2013 No. 044

A Statistical Investigation of the Effects of Computer Disruptions on Student and School Scores

Final Report

Prepared for: **Minnesota Department of Education**
1500 Highway 36 West
Roseville, MN 55113

Prepared under: **A85179/A94278**

Authors: **Bethany H. Bynum**
R. Gene Hoffman
Matthew S. Swain

Date: **August 15, 2013**

HumRRO®
Human Resources Research Organization

Human Resources Research Organization (HumRRO)
66 Canal Center Plaza, Suite 700, Alexandria, Virginia 22314-1578
Phone: 703.549.3611 | Fax: 703.549.9661 | www.humrro.org

A Statistical Investigation of the Effects of Computer Disruptions on Student and School Scores

Table of Contents

Defining Computer Disruption 1

Propensity Matching 3

2013 Score Differences 5

- Differences in Average 2013 Test Scores 5
- Examining the Predictability of 2013 Test Scores 6
- Examine Distributions of Predicted Student Scores 7
- Compare Predictions for Disrupted Students to Non-Disrupted Students 10
- Student-Level Summary 11

Impact of Disruption on School-Level Scores 11

- Distribution of School Disruptions 11
- School-Level Score Differences 13
- School-Level Classification Differences 18
- School-Level Summary 18

Conclusions 19

References 20

Appendix A 21

List of Tables

Table 1. Page Latency in Seconds 2

Table 2. Percentage of Students with Indicated Latencies 2

Table 3. Percentage of Students with Latencies Greater than 10 Seconds 3

Table 4. Mean Differences between Disrupted and Non-Disrupted Groups by Grade 5

Table 5. Incremental Validity Estimation of Disruption 6

Table 6. Predictability of 2013 Scores for Non-Disrupted and Disrupted Groups 7

Table 7. Distribution of the Difference between Predicted and Observed Reading Scores for Non-Disrupted and Disrupted Groups 8

Table 8. Distribution of the Difference between Predicted and Observed Math Scores for Non-Disrupted and Disrupted Groups 9

Table 9. Percent of Disrupted Students with Predicted and Observed Score Differences at the 5th, 10th, 90th and 95th Percentile of Non-Disrupted Students 11

Table 10. Sample Distribution of Schools 12

Table 11. Average Percent of Students Tested on April 16 and 23 for All Schools that Tested on April 16 and 23 13

Table 12. Mean Scores for Schools that Tested on April 16 or 23 15

Table 13. Mean Scores for Schools with Students in the Disrupted Sample 16

Table 14. Incremental Validity Estimates for Schools with Students in the Disrupted Sample....17
Table 15. Difference in School Classification for Schools with Students in the Disrupted Sample....18
Table A.1. Mean Covariate Differences for Reading Grade 3....21
Table A.2. Mean Covariate Differences for Reading Grade 4....22
Table A.3. Mean Covariate Differences for Reading Grade 5....23
Table A.4. Mean Covariate Differences for Reading Grade 6....24
Table A.5. Mean Covariate Differences for Reading Grade 7....25
Table A.6. Mean Covariate Differences for Reading Grade 8....26
Table A.7. Mean Covariate Differences for Reading Grade 10....27
Table A.8. Mean Covariate Differences for Math Grade 3....28
Table A.9. Mean Covariate Differences for Math Grade 4....29
Table A.10. Mean Covariate Differences for Math Grade 5....30
Table A.11. Mean Covariate Differences for Math Grade 6....31
Table A.12. Mean Covariate Differences for Math Grade 7....32
Table A.13. Mean Covariate Differences for Math Grade 8....33

A Statistical Investigation of the Effects of Computer Disruptions on Student and School Scores

In April 2013, the Human Resources Research Organization (HumRRO) was asked by the Minnesota Department of Education (MDE) to investigate the effects of computer disruptions for students taking online versions of the Minnesota Comprehensive Assessment Series III (MCA-III) and MCA-Modified. HumRRO has been MDE's independent psychometric quality assurance contractor since 2006. On April 16 and 23 many students testing online in reading, math, and science experienced some sort of computer-related testing disruptions, including delays in item delivery and unexpected logouts. In order to ensure that reported results are valid, reliable and fair, HumRRO investigated the impact of the computer disruptions on students' scores. We worked with MDE and the Technical Advisory Committee (TAC) to conduct a thorough analysis of the assessment data. Our investigation is based on the well-accepted premise that students' test scores tend to exhibit consistency over time; that is, students who do well in one year of the test tend to do well the subsequent year. Because most students did not experience computer-related disruptions, we used a set of variables to match disrupted and non-disrupted students to help estimate the impact on those who experienced computer-related issues. This document summarizes the results of the statistical investigation for MCA-III reading and math grades 3 – 8 and reading grade 10.

Defining Computer Disruption

HumRRO received student testing data from American Institutes of Research (AIR) for April 16, 23, and 30 test administrations that included latency information. Latency refers to the amount of time it takes for an online page to load. Two levels of latency were reported in the data: page-level latency and item-level latency. AIR's Test Delivery System collected latency by item group. An item group can consist of one discrete item or a group of items associated with a passage. In the event that there is more than one item in a group, AIR calculates item-level latency by taking the average across that group. That is, if there were five items on a page and the total page latency was 50 seconds, the item-level latency would be 10 seconds for each item. For science and math, each item appears on its own page, so page and item-level latency are the same. For reading, all items associated with a passage appear on the same page. Based on this information, HumRRO focused on the latency of each page. Again, for science and math a page includes one item, and for reading a page includes a group of items.

For many students, latencies data was incomplete. Latency is only retained by the system when a student pauses or completes a test. However, if a student is unexpectedly logged out of the testing session, then data is not recorded. Also, if a student paused the test and resumed on the same day, only the most recent latency was captured. For these reasons, incomplete latency data may or may not indicate computer disruptions. Identifying students with disruptions based on latency data is imperfect and should be considered a lower bound estimate of the numbers of students who were actually disrupted.

With this in mind, we examined the data to identify abnormally long latencies and identify the number of long latencies that could potentially be considered as disruptive. First, we examined the distribution of latencies for the two days with large numbers of disruptions (April 16 and 23) and for one day with no disruptions (April 30). Table 1 shows the distribution of latency times and Table 2 shows the percent of students with latencies greater than 30, 20, 15, 10 and 5 seconds.

Table 1. Page Latency in Seconds

Date	Latency Records	Mean	Median	90th	95th	99th
April 16[a]	129,980	4.46	0.20	2.20	23.41	107.80
April 23[a]	1,072,104	2.56	0.12	0.73	2.63	80.54
April 30[b]	2,176,700	0.20	0.10	0.35	0.55	1.29

Note. Latency Records count includes all records retained by AIR's system for a given day, excluding the first record which includes some start-up time and is not indicative of *true latency*. The total count includes student records that were captured but no latency time was recorded. The average number of latency records per student is larger on April 23 and 30 because a large proportion of students took math and science tests on those dates.
[a] Disruptions occurred.
[b] No disruptions occurred.

Table 2. Percentage of Students with Indicated Latencies

Date	N	> 30 sec	20-30 sec	15-20 sec	10-15 sec	5-10 sec
April 16[a]	23,077	11.08	2.61	1.37	1.62	3.48
April 23[a]	65,866	16.69	4.68	2.66	2.95	6.61
April 30[b]	95,025	0.25	0.09	0.11	0.19	1.15

[a] Disruptions occurred.
[b] No disruptions occurred.

Table 1 shows that on a normal testing day, April 30, the average latency is only 1/5th of a second and a long latency (99th percentile) is just over a second. In contrast, on April 16 and 23 the average latency was over four and two seconds respectively, with some students experiencing latencies over a minute. Table 2 shows the percentage of students who experienced long latencies on each testing day. On April 16, over 11 percent of students experienced a latency of 30 seconds or greater. On April 23, over 16 percent of students experienced a latency of 30 seconds or greater, with nearly 27 percent of students experiencing at least one latency of 10 seconds or longer. On a normal testing day, one percent of students experience latencies between five and ten seconds but substantially less than one percent of students experienced latencies greater than ten seconds. Based on this, we determined that latencies greater than 10 seconds were exceptionally abnormal.

In addition to latency length, the number of abnormal latencies could affect students' testing experience. For science and math, each page is one item; thus, students experience a latency for each item. A student who encounters 50 items and experiences only one abnormal latency is less likely to be affected than a student who goes through the same number of items and experiences multiple abnormal latencies. In contrast, the reading test has a set of items per page. Therefore, students do not see as many pages or experience as many page loads for the reading test. For reading, if a student is distracted by an abnormal latency and then has to read a passage and respond to a set of items, the entire set of items may be affected by the disruption. As a result, it is possible that one abnormal latency could have a larger effect in reading than in science or math.

Table 3 includes students with at least one latency greater than 10 seconds and, for those students, presents the percentage of the total latencies that were greater than 10 seconds based on the latency data available. For example, for math there are 8,180 students who experienced at least one latency greater than 10 seconds, and on average 16% of those student's total latencies were greater than 10 seconds.

