the two groups are more balanced on the covariates. Weights were scaled to have a mean of 1 for each condition, and then weights were trimmed so that the maximum possible weight for each condition was 50. After weighting, the two groups are more similar on the background variables (see Table 2 for comparison).

Statistical Analysis

A hierarchical linear regression model (Raudenbush & Bryk, 2002) was used to estimate the adjusted DiD and determine if it was significantly different than zero (e.g., if the mean improvement in growth for adoption schools was different than the mean improvement in growth for comparison schools). A separate model was fit for each condition, with students (level 1) nested within schools (level 2). Random school intercepts were used to account for within-school correlation, and the propensity score weights were applied. The regression model used student SGP as the dependent variable and included the same variables used for the propensity score model as covariates. The model also included the group indicator (adoption vs. comparison), a treatment indicator (=1 for adoption schools during the Interim adoption year, = 0 otherwise), and the interaction of the two indicators, which estimates the DiD.

After testing the overall effect of Interim adoption using the DiD model, we then examined dosage effects of the Periodic Assessments. To examine dosage effects for each condition, we examined whether student growth increased as the number of Interim Assessments (0, 1, 2, 3, 4 or more) and number of Classroom Assessments (0, 1-2, 3-4, 5-6, 7 or more) increased. The dosage effect analysis was limited to the adoption schools during the adoption year. Student SGP was regressed on the same covariates used for the DiD analysis and propensity scores, along with number of Interim and Classroom Assessments taken. Random school intercepts were used to account for within-school correlation.

Finally, we examined the extent that performance on the Interim Assessments was related to student growth. This analysis was limited to students in the adoption schools during the Interim adoption year that took at least one Interim Assessment. Student performance on the Interim Assessments was summarized with the predicted end-of-year Summative score. Within each condition, correlations between Interim performance and SGP were examined. Further, students were placed into quintiles based on their Interim performance. Student SGP was regressed on the same covariates used for the DiD analysis and propensity scores, along with Interim performance quintile. Random school intercepts were used to account for within-school correlation.

Results

Effects of Interim Adoption

Table 3 presents the results of the DiD analyses examining the effects of Interim adoption. Conditions where the estimate was positive and statistically significant are shaded in blue; conditions where the estimate was negative and statistically significant are shaded in orange.

Table 3. DiD Estimates of Effect of Interim Adoption

Subject	Grade Level pair	Effect of Interim adoption on Student Growth Percentile			School Level Effect Size	
		Estimate	Standard Error	P-value	*d*	Δ Growth Percentile
English	3-4	2.53	0.59	<.001	0.31	+12.0
	4-5	3.07	0.60	<.001	0.40	+15.4
	5-6	3.68	0.63	<.001	0.46	+17.9
	6-7	2.78	0.62	<.001	0.35	+13.5
	7-8	3.21	0.58	<.001	0.44	+17.0
	8-9	3.23	0.77	<.001	0.38	+14.8
	9-10	1.58	0.51	0.002	0.22	+8.9
Math	3-4	1.25	0.38	0.001	0.13	+5.2
	4-5	2.48	0.39	<.001	0.26	+10.1
	5-6	-0.66	0.39	0.087	-0.06	-2.5
	6-7	0.66	0.38	0.084	0.07	2.8
	7-8	-0.41	0.37	0.275	-0.04	-1.7
	8-9	2.93	0.92	0.001	0.33	+13.0
	9-10	1.13	0.56	0.042	0.15	+6.0
Reading	3-4	1.88	0.38	<.001	0.25	+9.7
	4-5	0.77	0.39	0.047	0.10	+4.0
	5-6	0.44	0.39	0.253	0.05	+2.1
	6-7	-1.08	0.38	0.005	-0.13	-5.2
	7-8	-1.61	0.38	<.001	-0.21	-8.2
	8-9	0.47	0.82	0.565	0.05	+2.2
	9-10	-1.51	0.52	0.004	-0.19	-7.5

Subject	Grade Level pair	Effect of Interim adoption on Student Growth Percentile			School Level Effect Size	
		Estimate	Standard Error	P-value	d	Δ Growth Percentile
Science	3-4	3.79	0.54	<.001	0.46	+17.9
	4-5	2.11	0.50	<.001	0.23	+9.1
	5-6	3.86	0.54	<.001	0.43	+16.6
	6-7	0.07	0.54	0.894	0.01	+0.3
	7-8	2.38	0.50	<.001	0.28	+11.1
	8-9	-0.27	0.84	0.747	-0.03	-1.3
	9-10	-1.83	0.54	0.001	-0.23	-9.2

d = Estimate / SD where SD is the standard deviation of school mean SGP, Δ Growth Percentile is the estimated increase in the school percentile of mean SGP.

The results varied considerably across the 28 conditions (4 subject areas x 7 grade level pairs). For English, there was strong evidence of positive effects of Interim adoption, with effect estimates ranging from 1.58 for grades 9-10 to 3.68 for grades 5-6 and were positive and statistically significant for all seven grade level pairs. For math, the effect estimates ranged from -0.66 for grades 5-6 to 2.93 for grades 8-9 and were positive and statistically significant for four of the seven grade level pairs. For science, positive and significant effects were observed for grades 3-4, 4-5, 5-6, and 7-8, and a negative and significant effect was observed for grades 9-10. For reading, positive and significant effects were observed for grades 3-4 and 4-5, but negative and significant effects were observed for grades 6-7, 7-8, and 9-10.

To help interpret the size of the effects of Interim adoption, the effect estimates are expressed as a school-level *d* statistic (Table 3). The *d* statistic is calculated as the estimated effect of Interim adoption on mean SGP, divided by the school-level standard deviation of mean SGP. For example, *d*=0.20 suggests that Interim adoption leads to improvement of 0.20 standard deviations in school-level growth. The school-level standard deviations of mean SGP (Appendix Table A2) were calculated using one year of data from each school that tested the majority of students with Aspire Summative in consecutive years. The *d* statistics can also be expressed using the percentile scale. The "Δ Growth Percentile" column of Table 3 represents the estimate of how much a school would improve on a school growth percentile scale (relative to a percentile of 50) with Interim adoption.

The *d* statistics ranged from -0.23 (for science grades 9-10) to 0.46 (for English grades 5-6 and science grades 3-4), with an average *d* of 0.16 across the 28 conditions. The effects of Interim adoption were generally larger for the lower grade levels. The change in school growth percentile ranged from -9.2 (for science grades 9-10) to +17.9 (for English grades 5-6 and science grades 3-4), with an average of +6.2 across the 28 conditions.

Periodic Assessment Dosage Effects

There was variation across students and schools in how many Interim Assessments were administered. Overall, among students in the adoption schools during the first year of adoption, 5% took four or more Interim Assessments, 47% took three, 26% took two, 19% took one, and 2% took no Interim Assessments. The number of Interim Assessments taken varied across

conditions (Table 4). The most Interim Assessments were taken at grades 8-9, followed by grades 3-4 and grades 4-5. Fewer assessments were taken at grades 9-10. Meta-analyzing across the 28 conditions, we find that the variation in assessments taken helps explain some of the variation in Interim adoption effect sizes (i.e., the *d* statistic): The correlation of mean number of Interim Assessments taken and Interim adoption effect size (from Table 3) was 0.25.