Table 3. Percentage of Students with Latencies Greater than 10 Seconds

	n	Mean Pages	Mean	Median	10th	90th
Math	8,180	26	16.38%	11.76%	3.85%	34.70%
Science	401	34	13.69%	10.00%	2.44%	33.33%
Reading	6,576	6	48.54%	44.44%	11.54%	100.00%

For science and math, we selected students for the score impact study by flagging those with abnormal latencies (> 10 seconds) for at least 10% of the testing data captured. Of the 4,882 students identified with this approach in math, 98% had at least one latency greater than 30 seconds. For science, of the 203 students identified with this approach, 196 students had at least one latency of 30 seconds or longer. These results suggest that students who experienced multiple abnormal latencies, tended to experience at least one latency longer than 30 seconds. For reading, we selected students with (a) abnormal latencies greater than 10% of their test or (b) at least one latency of 30 seconds or more.

In addition to long latencies, computer disruptions also included abnormal restarts, where a student was unexpectedly logged out of the testing environment during testing and required to log in again to resume the test. Any student who experienced an abnormal restart was included in the score impact study.

Finally, AIR provided data on students who were administratively paused on April 16 and 23. Administrative pauses are pauses to the test initiated by a teacher or the test proctor. Students also have the ability to pause their test for situations such as bathroom breaks. About 5% of pauses are administrative on a normal testing day. Although there were an abnormal number of administrative pauses on these two days, it was not clear why this occurred. Given there was not a direct technical computer interruption, these records were not included in the score impact study.

Overall, we selected students to include in the score impact study based on known disruptions. This is not to say that other students were not affected by the computer disruptions, directly and indirectly. However, if we were unable to detect an effect in the concentrated sample, then it is unlikely we would find an effect in a larger sample of students who experienced less disruption.

Propensity Matching

To determine the impact of the technical disruptions, we first matched the disrupted students with a group of students who tested on the non-affected administration dates. The disrupted sample included students who tested on April 16 or April 23 and had either abnormal latencies or an unexpected restart, as described above. The matched non-disrupted sample was chosen from students who did not test on either of those days.

Propensity matching is a statistical approach to match sample groups on a set of variables that are likely to be related to the outcome of interest when random group assignment is not possible. For this study, propensity matching attempted to statistically control for naturally occurring variables that were unrelated to the disruption, but might impact 2013 scores. The result was two groups of students who were as closely matched as we could make them except for their experience with computer disruptions and 2013 test scores. By matching groups on these variables, the difference in 2013 test scores between the two groups could be more strongly attributed to disruption.

The following variables were used for matching:

- 2012 math and reading theta scores
- Gender (0 = Female, 1 = Male)
- Ethnicity (Dummy coded for Asian, Hispanic, Black, White)
- Eligibility for Free/Reduced-price Lunch status (0 = No, 1 = Yes)
- Limited English Proficiency (LEP) status (0 = No, 1 = Yes)
- School-level percentage of Free/Reduced-price Lunch students
- School-level achievement (Average of 2012 student-level reading and math scores)

Additionally, for math, the number of times a student took the math test in 2012 was included as a predictor to account for any increase in test score that was due to taking the test multiple times. For grade 3, where there are no prior year test scores, current year score for the non-disrupted subject (reading or math) was used.

Prior to matching, we examined differences between the two samples on these variables using Cohen's *d*. There were a number of small to moderate differences (Cohen's $d > .10$ and $< .30$) between the disrupted and non-disrupted samples. Generally, the disrupted sample had slightly higher school-level achievement and lower percentages of free/reduced-price lunch students, although this was not the case for all grades and subjects. These results suggest that propensity matching is necessary to ensure the two samples are equivalent. Appendix A reports the table of differences prior to matching.

Next, using logistic regression, we regressed group membership (disrupted or not disrupted) onto the matching variables. Generally, the pseudo R^2 values of the logistic regression were small, ranging from .0012 to .0059. The small values suggest that overall, the combination of prior year student achievement, demographics, socioeconomic status (SES), school-level achievement or school-level SES had little relationship to the likelihood that a student experienced disruptions.

To match the two samples, the predicted disruption probabilities from the logistic regression analyses were saved. The predicted probabilities represent the probability that a student was in the disrupted sample. We used the nearest neighbor method to match the two samples. The predicted disruption probability for each student in the disrupted sample was matched to the student with the closest predicted disruption probability in the non-disrupted sample. The sampling was done without replacement so that each student in the disrupted sample was matched with a unique student in the non-disrupted sample. The average difference between a disrupted student's predicted probability and the matched non-disrupted student's predicted probability was .00004 for reading and .000002 for math. The largest difference was .04. Differences no larger than .20 have been shown to reduce bias and produce accurate group difference estimates (Austin, 2009; Connelly, Sackett, & Waters, 2013). The results suggest that everyone in the disrupted sample was matched with a student in the non-disrupted sample with a very similar predicted disruption probability. To further evaluate the closeness of the matched sample, we examined the mean difference of the matching variables. The average Cohen's *d* between the two samples was .002 and ranged from 0 to .119. All of the differences greater than .10 prior to matching were reduced. Cohen's *d* effect sizes near zero suggest that the samples were effectively balanced on prior year achievement, school-level achievement, gender, ethnicity, race, free/reduced-price lunch status, LEP status, and school-level free-

reduced percentage. A summary of the mean, standard deviations, and effect sizes before and after matching are found in Appendix A.

2013 Score Differences

Using the matched samples, we examined whether students' scores from tests with computer disruptions differed from students' scores from tests that were not disrupted. By matching the samples on variables that are likely to predict student scores, any difference between the two samples can be better attributed to the computer disruptions. We used a number of analyses to examine differences in scores. Scores from 2013 were maximum likelihood theta estimates for reading and math. MDE made a policy decision to allow all students who tested on April 16 and 23 to review all answers in the last segment of the test before their sessions ended on those days. Additionally, districts could request that a student be permitted to review all items on the test or take a new test. However, no data were captured to identify which, if any, students reviewed part (or all) of a test or whether any student retook the test. Only the final 2013 test scores were included in these analyses. It is possible that we might see scores that are lower or higher than expected for students who were disrupted.

Differences in Average 2013 Test Scores

If testing disruptions had no overall impact, then the averages of the 2013 test scores for the two matched groups would be expected to be nearly identical. On the other hand, differences in the average 2013 test scores would be evidence that computer disruption did impact test performance. We examined mean differences between the two samples using a *t*-test and Cohen's *d* effect size. Table 4 presents the results.

Table 4. Mean Differences between Disrupted and Non-Disrupted Groups by Grade

		Non-Disrupted		Disrupted			
Grade	*n*	Mean	*SD*	Mean	*SD*	*t*-value	*d*
Reading							
3	986	0.11	1.10	0.07	1.08	0.96	.04
4	980	-0.03	1.12	0.11	1.09	**-2.68**	-.12
5	843	0.18	1.09	0.12	1.05	1.27	.06
6	718	0.14	1.13	0.16	1.11	-0.33	-.02
7	882	0.11	1.05	0.11	0.97	-0.10	.00
8	1058	0.13	1.04	0.05	0.99	1.64	.07
10	2,265	0.14	0.97	0.18	0.97	-1.65	-.05
Math							
3	1287	-0.05	1.08	-0.01	1.05	-1.08	-.04
4	815	0.05	1.09	-0.07	1.03	**2.31**	.11
5	691	0.14	1.06	0.10	1.09	0.84	.05
6	659	0.20	1.01	0.14	0.98	1.00	.06
7	675	0.23	0.99	0.21	1.07	0.33	.02
8	740	0.13	1.07	0.11	0.97	0.38	.02

Note. Bolded values indicate statistical significance at $p < .05$.

For reading, the effect sizes ranged from -.12 to .07, suggesting no overall directional effect. The mean difference was statistically significant for grade 4, where the disrupted sample had

higher theta scores than the non-disrupted sample. However, the overall effect size was small (*d* = .12). For math, the effect sizes ranged from -.04 to .11. With the exception of grade 3, the non-disrupted group had higher theta scores than the disrupted group, but generally the differences were small. There was a statistically significant difference for grade 4 but the Cohen's *d* of .11 indicates a small effect.

Examining the Predictability of 2013 Test Scores

First, to examine predictability of 2013 test scores, we examined the incremental variance that disruption added to the prediction of the 2013 test scores beyond other known indicators of performance (including prior year achievement, free/reduced-price lunch status, LEP status, race, ethnicity, gender, school-level achievement, and school-level percentage of free-reduced lunch). If the inclusion of disruption in the multiple regression models adds to the estimation of 2013 scores, then this would supply evidence that disruption impacted 2013 test scores. For this model, the two groups were included in the regression models together.

Table 5 reports the R^2 values for each model and the R^2 change between the two models. With the inclusion of disruption in the model, the R^2 value changed very little or not at all, indicating that disruption did not add to the prediction of 2013 scores.

Table 5. Incremental Validity Estimation of Disruption

Grade	*n*	Covariates Only R^2	Covariates + Disruption R^2	ΔR^2
Reading				
3	1972	.580	.581	.001
4	1960	.652	.652	.000
5	1686	.643	.643	.000
6	1436	.628	.628	.000
7	1764	.616	.616	.000
8	2116	.641	.642	.001
10	4,530	.643	.643	.000
Math				
3	2574	.615	.615	.000
4	1630	.751	.754	.002
5	1382	.728	.729	.000
6	1318	.736	.737	.001
7	1350	.768	.768	.000
8	1480	.737	.737	.000

Note. ΔR^2 is the difference in the R^2 values between the Covariates Only model and the Covariates + Disruption model.