Table 4. Frequency Distribution of Number of Interim Assessments Taken

Subject	Grade level pair	Number of Interim Assessments taken, %				
		0	1	2	3	4+
English	3-4	2.7	20.5	29.3	46.0	1.5
	4-5	1.5	15.5	33.4	47.4	2.3
	5-6	2.7	14.7	22.0	56.4	4.2
	6-7	3.0	17.8	23.3	50.5	5.3
	7-8	2.5	17.8	24.1	50.4	5.1
	8-9	2.5	12.6	22.6	62.1	0.1
	9-10	2.7	31.7	28.1	37.3	0.2
Math	3-4	2.0	15.9	16.8	52.0	13.2
	4-5	2.2	17.9	20.1	48.7	11.0
	5-6	2.7	20.6	26.6	45.4	4.7
	6-7	3.2	21.4	29.1	44.4	2.0
	7-8	2.8	20.7	29.4	43.1	4.0
	8-9	1.3	7.3	25.5	65.7	0.2
	9-10	2.5	26.9	31.2	38.4	1.0
Reading	3-4	1.7	16.4	21.8	48.0	12.1
	4-5	1.7	16.4	22.1	47.9	11.9
	5-6	2.1	21.8	24.0	46.0	6.1
	6-7	2.9	20.7	27.1	46.6	2.7
	7-8	2.6	21.1	28.0	46.2	2.1
	8-9	1.0	7.4	23.8	67.6	0.2
	9-10	2.2	22.3	35.2	37.6	2.7
Science	3-4	2.3	21.4	19.0	53.6	3.7
	4-5	1.9	16.9	27.7	52.9	0.7
	5-6	2.9	20.9	30.1	40.0	6.1
	6-7	3.8	25.0	26.5	42.8	2.0
	7-8	3.1	20.9	38.4	35.0	2.5
	8-9	2.1	14.1	24.2	55.8	3.9
	9-10	2.5	20.7	30.6	43.5	2.7

The variation in Interim Assessments taken allows us to examine dosage effects of Interim Assessments. Figure 2 shows the adjusted mean SGP, by number of Interim Assessments taken. The numbers presented in Figure 2 represent the average results across the 28 conditions and are adjusted for the model covariates and number of Classroom Assessments taken. Within the adoption schools during the first year of adoption, academic growth generally increased as students took more Interim Assessments. For example, students taking three

Interim Assessments had an average SGP of 51.9, while those taking one Interim Assessment had an average SGP of 48.3.

Figure 2. Number of Interim Assessments Taken and Average Student Growth Percentile
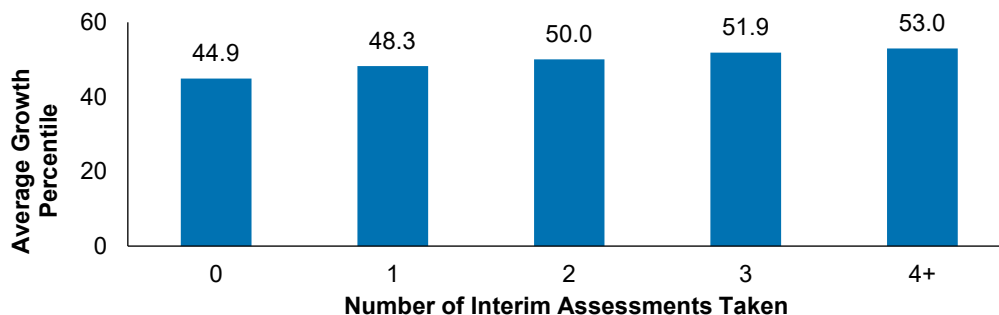

There was also variation across students and schools in how many Classroom Assessments were administered. Overall, among students in the adoption schools during the first year of adoption, 0.8% took seven to 10 Classroom Assessments, 0.9% took five or six, 1.8% took three or four, 9.4% took one or two, and 87.0% took no Classroom Assessments. Use of the Classroom Assessments varied across the 20 conditions[7] (Table 5). Use of the Classroom Assessments decreased with grade level. Meta-analyzing across the 20 conditions, the correlation of mean number of Classroom Assessments taken and Interim adoption effect size was 0.42. However, very few Classroom Assessments were administered within the Interim adoption schools, so it seems unlikely that greater use of Classroom Assessments caused an improved effect of Interim adoption.

Table 5. Frequency Distribution of Number of Classroom Assessments Taken

Subject	Grade level pair	Number of Classroom assessments taken, %				
		0	1-2	3-4	5-6	7+
English	3-4	75.7	23.3	1.0	0.0	0.0
	4-5	90.2	6.1	1.4	0.0	2.3
	5-6	84.7	9.8	5.0	0.5	0.0
	6-7	88.5	9.8	1.1	0.6	0.0
	7-8	86.1	10.9	2.7	0.3	0.0
Math	3-4	84.1	10.7	1.4	1.2	2.6
	4-5	84.3	9.3	3.0	1.2	2.3
	5-6	85.8	10.9	2.6	0.4	0.3
	6-7	92.0	6.1	0.7	1.0	0.3
	7-8	89.9	9.0	0.4	0.3	0.4
Reading	3-4	81.1	11.2	4.3	0.9	2.6
	4-5	82.6	11.9	2.2	1.8	1.6
	5-6	90.2	6.6	2.2	0.7	0.2
	6-7	93.7	5.3	0.6	0.4	0.0
	7-8	90.8	8.4	0.8	0.0	0.0
Science	3-4	86.6	7.0	1.2	5.2	0.0
	4-5	90.7	5.9	1.1	2.2	0.0
	5-6	88.1	8.3	1.6	2.0	0.0
	6-7	85.2	11.3	2.9	0.5	0.0
	7-8	80.9	17.8	1.0	0.3	0.0

Among students in the adoption schools during the first year of adoption of Aspire Interim, we also examined whether taking more Classroom Assessments was associated with higher academic growth. Figure 3 provides the adjusted mean SGP, by number of Classroom Assessments taken. These results are averaged across the 20 subject/grade level combinations, and the average growth percentiles are adjusted for number of Interim Assessments taken, as well as the other model covariates. As number of Classroom Assessments taken increased, academic growth generally increased. For students who took five or six Classroom Assessments, the average SGP was 54.6, compared to 47.1 for students who took no Classroom Assessments. The average SGP was lower for students who took seven or more Classroom Assessments, relative to those who took five or six assessments.

Figure 3. Number of Classroom Assessments Taken and Average Student Growth Percentile
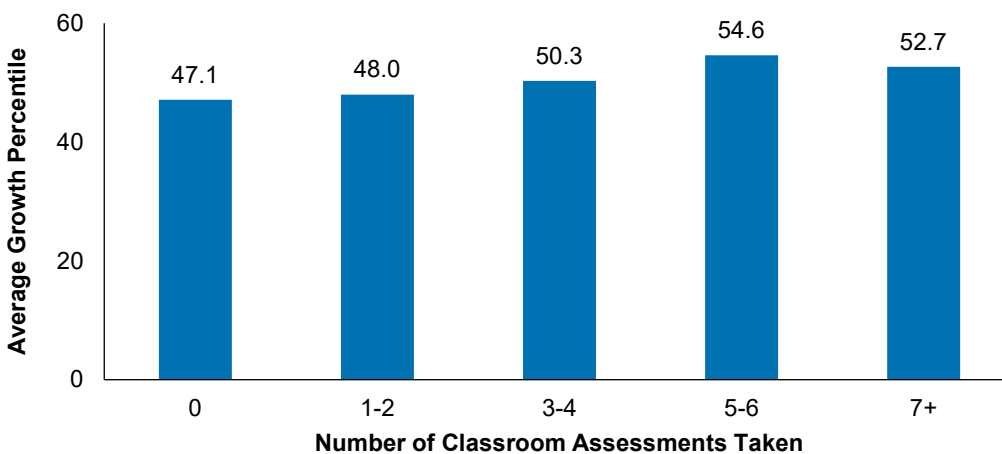

Performance on Interim Assessments

Among students in the adoption schools during the first year of adoption of Aspire Interim, we also examined whether performance on the Interim Assessments was associated with higher academic growth. First, we examine correlations of Interim performance (measured by predicted end-of-year Summative score) with prior year summative score, current year summative score, and SGP (Table 6). Across the 28 conditions, the average correlation of Interim performance with current year summative score was 0.79, slightly higher than the average correlation of Interim performance with prior year summative score (0.76). Correlations of Interim performance with student growth (SGP) ranged from 0.22 (math, grade 8-9) to 0.39 (English, grade 5-6), with an average correlation of 0.31. The positive correlations of Interim performance and SGP suggest that students who perform better on the Interim Assessments are more likely to demonstrate higher growth on the Summative Assessments.