Next, using all available data to create equations that predicted 2013 test scores, we can statistically determine whether disrupted students scored differently than expected. Prediction equations were estimated for the disrupted students and separately for the matched sample of non-disrupted students. In addition to prediction equations, this technique gave us multiple regression coefficients (R^2). R^2 are interpreted like correlation coefficients to tell us how well 2013 test scores can actually be predicted. If students' performance was affected, the strength of the prediction for the disrupted students should be less than for the non-disrupted students as shown by lower multiple regression coefficients. A lower coefficient means that students' performance in the disrupted group was not as predictable as students' performance in the non-

disrupted group. This would supply another piece of evidence about the impact of the computer disruptions.

Table 6 presents the multiple R^2 values for the disrupted and non-disrupted groups. Overall, 2013 test scores were well predicted for both samples, with 57% to 66% of the variance accounted for by the predictor variables for reading and 60% to 77% for math. Note that prediction is lower for grade 3 in both subjects, because students in grade 3 did not have prior year test scores, the strongest predictor of 2013 test scores. Generally, there were slightly higher R^2 values for the non-disrupted group, although the difference in variance accounted for was practically small, ranging from .3% to 5%.

Table 6. Predictability of 2013 Scores for Non-Disrupted and Disrupted Groups

		Non-Disrupted	Disrupted	
Grade	*n*	R^2	R^2	R^2 Difference
Reading				
3	986	.626	.575	.051
4	980	.663	.640	.023
5	843	.653	.635	.018
6	718	.654	.606	.047
7	882	.624	.612	.012
8	1058	.642	.645	-.003
10	2,265	.651	.659	.012
Math				
3	1287	.626	.606	.020
4	815	.774	.732	.042
5	691	.735	.726	.009
6	659	.751	.747	.003
7	675	.772	.769	.003
8	740	.763	.709	.054

Note. R^2 Difference is the difference between the Non-disrupted R^2 and the Disrupted R^2.

Examine Distributions of Predicted Student Scores

The prediction equations for the non-disrupted students give us a statistical statement about what to normally expect for students testing under non-disrupted conditions. The prediction is not perfect, but given the high R^2 values we can use the prediction equation to calculate how disrupted students might have scored had they not been disrupted. For each disrupted student, we computed their 2013 predicted score using the regression equation computed for the non-disrupted students. Next, we computed the difference between the predicted score and observed score, where positive values indicate higher predicted scores than observed and negative values indicate higher observed scores than predicted. Table 7 and 8 present the distribution of observed and predicted scores for the disrupted and non-disrupted sample for reading and math, respectively. The difference between observed and predicted scores is also reported.[1]

[1] Technical Note: When a prediction equation is derived on one sample and applied to a second sample, the variance of the residuals is expected to be larger due to shrinkage. Our predictions of performance for the disruption group were slightly weaker than would be expected based on the shrinkage associated with applying the prediction equation to a randomly equivalent sample. Given that our second sample was not randomly equivalent, but differs by the computer disruption, the small difference suggests that our prediction utility is not severely reduced in the disrupted sample. However, we did investigate differences in predicted scores for the two samples further.

Table 7. Distribution of the Difference between Predicted and Observed Reading Scores for Non-Disrupted and Disrupted Groups

	Non-Disrupted						Disrupted					
	Mean	*SD*	5th	10th	90th	95th	Mean	*SD*	5th	10th	90th	95th
Grade 3 (*n* = 986)												
Difference	0.062	0.721	-1.092	-0.814	0.984	1.220	0.138	0.727	-1.065	-0.785	1.048	1.314
Predicted Theta	0.177	0.802	-1.207	-0.825	1.163	1.392	0.205	0.803	-1.250	-0.849	1.174	1.361
Observed Theta	0.115	1.099	-1.840	-1.265	1.508	1.865	0.068	1.079	-1.763	-1.359	1.400	1.789
Grade 4 (*n* = 980)												
Difference	0.000	0.649	-1.085	-0.796	0.798	1.058	-0.032	0.656	-1.144	-0.826	0.782	1.027
Predicted Theta	-0.027	0.910	-1.634	-1.185	1.034	1.404	0.075	0.866	-1.451	-1.022	1.159	1.537
Observed Theta	-0.027	1.117	-2.044	-1.450	1.310	1.701	0.107	1.086	-1.686	-1.241	1.479	1.856
Grade 5 (*n* = 843)												
Difference	0.000	0.640	-1.013	-0.755	0.790	1.023	0.012	0.640	-1.058	-0.782	0.787	1.011
Predicted Theta	0.184	0.879	-1.301	-0.973	1.255	1.665	0.130	0.865	-1.299	-0.953	1.211	1.512
Observed Theta	0.184	1.087	-1.648	-1.216	1.487	1.857	0.118	1.045	-1.643	-1.171	1.412	1.872
Grade 6 (*n* = 718)												
Difference	0.000	0.662	-1.033	-0.781	0.878	1.096	-0.039	0.701	-1.197	-0.959	0.818	1.064
Predicted Theta	0.136	0.910	-1.378	-1.012	1.247	1.613	0.117	0.854	-1.378	-0.976	1.163	1.447
Observed Theta	0.136	1.125	-1.750	-1.294	1.508	1.952	0.156	1.105	-1.669	-1.181	1.532	2.035
Grade 7 (*n* = 882)												
Difference	0.000	0.646	-0.995	-0.809	0.803	1.071	-0.023	0.615	-1.030	-0.795	0.740	0.956
Predicted Theta	0.109	0.832	-1.275	-0.946	1.227	1.484	0.091	0.807	-1.231	-0.905	1.168	1.428
Observed Theta	0.109	1.053	-1.750	-1.235	1.401	1.839	0.114	0.972	-1.417	-1.080	1.305	1.645
Grade 8 (*n* = 1,058)												
Difference	0.000	0.623	-0.960	-0.742	0.747	1.004	0.056	0.596	-0.936	-0.703	0.811	1.054
Predicted Theta	0.127	0.833	-1.298	-0.912	1.110	1.488	0.110	0.792	-1.175	-0.901	1.133	1.413
Observed Theta	0.127	1.040	-1.587	-1.201	1.401	1.808	0.055	0.990	-1.571	-1.156	1.385	1.678
Grade 10 (*n* = 2,265)												
Difference	-0.006	0.576	-0.922	-0.725	0.697	0.944	-0.022	0.590	-1.046	-0.767	0.710	0.929
Predicted Theta	0.131	0.786	-1.215	-0.883	1.088	1.368	0.163	0.781	-1.185	-0.824	1.106	1.350
Observed Theta	0.137	0.975	-1.499	-1.052	1.336	1.634	0.185	0.973	-1.471	-1.042	1.386	1.727

Note. Bolded values indicate significant difference in standard deviation at $p < .05$. For the non-disrupted group, if all variables are normally distributed, the mean difference between the predicted and observed theta should be zero.

Table 8. Distribution of the Difference between Predicted and Observed Math Scores for Non-Disrupted and Disrupted Groups

	Non-Disrupted						Disrupted					
	Mean	*SD*	5th	10th	90th	95th	Mean	*SD*	5th	10th	90th	95th
Grade 3 (*n* = 1,287)												
Difference	-0.003	0.658	-1.053	-0.827	0.852	1.092	-0.049	0.662	-1.140	-0.882	0.783	1.021
Predicted Theta	-0.056	0.855	-1.554	-1.188	0.994	1.260	-0.057	0.814	-1.529	-1.145	0.939	1.201
Observed Theta	-0.054	1.076	-1.971	-1.447	1.323	1.585	-0.009	1.047	-1.853	-1.390	1.250	1.613
Grade 4 (*n* = 815)												
Difference	0.000	0.516	-0.837	-0.664	0.620	0.848	0.100	0.540	-0.731	-0.572	0.739	1.019
Predicted Theta	0.050	0.955	-1.716	-1.207	1.205	1.566	0.029	0.902	-1.646	-1.155	1.114	1.413
Observed Theta	0.050	1.085	-1.913	-1.323	1.392	1.791	-0.071	1.032	-1.925	-1.426	1.200	1.546
Grade 5 (*n* =691)												
Difference	0.000	0.544	-0.912	-0.684	0.617	0.881	0.026	0.578	-0.900	-0.682	0.768	0.978
Predicted Theta	0.144	0.906	-1.448	-1.082	1.242	1.629	0.122	0.957	-1.587	-1.166	1.335	1.646
Observed Theta	0.144	1.057	-1.584	-1.207	1.485	1.806	0.096	1.089	-1.940	-1.281	1.462	1.765
Grade 6 (*n* =659)												
Difference	0.000	0.505	-0.792	-0.612	0.617	0.847	0.067	0.536	-0.792	-0.611	0.757	0.952
Predicted Theta	0.195	0.877	-1.299	-0.970	1.284	1.533	0.207	0.887	-1.380	-0.990	1.258	1.565
Observed Theta	0.195	1.012	-1.545	-1.183	1.408	1.762	0.140	0.984	-1.495	-1.152	1.410	1.716
Grade 7 (*n* =675)												
Difference	0.000	0.474	-0.736	-0.604	0.592	0.779	-0.001	0.527	-0.780	-0.609	0.598	0.862
Predicted Theta	0.232	0.872	-1.209	-0.891	1.296	1.563	0.213	0.889	-1.312	-0.936	1.256	1.537
Observed Theta	0.232	0.992	-1.465	-1.088	1.527	1.791	0.213	1.071	-1.731	-1.205	1.489	1.795
Grade 8 (*n* =740)												
Difference	0.000	0.521	-0.824	-0.638	0.655	0.850	0.021	0.531	-0.827	-0.599	0.706	0.941
Predicted Theta	0.133	0.935	-1.593	-1.054	1.210	1.522	0.134	0.871	-1.477	-0.951	1.164	1.447
Observed Theta	0.133	1.070	-1.848	-1.261	1.378	1.752	0.112	0.972	-1.629	-1.059	1.272	1.649

Note. Bolded values indicate significant difference in standard deviation at $p < .05$. For the non-disrupted group, if all variables are normally distributed, the mean difference between the predicted and observed theta should be zero.

Compare Predictions for Disrupted Students to Non-Disrupted Students

Large numbers of students with *notable differences* between observed and predicted scores provides another piece of evidence about the impact of the computer disruptions. We defined large number and notable difference by comparing the difference in observed and predicted scores between the non-disrupted and disrupted groups. The non-disrupted group represented the baseline: what would be expected under normal testing conditions.

First, we examined the difference in the standard deviation for the disrupted and non-disrupted group using the F-test to test the equality of the two variances. This provides information on whether the spread of differences in predicted and observed scores is statistically different for the two groups. Results suggest that the standard deviations are not statistically different.

Next, we computed the difference in observed and predicted scores at the 5th, 10th, 90th and 95th percentile for the non-disrupted group and determined the number of students in the disrupted group who were at or below the same cut point for the 5th and 10th percentile and those that were at or above the cut point for the 90th and 95th percentile. If more than 5% and 10% of the disrupted students were below the 5th and 10th percentile cuts, respectively, then more students in the disrupted group scored higher than expected. If more than 10% and 5% of the disrupted students were above the 90th and 95th percentile cut, respectively, then more students in the disrupted group scored lower than expected. Either case would provide evidence that the computer disruption had an impact on scores. Table 9 presents the percent of students in the disrupted group below the 5th and 10th percentile cuts and above the 90th and 95th percentile cuts.

For reading, the results show a slightly higher percent of students above the 90th and 95th cut point for grades 3 and 8. These results suggest that for grade 3, approximately 6.69% of students had higher predicted scores than observed score; 1.69% higher than what would have been expected based on the non-disrupted sample. For grade 8, 12.29% of students had higher predicted scores than observed scores; 2.29% higher than what would have been expected. In contrast, there were a higher percentage of students below the 5th and 10th percentile cut point for reading grades 4, 5, and 6. The largest differences were in grade 6, where 8% of the sample was below the 5th percentile cut and nearly 15% of the sample was below the 10th percentile cut. This suggests that nearly 5% more students had higher observed scores than predicted scores than what have been expected, providing evidence that the disruption may have advantaged these students.

The results do not follow the same pattern for math. For grades 3 and 7, there were higher percentages of students below the 5th and 10th percentile cuts, indicating higher observed than predicted scores. For grades, 4, 5, 6, 7 and 8 there were higher percentages of students above the 90th and 95th percentile cut, indicating a number of students had larger differences than would be expected based on the non-disrupted sample. The percentages range from .4 – 5.1% above the expected 10% at the 90th percentile cut and 1.2 – 2.7% above the expected 5% at the 95th percentile cut. It is important to note, that given our sample sizes, these discrepancies amount to unexpected differences between predicted and observed scores for 10 to 40 students depending on the grade.

Table 9. Percent of Disrupted Students with Predicted and Observed Score Differences at the 5th, 10th, 90th and 95th Percentile of Non-Disrupted Students

		5th	10th	90th	95th
Grade	*n*	Larger than 5% or 10% = Advantaged		Larger than 10% or 5% = Disadvantaged	
Reading					
3	986	4.46%	9.33%	10.95%	6.69%
4	980	5.61%	11.02%	9.69%	4.69%
5	843	6.05%	10.56%	9.85%	4.98%
6	718	8.08%	14.76%	8.36%	4.04%
7	882	5.67%	9.64%	7.48%	3.74%
8	1058	4.63%	8.41%	12.29%	5.58%
10	2,265	6.75%	11.43%	10.24%	4.90%
Math					
3	1287	6.84%	11.50%	8.47%	3.73%
4	815	3.31%	6.75%	15.09%	7.24%
5	691	4.78%	9.84%	14.76%	7.09%
6	659	5.01%	9.86%	14.42%	7.74%
7	675	5.93%	10.22%	10.38%	6.22%
8	740	5.00%	8.65%	11.08%	6.62%

Student-Level Summary

A number of analyses were conducted to examine the impact of computer disruptions on student scores. The statistical evidence provided in this report is intended to inform MDE and the TAC about whether the computer disruption systematically impacted student test scores. The evidence shows that there were disruption effects; however, the effects on students' scores were neither widespread nor large. In addition, the evidence shows that disruptions were not consistently detrimental, but at times were beneficial. The lack of widespread, consistently negative effects from disruptions may well have been prevented by MDE policy allowing students to review items or retake the test.

Impact of Disruption on School-Level Scores

School-level accountability is based on the aggregation of student-level scores. While computer disruptions did not have a large impact on student-level scores, there is concern that once aggregated the disruption could impact school-level accountability results. We investigated the impact of computer disruption on school scores using the following set of analyses.

Distribution of School Disruptions

We defined schools as disrupted or not disrupted. Not disrupted (non-impacted) schools were those that tested online but did not test on April 16 or 23 or schools that tested on paper. Disrupted schools were those tested online on April 16 or 23; schools within this categorization were more problematic to define.

As described above, a disrupted *sample of students* was defined as testing on April 16 or 23 and had either abnormal latencies or abnormal restarts. However, because of the amount of

missing latency data, we could not conclude that students who tested on April 16 or 23, who were not included in the disrupted sample, were not disrupted. A number of these students could have been disrupted but missing data prevented us from identifying them definitively. Therefore, we divided schools that tested on April 16 and 23 into (a) schools with known disruptions, where some proportion of their students was included in the disrupted sample, and (b) schools with potential disruptions.[2] Table 10 presents the sample of schools in each category. Of the schools that tested online in Reading, 40% to 60% of the schools tested students on April 16 or 23 and 15% to 23% of schools had known disruptions. For math, 22% to 34% of schools tested on April 16 or 23 and 9% to 15% of schools had known disruptions.

Table 10. Sample Distribution of Schools

		Schools Tested on April 16 or 23		Non-Impacted Schools	
Grade	Total # Schools	Potentially Disrupted Sample	Known Disrupted Sample	Did Not Test on April 16 or 23	Tested on Paper
Reading					
3	938	102	92	207	537
4	923	105	93	200	525
5	910	109	67	209	525
6	742	101	57	183	401
7	668	85	49	188	346
8	685	74	56	204	351
10	727	120	81	137	389
Math					
3	922	151	133	556	82
4	917	154	88	591	84
5	898	143	81	587	87
6	735	87	59	509	80
7	652	71	56	433	92
8	671	73	56	448	94

Note. For schools in the known disrupted sample, a proportion of their students were included in the disrupted sample for the student-level analyses. Schools in the potentially disrupted sample may or may not have had disrupted students, but none were identified based on the available data.

Table 11 provides a summary of the average percentage of students who tested on April 16 or 23. On average, for schools that tested on the affected days, 38% to 54% of students within a grade tested on those days, with the exception of grade 10 reading where on average over 70% of students within a school tested on those days. More notable, is the range of percentages. Some schools tested less than one percent on those days while other schools tested every student within a grade on those days.

Of the students who tested on those days, a proportion were actually disrupted by abnormal latencies and restarts. Since all students included in the disrupted sample are students with known disruptions, the percentages reported can be considered a lower bound estimate of the actual percent of students within a school that were disrupted. On average, for a school that tested on the disrupted days, 3% to 14% of students experienced a known disruption. The percent of students within a school with known disruptions ranged from 0% to 97%.

[2] Schools with no known disruptions may or may not have had disrupted students.

Table 11. Average Percent of Students Tested on April 16 and 23 for All Schools that Tested on April 16 and 23

	Average % Tested on April 16 or April 23 within a School				
Grade	*n*	Mean	*SD*	Min	Max
Reading					
3	194	44.96%	33.63%	0.68%	100%
4	198	45.05%	35.28%	0.60%	100%
5	176	41.38%	35.40%	0.43%	100%
6	158	48.26%	39.42%	0.34%	100%
7	134	52.27%	38.91%	0.41%	100%
8	130	54.83%	38.13%	0.25%	100%
10	201	70.70%	29.77%	0.40%	100%
Math					
3	284	44.29%	30.44%	0.63%	100%
4	242	38.19%	30.61%	0.49%	100%
5	224	42.52%	32.31%	0.70%	100%
6	146	47.54%	35.68%	0.27%	100%
7	127	44.81%	37.13%	0.25%	100%
8	129	45.22%	36.17%	0.35%	100%

School-Level Score Differences

To evaluate the impact of the disruption on school-level means it is informative to consider possible ways that the school-level means could be computed, taking into consideration the disruptions. We do not have the ability to undo the disruptions and we do not have the ability to know what a student would have scored if there were no disruptions. But, we can consider alternative ways of computing school means and evaluate the differences in those means compared to using students' observed scores. We computed school-level means using three alternatives.