Table 6. Correlations of Aspire Interim Scores with Aspire Summative Scores and Growth

| Subject | Grade level | N | Correlations with Interim Performance | | |
			Prior year Summative score	Current year Summative score	Summative SGP
English	3-4	3,146	0.746	0.787	0.368
	4-5	2,664	0.775	0.799	0.357
	5-6	3,593	0.748	0.806	0.386
	6-7	3,526	0.749	0.802	0.348
	7-8	3,828	0.757	0.794	0.366
	8-9	3,653	0.817	0.855	0.322
	9-10	6,675	0.828	0.850	0.284
Math	3-4	14,243	0.698	0.768	0.323
	4-5	13,939	0.709	0.773	0.332
	5-6	13,384	0.693	0.709	0.265
	6-7	13,574	0.736	0.780	0.308
	7-8	14,422	0.758	0.809	0.289
	8-9	3,348	0.796	0.795	0.217
	9-10	7,061	0.803	0.807	0.242
Reading	3-4	14,867	0.746	0.783	0.331
	4-5	14,716	0.756	0.767	0.289
	5-6	14,241	0.730	0.758	0.299
	6-7	14,314	0.743	0.761	0.304
	7-8	14,694	0.722	0.751	0.315
	8-9	4,734	0.739	0.766	0.346
	9-10	8,758	0.744	0.756	0.286
Science	3-4	5,228	0.778	0.802	0.316
	4-5	5,269	0.771	0.796	0.324
	5-6	4,983	0.761	0.798	0.316
	6-7	4,740	0.762	0.800	0.307
	7-8	6,350	0.762	0.783	0.295
	8-9	3,963	0.807	0.813	0.265
	9-10	6,499	0.798	0.812	0.281

Figure 4 provides the mean SGP, by quintile of performance on the Interim Assessments. The results represent the average across the 28 conditions. As performance on the Interim Assessments increased, academic growth increased substantially. For students who performed in the top quintile on the Interim Assessments, the average student growth percentile was 61.0, compared to 37.5 for students in the bottom quintile.

Figure 4. Performance on Interim Assessments and Average Student Growth Percentile
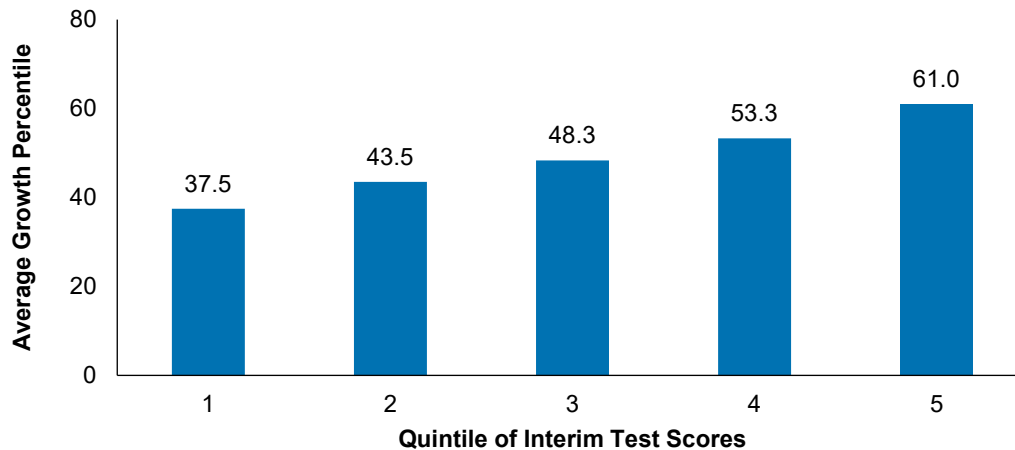
Discussion

Summary

Adoption of ACT Aspire Interim Assessments led to positive improvements in student growth for most subject area/grade level combinations. The effects of adoption were strongest for English and weakest for reading. On average, adoption of the Interim Assessments improved student growth by 1.3 student growth percentile (SGP) units. At the school level, this translates to an effect size of 0.16 on school-level growth, which is like a school at the 50th percentile of school growth improving to the 56th percentile of school growth.