First, the students who were identified in the disruption sample were excluded from the school-level means. For some schools, this effectively excluded less than one percent of students and for others it excluded more than 90% of students. Again, we did not exclude all students who may have been disrupted, but we excluded a concentrated set of students who were. Second, all students who tested on April 16 or 23 were excluded from the school-level mean calculation. This approach takes into consideration that we do not have complete data to determine whether students were directly or indirectly affected on the two days, so we removed all students potentially affected. Finally, we computed school-level means using the predicted score from the student-level analyses described above for students who were disrupted. It is important to understand that differences in the observed and predicted scores do not *only* capture the disruption effect. There are other uncertainties in test scores (e.g., some students may not do as well as they should because they didn't eat breakfast or they feel sick) that can yield differences between observed and predicted scores. If school-level means differ when using the predicted scores, we cannot attribute the difference to the disruption. However if the scores are not different then evidence is provided that the disruptions did not adversely impact school-level means.

Tables 12 and 13 present the results comparing the school-level means using observed scores and school-level means removing students who tested on April 16 and 23 (Table 12), removing

students who were identified in the disruption sample (Table 13), and using the predicted score in place of the observed score for students in the disrupted sample (Table 13). The results suggest that there are small to moderate differences depending on the alternative approach for calculating school means. The largest differences in school-level means were seen when excluding all students who tested on April 16 and 23. The Cohen's *d*'s ranged from .01 to .26. When these students were excluded, the school-level means dropped on average by .09 theta points and the differences ranged from .01 to .19. This suggests removing all students who tested on April 16 or 23 adversely affects school-level scores.

The school-level mean differences, after removing students in the disrupted sample, were small for most grades. The Cohen's *d* ranged from -.002 to .22, where the means increased slightly when the disrupted students were removed. The largest difference was in grade 10 reading, where on average 35% of students' scores were excluded from the school mean. Finally, when the predicted scores were substituted for the observed scores the differences in school-level means varied by grade and subject. For reading grades 3, 5, and 8 and math grades 4, 5, and 7 the school means decreased. For the remaining grades the means increased. The effect size for all of the grades and subjects were small ranging from .002 to .09.

We also examined the predictability of the 2013 school-level means. First, using multiple regression, we examined the relationship between 2012 school-level means and the 2013 school-level means for all schools that tested online. Next, we included the percent of students disrupted into the regression equation and examined whether the predictability of 2013 scores changed. If the disruption had a negative impact on school-level means, we would expect to see a significant negative relationship between 2013 school-level scores and percent of students disrupted. Additionally, if the disruption impacted school means we would expect to see an increase in the predictability of 2013 scores when disruption is considered. Lastly, we included the interaction term between the 2012 school-level mean and the percent of students disrupted. If disruption had an impact, the predictability of 2013 scores from 2012 scores was expected to be weaker for schools with a large percentage of disrupted students.

The regression results are presented in Table 14. The percent of disruption was a non-significant predictor of 2013 scores and the incremental variance accounted for when adding disruption to the model was small. Additionally, the interaction term was non-significant for all grades in reading and math. The predictability of 2013 scores was not improved by accounting for the percentage of students in a school with known disruption. This provides evidence that the disruption did not impact 2013 school scores.

Table 12. Mean Scores for Schools that Tested on April 16 or 23

	Avg. School Scores for All Schools that Tested on April 16 or 23			Avg. School Scores Removing Schools that Tested Everyone on April 16 or 23			Avg. School Scores Removing Students Who Tested on April 16 and 23					
Grade	*n*	Mean θ	*SD*	*n*	Mean θ	*SD*	*n*	Mean θ	*SD*	Avg. % of Students Removed	θ Difference	*d*[a]
Reading												
3	195	-0.034	0.543	179	0.007	0.498	179	-0.001	0.580	40%	0.008	0.015
4	198	-0.058	0.560	181	-0.033	0.483	181	-0.083	0.570	40%	0.050	0.095
5	177	-0.110	0.592	165	-0.050	0.485	165	-0.114	0.657	37%	0.064	0.112
6	159	-0.198	0.668	142	-0.103	0.567	142	-0.190	0.731	42%	0.087	0.135
7	134	-0.237	0.650	117	-0.200	0.630	117	-0.332	0.754	45%	0.132	0.192
8	130	-0.274	0.660	119	-0.221	0.613	119	-0.334	0.773	51%	0.113	0.163
10	203	-0.337	0.670	174	-0.305	0.625	174	-0.499	0.870	65%	0.194	0.260
Math												
3	284	-0.129	0.599	272	-0.118	0.601	272	-0.193	0.715	42%	0.075	0.113
4	242	-0.095	0.621	231	-0.098	0.628	231	-0.168	0.747	35%	0.070	0.101
5	224	0.000	0.570	212	0.001	0.564	212	-0.077	0.708	39%	0.078	0.122
6	146	-0.089	0.624	135	-0.048	0.557	135	-0.168	0.663	43%	0.120	0.196
7	127	-0.171	0.684	118	-0.146	0.688	118	-0.277	0.825	41%	0.131	0.173
8	129	-0.261	0.702	117	-0.269	0.720	117	-0.358	0.844	40%	0.089	0.113

[a]Cohen's *d* comparing the average school scores removing schools that tested everyone on April 16 or 23 and the average school scores removing students who tested on April 16 and 23.

Table 13. Mean Scores for Schools with Students in the Disrupted Sample

		Avg. School Scores for Schools with Students in the Disrupted Sample		Avg. School Scores Replacing Observed Scores with Predicted Scores for Disrupted Students			Avg. School Scores Removing Students with Disruptions		Avg. % of Students Removed	
Grade	*n*	Mean θ	*SD*	Mean θ	*SD*	*d*	Mean θ	*SD*		*d*
Reading										
3	92	0.016	0.456	0.054	0.391	-0.089	0.017	0.440	14%	-0.002
4	93	0.062	0.401	0.053	0.389	0.023	0.046	0.404	17%	0.040
5	67	0.053	0.419	0.058	0.400	-0.013	0.039	0.443	16%	0.031
6	57	0.080	0.339	0.074	0.337	0.020	0.071	0.365	14%	0.028
7	49	0.088	0.262	0.077	0.262	0.041	0.066	0.309	18%	0.078
8	56	0.039	0.304	0.047	0.300	-0.028	0.026	0.351	18%	0.041
10	81	0.043	0.365	0.034	0.335	0.025	-0.047	0.458	35%	0.219
Math										
3	133	-0.055	0.569	-0.058	0.564	0.006	-0.063	0.586	13%	0.013
4	88	-0.045	0.493	-0.033	0.489	-0.025	-0.032	0.494	11%	-0.026
5	81	0.077	0.453	0.084	0.446	-0.014	0.077	0.448	12%	0.015
6	59	0.066	0.424	0.063	0.419	0.007	0.060	0.459	10%	0.014
7	56	0.035	0.507	0.036	0.503	-0.003	0.030	0.512	12%	0.009
8	56	-0.027	0.441	-0.028	0.435	0.002	-0.030	0.455	7%	0.008

Table 14. Incremental Validity Estimates for Schools with Students in the Disrupted Sample

Grade	*n*	R^2 2012 School Scores Only	R^2 2012 School Scores + % Disrupted	β % Disrupted	ΔR^2	R^2 + Interaction (2012 School Scores x % Disrupted)	β Interaction Term	ΔR^2
Reading								
3	397	0.617	0.618	0.198	0.001	0.618	-0.106	0.000
4	397	0.555	0.559	0.294	0.004	0.560	0.213	0.000
5	382	0.442	0.443	0.131	0.001	0.444	0.454	0.001
6	337	0.447	0.449	0.317	0.003	0.449	-0.117	0.000
7	286	0.475	0.478	0.340	0.003	0.482	-0.663	0.004
8	295	0.476	0.478	0.300	0.002	0.478	0.061	0.000
10	320	0.631	0.634	0.183	0.003	0.637	-0.430	0.003
Math								
3	791	0.765	0.765	-0.095	0.000	0.766	0.564	0.001
4	805	0.763	0.763	0.086	0.000	0.763	-0.184	0.000
5	770	0.734	0.734	0.219	0.000	0.735	-0.720	0.001
6	588	0.727	0.727	0.228	-0.001	0.728	0.435	0.001
7	517	0.743	0.742	0.116	-0.001	0.744	0.283	0.001
8	528	0.732	0.732	0.344	0.000	0.733	-1.584	0.001

Note. Regression parameters for % Disrupted and the Interaction Term were all non-significant.

School-Level Classification Differences

Finally, we examined the impact of disruptions on the percent of students classified as proficient (Meets Expectations) for schools by grade. We calculated percent proficient using the observed scores and replacing observed scores with predicted scores for students with known disruptions. Again, it is important to interpret these results with the understanding that differences in these estimates are not *only* a result of disruptions. The predicted scores account for the general uncertainties in the estimation of student's ability. Therefore, differences cannot be attributed solely to the disruption. However if the percentage of proficient students are not different, then evidence is provided that the disruptions did not adversely impact classification. Table 15 presents the results. The results show that, on average, schools' classifications were not adversely impacted by student disruptions. The differences in the means are small with Cohen's *d* estimates ranging from -.076 to .057.

Table 15. Difference in School Classification for Schools with Students in the Disrupted Sample

		Observed % Proficient		Predicted % Proficient		
Grade	*n*	Mean	*SD*	Mean	*SD*	*d*
Reading						
3	89	58.2%	16.4%	59.5%	16.0%	-0.076
4	89	55.9%	15.5%	56.0%	16.0%	-0.010
5	64	64.8%	13.9%	65.8%	13.2%	-0.075
6	55	59.8%	13.6%	60.5%	13.3%	-0.054
7	48	56.0%	10.9%	55.4%	11.3%	0.057
8	54	54.5%	12.1%	55.1%	12.1%	-0.053
10	76	65.0%	13.2%	65.5%	13.6%	-0.043
Math						
3	128	70.0%	19.7%	70.4%	20.3%	-0.016
4	87	67.2%	18.1%	67.8%	17.9%	-0.035
5	80	56.7%	18.1%	56.9%	18.1%	-0.006
6	58	54.2%	16.4%	53.9%	16.4%	0.019
7	56	52.6%	17.9%	52.6%	17.9%	0.000
8	54	52.3%	17.0%	52.6%	16.6%	-0.020

Note. School samples less than 20 were excluded.

School-Level Summary

Overall, the impact of disruption on school-level scores appears to be very small. Taking into consideration the students with known disruptions, generally school-level scores were not affected. School-level scores using predicted scores for students with known disruptions and excluding students with known disruptions from school-level scores were not largely different from the observed school-level scores. Additionally, the percent of students disrupted did not impact the prediction of 2013 school performance. If the state were to take a broader approach to adjusting scores and excluded all students who tested on April 16 or 23, then schools would see lower overall test scores. Based on the data available, school-level analyses suggest that there may be a small impact on school-level means for schools that experienced disruption, but the direction of the impact is not consistent and adjusting school-level scores based on this data would be beneficial for some schools and detrimental to others. Finally, the regression analysis

suggests that the relationship between 2012 school means and 2013 school means was not altered by the known prevalence of disruption within schools.

Conclusions

This report provides a statistical investigation of computer disruptions on student and school-level scores for the Minnesota Comprehensive Assessment. A number of analyses were conducted at both the student-level and the school-level to investigate whether computer disruptions impacted scores. While there is some evidence to suggest that there were effects from the disruption, nothing emerged in a systematic way across grades, subjects, or methodologies. That is, there is no statistical evidence to suggest that the disruption, on average, adversely impacted students and schools that tested on April 16 or 23. This report is intended to inform the MDE and the Minnesota TAC of the statistical impact of computer disruption and to be used as a piece of evidence in considering whether policy actions are appropriate.

References

Austin, P. C. (2009). Some methods of propensity-score matching had superior performance to others: Results of an empirical investigation and Monte Carlo simulation. *Biometrical Journal, 51,* 171-184. doi: 10.1002/bimj.200810488

Connelly, B. S., Sackett, P.R., & Waters, S. D. (2013). Balancing treatment and control groups in quasi-experiments: An Introduction to propensity scoring. *Personnel Psychology, 66,* 407-442. doi: 10.1111/peps.12020

Appendix A

Mean Differences between Non-Disrupted and Disrupted Samples before and after Propensity Matching

Table A.1. Mean Covariate Differences for Reading Grade 3

	Before Matching				
	Non-Disrupted		Disrupted		
	Mean	*SD*	Mean	*SD*	*d*
Predicted Probability	0.072	0.020	0.077	0.018	**-0.281**
2013 Math Theta	0.030	1.027	0.066	0.988	-0.035
School Free/Reduced Lunch	0.402	0.205	0.350	0.164	**0.280**
School Achievement	0.071	0.428	0.134	0.356	**-0.160**
Indian	0.036	0.185	0.027	0.162	0.049
Asian	0.042	0.200	0.049	0.215	-0.034
Hispanic	0.075	0.263	0.076	0.265	-0.003
Black	0.063	0.243	0.042	0.201	0.094
White	0.771	0.420	0.805	0.396	-0.083
Male	0.502	0.500	0.532	0.499	-0.061
Free/ Reduce Lunch	0.400	0.490	0.370	0.483	0.060
Limited English Proficiency	0.068	0.252	0.063	0.243	0.021
	After Matching				
	Non-Disrupted		Disrupted		
	Mean	*SD*	Mean	*SD*	*d*
Predicted Probability	0.077	0.018	0.077	0.018	0.000
2013 Math Theta	0.043	1.013	0.079	0.998	-0.036
School Free/Reduced Lunch	0.345	0.171	0.346	0.160	-0.001
School Achievement	0.144	0.372	0.139	0.346	0.013
Indian	0.022	0.148	0.025	0.157	-0.020
Asian	0.055	0.228	0.053	0.224	0.009
Hispanic	0.092	0.290	0.079	0.270	0.047
Black	0.046	0.209	0.046	0.209	0.000
White	0.785	0.411	0.797	0.402	-0.030
Male	0.518	0.500	0.529	0.499	-0.022
Free/Reduced Lunch	0.354	0.478	0.363	0.481	-0.019
Limited English Proficiency	0.078	0.268	0.068	0.252	0.039

Note. Bolded values indicate Cohen's *d* values greater than .10 or -.10. Sample sizes before matching ranges from for 12,772 – 14,482 for the non-disrupted group and 990 – 1,110 for the disrupted group. Sample size is 986 after matching.

Table A.2. Mean Covariate Differences for Reading Grade 4

	Before Matching				
	Non-Disrupted		Disrupted		
	Mean	*SD*	Mean	*SD*	*d*
Predicted Probability	0.063	0.013	0.066	0.017	**-0.245**
2012 Math Theta	0.079	0.954	0.122	0.934	-0.046
2012 Reading Theta	0.038	1.089	0.077	1.052	-0.037
School Free/Reduced Lunch	0.372	0.202	0.372	0.196	-0.003
School Achievement	0.077	0.356	0.128	0.397	**-0.136**
Indian	0.032	0.176	0.034	0.181	-0.010
Asian	0.047	0.211	0.049	0.216	-0.010
Hispanic	0.065	0.246	0.049	0.216	0.069
Black	0.057	0.232	0.046	0.210	0.050
White	0.763	0.425	0.787	0.410	-0.056
Male	0.500	0.500	0.500	0.500	-0.002
Free/Reduced Lunch	0.388	0.487	0.363	0.481	0.051
Limited English Proficiency	0.069	0.253	0.054	0.227	0.060
	After Matching				
	Non-Disrupted		Disrupted		
	Mean	*SD*	Mean	*SD*	*d*
Predicted Probability	0.066	0.017	0.066	0.017	-0.004
2012 Math Theta	0.115	0.906	0.135	0.922	-0.022
2012 Reading Theta	0.133	1.071	0.102	1.057	0.029
School Free/Reduced Lunch	0.383	0.212	0.371	0.193	0.059
School Achievement	0.120	0.409	0.127	0.398	-0.017
Indian	0.049	0.216	0.036	0.186	0.066
Asian	0.048	0.214	0.050	0.218	-0.009
Hispanic	0.071	0.256	0.047	0.212	**0.101**
Black	0.048	0.214	0.045	0.207	0.015
White	0.784	0.412	0.822	0.383	-0.095
Male	0.484	0.500	0.501	0.500	-0.033
Free/Reduced Lunch	0.386	0.487	0.356	0.479	0.061
Limited English Proficiency	0.056	0.231	0.050	0.218	0.027

Note. Bolded values indicate Cohen's *d* values greater than .10 or -.10. Sample sizes before matching ranges from for 14,917 – 16,307 for the non-disrupted group and 989-1,125 for the disrupted group. Sample size is 977 after matching.