We also found positive dosage effects of Interim and Classroom Assessments: As more assessments were taken, average SGP increased. Performance on the Interim Assessments was strongly related to SGPs, indicating that students who perform well on the Interim Assessments are more likely to demonstrate high growth on the Summative Assessments.

Limitations

A limitation of the study is that we did not account for other periodic assessments (outside of ACT Aspire) that may have been used by study schools. Many of the schools likely administered periodic assessments from other assessment providers. This seems especially likely for comparison group schools and for adoption schools during the first growth period (e.g., prior to Aspire Periodic adoption). It's also possible that some adoption schools administered both the Aspire Periodic Assessment and other periodic assessments during the second growth period. Thus, our study does not provide a clean comparison of "Aspire Periodic versus no Periodic," but rather is more likely "Aspire Periodic" versus "Some other periodic assessments." It's possible that the effect of Aspire Interim adoption would have been larger had the former comparison been possible.

Another limitation is that only one outcome was examined: student growth as measured by performance on the Aspire Summative Assessments. While it's encouraging that Interim adoption led to improvements in student growth, a more robust analysis would examine distal outcomes outside of the Aspire Assessment system. Ultimately, Aspire Interim adoption hopes to improve higher-order thinking skills and transfer of knowledge and skills to real-world situations.

While the study examined use of Aspire Classroom Assessments, low usage of the Classroom Assessments makes it difficult to draw strong conclusions about their effects. On average, only 13% of students in adoption schools took one or more Classroom Assessments. Had this percentage been higher, we might have observed larger differences between adoption and comparison schools.

Future Directions

While the study showed positive effects of ACT Aspire Periodic Assessment, it did not address how the assessment data was used, or how variation in assessment use related to differences in improvement. In theory, use of Periodic Assessments has the potential to improve teaching and learning. However, work by teachers and school/district leadership is needed to realize the potential. Periodic Assessment should not be viewed as a passive act, but rather one that requires a commitment to creating a school culture of professional learning and other supports needed for optimal use of assessments (c.f., Goren, 2010 for more discussion). One idea for additional research would be to survey schools that have used Aspire's Periodic Assessments to understand variation in how the assessment data are used and whether different usage types are related to student growth. Such a survey could also gather data on use of other (non-Aspire) periodic assessments and gather information to help understand the low usage of the Classroom Assessments.

The study revealed variation across conditions (subject areas and grade levels) in the effects of Interim adoption. Some of the variation is explained by differences in how many Periodic Assessments were administered. But additional research is needed to better understand why Interim adoption appears to have a positive effect for some, but not all, conditions. The user survey discussed earlier could also help address questions around differences across conditions.

Notes

1. Testing time is usually 45 minutes or less for English, Reading, and Science; longer testing times are more common for Math.
2. Note that Aspire Interim Assessments are available for grades 3-10 and can therefore impact growth between summative assessments for grades 3-4 through 9-10.
3. Some schools that have administered ACT Aspire could not be matched to either school database, either because they are not included in the school database or because we were unable to match using the data available from the Aspire administration (e.g., school name and district name).
4. https://mdreducation.com/
5. For example, Alabama administered ACT Aspire Summative for grades 3-8 and grade 10 (but not grade 9) and Wisconsin administered ACT Aspire Summative for grades 9-10 (but not grades 3-8).
6. The SGP tables are documented at *https://www.act.org/content/act/en/research/act-growth-modeling-resources.html*.
7. Classroom assessments are available for grades 3-8 (five growth periods) and four subject areas.

References

ACT. (2017). *ACT Aspire technical manual* (Version 4). Iowa City, IA: ACT.

ACT. (2018). *ACT Aspire Periodic technical manual* (Spring 2018 Version 2). Iowa City, IA: ACT.

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research, 46*(3), 399-424.

Burch, P. (2010). The bigger picture: Institutional perspectives on interim assessment technologies. *Peabody Journal of Education, 85*(2), 147–162.

Glander, M. (2016). *Documentation to the 2014–15 Common Core of Data (CCD) universe files* (NCES 2016-077). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved from https://nces.ed.gov/ccd/pdf/2016077_Documentation_062916.pdf.