Table A.3. Mean Covariate Differences for Reading Grade 5

	Before Matching				
	Non-Disrupted		Disrupted		
	Mean	*SD*	Mean	*SD*	*d*
Predicted Probability	0.055	0.013	0.058	0.012	**-0.241**
2012 Math Theta	0.147	0.913	0.205	0.919	-0.063
2012 Reading Theta	0.111	1.026	0.134	0.995	-0.022
School Free/Reduced Lunch	0.360	0.189	0.372	0.190	-0.062
School Achievement	0.157	0.325	0.198	0.336	**-0.124**
Indian	0.023	0.151	0.027	0.162	-0.024
Asian	0.048	0.214	0.042	0.201	0.029
Hispanic	0.065	0.247	0.048	0.213	0.077
Black	0.051	0.221	0.062	0.240	-0.044
White	0.766	0.423	0.787	0.409	-0.051
Male	0.495	0.500	0.526	0.500	-0.062
Free/Reduced Lunch	0.368	0.482	0.387	0.487	-0.039
Limited English Proficiency	0.065	0.247	0.051	0.221	0.058
	After Matching				
	Non-Disrupted		Disrupted		
	Mean	*SD*	Mean	*SD*	*d*
Predicted Probability	0.058	0.012	0.058	0.012	0.000
2012 Math Theta	0.259	0.862	0.262	0.893	-0.003
2012 Reading Theta	0.221	0.958	0.191	0.961	0.031
School Free/Reduced Lunch	0.361	0.183	0.363	0.183	-0.014
School Achievement	0.207	0.312	0.198	0.337	0.028
Indian	0.017	0.130	0.023	0.151	-0.044
Asian	0.048	0.214	0.043	0.203	0.024
Hispanic	0.046	0.209	0.046	0.209	0.000
Black	0.049	0.216	0.057	0.231	-0.033
White	0.840	0.367	0.831	0.375	0.023
Male	0.525	0.500	0.529	0.499	-0.007
Free/Reduced Lunch	0.344	0.475	0.358	0.480	-0.028
Limited English Proficiency	0.041	0.197	0.049	0.216	-0.042

Note. Bolded values indicate Cohen's *d* values greater than .10 or -.10. Sample sizes before matching ranges from for 14,121 – 15,727 for the non-disrupted group and 825-926 for the disrupted group. Sample size is 813 after matching.

Table A.4. Mean Covariate Differences for Reading Grade 6

	Before Matching				
	Non-Disrupted		Disrupted		
	Mean	*SD*	Mean	*SD*	*d*
Predicted Probability	0.048	0.013	0.052	0.015	**-0.305**
2012 Math Theta	0.202	0.968	0.259	0.964	-0.059
2012 Reading Theta	0.144	1.040	0.180	1.034	-0.035
School Free/Reduced Lunch	0.360	0.186	0.351	0.168	0.046
School Achievement	0.201	0.341	0.274	0.380	**-0.201**
Indian	0.022	0.146	0.012	0.109	0.076
Asian	0.040	0.196	0.032	0.175	0.046
Hispanic	0.063	0.243	0.062	0.241	0.005
Black	0.055	0.228	0.045	0.207	0.046
White	0.778	0.415	0.804	0.397	-0.062
Male	0.503	0.500	0.505	0.500	-0.005
Free/Reduced Lunch	0.371	0.483	0.332	0.471	0.083
Limited English Proficiency	0.051	0.220	0.038	0.192	0.063
	After Matching				
	Non-Disrupted		Disrupted		
	Mean	*SD*	Mean	*SD*	*d*
Predicted Probability	0.052	0.015	0.052	0.015	0.002
2012 Math Theta	0.337	0.970	0.337	0.930	0.000
2012 Reading Theta	0.247	0.995	0.246	1.029	0.001
School Free/Reduced Lunch	0.352	0.183	0.343	0.163	0.053
School Achievement	0.275	0.379	0.280	0.379	-0.013
Indian	0.003	0.054	0.006	0.076	-0.045
Asian	0.033	0.180	0.033	0.180	0.000
Hispanic	0.064	0.245	0.061	0.239	0.012
Black	0.042	0.201	0.046	0.211	-0.021
White	0.858	0.350	0.853	0.354	0.012
Male	0.531	0.499	0.508	0.500	0.046
Free/Reduced Lunch	0.327	0.469	0.305	0.461	0.047
Limited English Proficiency	0.044	0.204	0.030	0.172	0.069

Note. Bolded values indicate Cohen's *d* values greater than .10 or -.10. Sample sizes before matching ranges from for 13,921 – 15,553 for the non-disrupted group and 701-825 for the disrupted group. Sample size is 689 after matching.

Table A.5. Mean Covariate Differences for Reading Grade 7

	Before Matching				
	Non-Disrupted		Disrupted		
	Mean	*SD*	Mean	*SD*	*d*
Predicted Probability	0.055	0.014	0.059	0.013	-0.283
2012 Math Theta	0.233	0.904	0.226	0.853	0.008
2012 Reading Theta	0.225	1.107	0.245	1.066	-0.019
School Free/Reduced Lunch	0.331	0.169	0.344	0.148	-0.079
School Achievement	0.266	0.312	0.238	0.294	0.092
Indian	0.022	0.148	0.014	0.118	0.062
Asian	0.044	0.205	0.026	0.160	0.098
Hispanic	0.050	0.217	0.029	0.169	**0.106**
Black	0.047	0.213	0.026	0.160	**0.114**
White	0.787	0.409	0.861	0.346	**-0.197**
Male	0.487	0.500	0.522	0.500	-0.069
Free/Reduced Lunch	0.347	0.476	0.344	0.475	0.006
Limited English Proficiency	0.043	0.202	0.016	0.125	**0.164**
	After Matching				
	Non-Disrupted		Disrupted		
	Mean	*SD*	Mean	*SD*	*d*
Predicted Probability	0.059	0.013	0.059	0.013	0.000
2012 Math Theta	0.273	0.899	0.277	0.825	-0.004
2012 Reading Theta	0.286	1.110	0.310	1.047	-0.022
School Free/Reduced Lunch	0.351	0.170	0.340	0.146	0.069
School Achievement	0.227	0.333	0.239	0.295	-0.037
Indian	0.015	0.123	0.013	0.113	0.020
Asian	0.034	0.182	0.027	0.163	0.041
Hispanic	0.025	0.156	0.025	0.156	0.000
Black	0.028	0.166	0.024	0.152	0.030
White	0.897	0.304	0.911	0.285	-0.048
Male	0.541	0.499	0.530	0.499	0.024
Free/Reduced Lunch	0.361	0.481	0.328	0.470	0.070
Limited English Proficiency	0.019	0.136	0.015	0.123	0.027

Note. Bolded values indicate Cohen's *d* values greater than .10 or -.10. Sample sizes before matching ranges from for 14,522 – 16,152 for the non-disrupted group and 853-941 for the disrupted group. Sample size is 844 after matching.

Table A.6. Mean Covariate Differences for Reading Grade 8

	Before Matching				
	Non-Disrupted		Disrupted		
	Mean	*SD*	Mean	*SD*	*d*
Predicted Probability	0.067	0.010	0.069	0.010	**-0.164**
2012 Math Theta	0.200	0.939	0.201	0.862	-0.001
2012 Reading Theta	0.226	0.993	0.223	0.913	0.003
School Free/Reduced Lunch	0.319	0.162	0.310	0.161	0.058
School Achievement	0.259	0.292	0.256	0.299	0.010
Indian	0.023	0.150	0.021	0.144	0.014
Asian	0.034	0.180	0.046	0.209	-0.062
Hispanic	0.054	0.225	0.049	0.216	0.020
Black	0.045	0.207	0.040	0.197	0.021
White	0.813	0.390	0.826	0.379	-0.033
Male	0.491	0.500	0.522	0.500	-0.063
Free/Reduced Lunch	0.331	0.471	0.340	0.474	-0.019
Limited English Proficiency	0.036	0.185	0.028	0.164	0.045
	After Matching				
	Non-Disrupted		Disrupted		
	Mean	*SD*	Mean	*SD*	*d*
Predicted Probability	0.069	0.010	0.069	0.010	-0.001
2012 Math Theta	0.311	0.888	0.279	0.802	0.038
2012 Reading Theta	0.315	0.963	0.286	0.875	0.032
School Free/Reduced Lunch	0.311	0.146	0.307	0.158	0.027
School Achievement	0.265	0.272	0.258	0.300	0.026
Indian	0.006	0.080	0.011	0.103	-0.047
Asian	0.053	0.223	0.043	0.203	0.045
Hispanic	0.046	0.210	0.047	0.212	-0.005
Black	0.050	0.219	0.040	0.195	0.052
White	0.845	0.362	0.860	0.348	-0.042
Male	0.523	0.500	0.518	0.500	0.011
Free/Reduced Lunch	0.313	0.464	0.312	0.464	0.002
Limited English Proficiency	0.024	0.152	0.023	0.148	0.007

Note. Bolded values indicate Cohen's *d* values greater than .10 or -.10. Sample sizes before matching ranges from for 13,024 – 14,561 for the non-disrupted group and 945 – 1,139 for the disrupted group. Sample size is 933 after matching.