Goren. P. (2010). Interim assessments as a strategy for improvement: Easier said than done. *Peabody Journal of Education, 85*(2), 125-129. DOI: 10.1080/01619561003673938.

Li, Y., Marion, S., Perie, M., & Gong, B. (2010). An approach for evaluating the technical quality of interim assessments. *Peabody Journal of Education, 85(2)*, 163–185.

Meyer, B. (1995). Natural and quasi-experiments in economics. *Journal of Business & Economic Statistics, 13*(2), 151-161.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Appendix

Table A1. Propensity Score Model Significant Predictors

Subject	Grade level pair	School/unit variables					State		Race/ethnicity						Growth Period			
		N	%RFL	Public	%tTested	Mean prior	AR	AL	Black	Missing	Other	White	Hispanic	Asian	2014-15	2015-16	2016-17	2017-18
English	3-4	-	+	-	-	-			-	-	-				-	-		
	4-5		+		+	-	-	-	-	-		-				+	+	+
	5-6	+	+		+	+	-	-	-	-			+				+	+
	6-7	-	+	-		+	+	+	-				+					-
	7-8	-	+	+	-	+	-	-	+				-	+				
	8-9	-	-	+	-	-	+				+		+		+	+		
	9-10	-	-	+	-	-	-									+	+	+
Math	3-4	+	+	+	+	-	-	+	+		-	+	+			+	+	+
	4-5	+	-	+	+	-	-			-		-			-	-		
	5-6	+		+	+	-	-		+	-					-		+	+
	6-7	+		+	+	-	+	+	+	-			+	+	-		+	
	7-8	+	+	+					+	-	-		+			+	+	+
	8-9		+	+														
	9-10	-	-	+	-	-	-						-			+	+	+
Reading	3-4	+	-	+	+	-	-	+	-	-	-			-		+	+	+
	4-5	+	-	+	+		-	-	-	-	-		+		-	-		
	5-6	+		+	+	-	-		+	-	-					+	+	+
	6-7	+	-	+	+		+	+	+	-	-		+	-		-	+	
	7-8	+	-	+	-		-		+	-	-	-				+	+	+
	8-9		-	+	-	+			+					+				
	9-10	+	-	+	-	-	-			+		-			-	-	-	
Science	3-4	+	+		-	+	+	-		-	-	+	+		+	+		
	4-5	+	+		+	-			-	-	-		+		+	+		
	5-6	-	+	+	-	+			+		-	-			+	+	+	
	6-7	-	+	-	-	+	+		+	+	-		+		+	+		
	7-8	+	+	+	+	-		-	+		-		-		-	+	+	
	8-9	-	-	+	-	+	+		-	+	+		-			+		-
	9-10	-	-	+	-	+			-	+		+	+		-	-		+

+ indicates positive coefficient, - indicates negative coefficient

Table A2. School-level Standard Deviations of Mean SGP

Subject	Grade level	N schools	SD MGP
English	3-4	901	8.3
	4-5	898	7.8
	5-6	676	7.9
	6-7	697	8.0
	7-8	756	7.3
	8-9	519	8.5
	9-10	1,273	7.0
Math	3-4	1,305	9.7
	4-5	1,295	9.7
	5-6	999	10.6
	6-7	970	9.5
	7-8	1,007	9.5
	8-9	532	8.8
	9-10	1,277	7.5
Reading	3-4	1,306	7.6
	4-5	1,295	7.7
	5-6	999	8.3
	6-7	968	8.2
	7-8	1,005	7.8
	8-9	527	8.6
	9-10	1,269	8.0
Science	3-4	962	8.2
	4-5	1,012	9.2
	5-6	780	9.0
	6-7	783	8.6
	7-8	815	8.4
	8-9	516	8.3
	9-10	1,252	7.9

SD = standard deviation, MGP = mean growth percentile

ACT is an independent, nonprofit organization that provides assessment, research, information, and program management services in the broad areas of education and workforce development. Each year, we serve millions of people in high schools, colleges, professional associations, businesses, and government agencies, nationally and internationally. Though designed to meet a wide array of needs, all ACT programs and services have one guiding purpose—helping people achieve education and workplace success.

ACT.org/research