Table A.7. Mean Covariate Differences for Reading Grade 10

	Before Matching				
	Non-Disrupted		Disrupted		
	Mean	*SD*	Mean	*SD*	*d*
Predicted Probability	0.217	0.067	0.241	0.059	**-0.393**
2011 Reading Scale Scores	854.738	13.964	855.251	13.739	-0.037
2011 Math Scale Scores	849.252	13.776	850.156	12.959	-0.068
School Free/Reduced Lunch	0.377	0.202	0.311	0.156	**0.368**
School Achievement	852.651	4.651	853.384	4.465	**-0.161**
Indian	0.037	0.189	0.018	0.131	**0.122**
Asian	0.053	0.224	0.036	0.187	0.083
Hispanic	0.067	0.251	0.064	0.245	0.013
Black	0.089	0.285	0.045	0.207	**0.181**
White	0.748	0.434	0.837	0.369	**-0.220**
Male	0.512	0.500	0.501	0.500	0.021
Free/Reduced Lunch	0.383	0.486	0.317	0.466	**0.138**
Limited English Proficiency	0.045	0.207	0.028	0.165	0.089
	After Matching				
	Non-Disrupted		Disrupted		
	Mean	*SD*	Mean	*SD*	*d*
Predicted Probability	0.241	0.057	0.241	0.059	-0.015
2011 Reading Scale Scores	855.403	13.361	855.913	13.546	-0.038
2011 Math Scale Scores	850.478	13.202	850.906	12.675	-0.033
School Free/Reduced Lunch	0.310	0.147	0.304	0.148	0.041
School Achievement	853.273	3.973	853.390	4.373	-0.028
Indian	0.009	0.096	0.011	0.107	-0.022
Asian	0.035	0.185	0.037	0.188	-0.007
Hispanic	0.064	0.244	0.058	0.234	0.022
Black	0.027	0.163	0.033	0.179	-0.034
White	0.864	0.342	0.860	0.347	0.012
Male	0.485	0.500	0.494	0.500	-0.018
Free/Reduced Lunch	0.325	0.468	0.288	0.453	0.080
Limited English Proficiency	0.021	0.143	0.021	0.143	0.000

Note. Bolded values indicate Cohen's *d* values greater than .10 or -.10. Sample sizes before matching ranges from for 8,560 – 9,638 for the non-disrupted group and 989-1,125 for the disrupted group. Sample size is 2,265 after matching.

Table A.8. Mean Covariate Differences for Math Grade 3

	Before Matching				
	Non-Disrupted		Disrupted		
	Mean	*SD*	Mean	*SD*	*d*
Predicted Probability	0.033	0.007	0.034	0.007	**-0.218**
2013 Reading Theta	0.022	1.086	-0.017	1.031	0.037
School Free/Reduced Lunch	0.403	0.235	0.408	0.261	-0.019
School Achievement	0.068	0.449	0.032	0.446	0.081
Indian	0.027	0.161	0.022	0.147	0.030
Asian	0.075	0.264	0.076	0.265	-0.002
Hispanic	0.081	0.274	0.084	0.277	-0.008
Black	0.115	0.319	0.136	0.343	-0.064
White	0.696	0.460	0.672	0.470	0.053
Male	0.508	0.500	0.495	0.500	0.025
Free/Reduced Lunch	0.408	0.491	0.388	0.487	0.040
Limited English Proficiency	0.109	0.311	0.121	0.326	-0.038
	After Matching				
	Non-Disrupted		Disrupted		
	Mean	*SD*	Mean	*SD*	*d*
Predicted Probability	0.034	0.007	0.034	0.007	0.000
2013 Reading Theta	0.004	1.072	0.000	1.033	0.004
School Free/Reduced Lunch	0.395	0.241	0.398	0.261	-0.011
School Achievement	0.039	0.456	0.031	0.447	0.017
Indian	0.012	0.107	0.015	0.121	-0.027
Asian	0.084	0.277	0.088	0.283	-0.014
Hispanic	0.075	0.264	0.072	0.259	0.012
Black	0.152	0.359	0.144	0.351	0.022
White	0.678	0.468	0.681	0.466	-0.007
Male	0.493	0.500	0.498	0.500	-0.009
Free/Reduced Lunch	0.391	0.488	0.384	0.487	0.014
Limited English Proficiency	0.124	0.330	0.124	0.330	0.000

Note. Bolded values indicate Cohen's *d* values greater than .10 or -.10. Sample sizes before matching ranges from for 38,722 – 43,161 for the non-disrupted group and 1,367 – 1,492 for the disrupted group. Sample size is 1,287 after matching.

Table A.9. Mean Covariate Differences for Math Grade 4

	Before Matching				
	Non-Disrupted		Disrupted		
	Mean	*SD*	Mean	*SD*	*d*
Predicted Probability	0.019	0.007	0.022	0.009	-0.357
2012 Math Theta	0.068	1.008	-0.069	1.023	**0.135**
2012 Reading Theta	0.050	1.099	-0.126	1.071	**0.163**
School Free/Reduced Lunch	0.391	0.232	0.407	0.219	-0.073
School Achievement	0.065	0.413	-0.022	0.403	**0.212**
Indian	1.590	0.717	1.592	0.672	-0.003
Asian	0.024	0.153	0.026	0.159	-0.013
Hispanic	0.072	0.258	0.102	0.302	**-0.107**
Black	0.071	0.258	0.056	0.230	0.062
White	0.101	0.301	0.107	0.309	-0.020
Male	0.695	0.460	0.675	0.469	0.043
Free/Reduced Lunch	0.503	0.500	0.542	0.499	-0.077
Limited English Proficiency	0.396	0.489	0.439	0.497	-0.087
	After Matching				
	Non-Disrupted		Disrupted		
	Mean	*SD*	Mean	*SD*	*d*
Predicted Probability	0.022	0.009	0.022	0.009	0.000
2012 Math Theta	-0.040	0.999	-0.061	1.024	0.021
2012 Reading Theta	-0.133	1.047	-0.128	1.076	-0.005
School Free/Reduced Lunch	0.395	0.239	0.402	0.220	-0.031
School Achievement	-0.012	0.431	-0.017	0.398	0.012
Indian	1.581	0.667	1.597	0.668	-0.024
Asian	0.022	0.147	0.024	0.155	-0.016
Hispanic	0.108	0.310	0.109	0.312	-0.004
Black	0.051	0.221	0.061	0.240	-0.042
White	0.106	0.309	0.109	0.312	-0.008
Male	0.712	0.453	0.696	0.460	0.035
Free/Reduced Lunch	0.512	0.500	0.540	0.499	-0.056
Limited English Proficiency	0.412	0.493	0.425	0.495	-0.025

Note. Bolded values indicate Cohen's *d* values greater than .10 or -.10. Sample sizes before matching ranges from for 42,337 – 44,643 for the non-disrupted group and 828 – 925 for the disrupted group. Sample size is 817 after matching.

Table A.10. Mean Covariate Differences for Math Grade 5

	Before Matching				
	Non-Disrupted		Disrupted		
	Mean	*SD*	Mean	*SD*	*d*
Predicted Probability	0.017	0.005	0.019	0.006	-0.281
2012 Math Theta	0.121	0.962	0.063	1.019	0.059
2012 Reading Theta	0.083	1.055	0.029	1.105	0.050
School Free/Reduced Lunch	0.383	0.227	0.412	0.230	**-0.127**
School Achievement	0.108	0.405	0.096	0.415	0.028
Indian	1.612	0.722	1.658	0.790	-0.060
Asian	0.021	0.144	0.037	0.188	-0.095
Hispanic	0.072	0.259	0.068	0.253	0.014
Black	0.071	0.256	0.089	0.286	-0.069
White	0.095	0.293	0.093	0.291	0.005
Male	0.701	0.458	0.683	0.466	0.040
Free/Reduced Lunch	0.507	0.500	0.532	0.499	-0.050
Limited English Proficiency	0.386	0.487	0.402	0.491	-0.033
	After Matching				
	Non-Disrupted		Disrupted		
	Mean	*SD*	Mean	*SD*	*d*
Predicted Probability	0.019	0.006	0.019	0.006	0.000
2012 Math Theta	0.121	0.972	0.086	1.005	0.036
2012 Reading Theta	0.076	1.076	0.054	1.097	0.020
School Free/Reduced Lunch	0.413	0.235	0.406	0.225	0.027
School Achievement	0.082	0.413	0.097	0.415	-0.035
Indian	1.684	0.769	1.663	0.787	0.028
Asian	0.046	0.209	0.037	0.189	0.043
Hispanic	0.069	0.253	0.070	0.255	-0.006
Black	0.107	0.310	0.096	0.294	0.038
White	0.071	0.258	0.089	0.284	-0.063
Male	0.707	0.455	0.709	0.455	-0.003
Free/Reduced Lunch	0.521	0.500	0.534	0.499	-0.026
Limited English Proficiency	0.356	0.479	0.384	0.487	-0.059

Note. Bolded values indicate Cohen's *d* values greater than .10 or -.10. Sample sizes before matching ranges from for 40,287 – 42,599 for the non-disrupted group and 708 – 760 for the disrupted group. Sample size is 700 after matching.

Table A.11. Mean Covariate Differences for Math Grade 6

	Before Matching				
	Non-Disrupted		Disrupted		
	Mean	*SD*	Mean	*SD*	*d*
Predicted Probability	0.016	0.004	0.017	0.005	-0.271
2012 Math Theta	0.161	0.996	0.218	1.034	-0.057
2012 Reading Theta	0.132	1.051	0.152	1.066	-0.019
School Free/Reduced Lunch	0.397	0.226	0.433	0.217	**-0.160**
School Achievement	0.156	0.396	0.111	0.390	**0.115**
Indian	1.722	0.754	1.743	0.782	-0.027
Asian	0.024	0.154	0.044	0.206	**-0.112**
Hispanic	0.068	0.251	0.087	0.282	-0.073
Black	0.071	0.257	0.047	0.211	**0.103**
White	0.101	0.302	0.101	0.302	-0.001
Male	0.697	0.460	0.654	0.476	0.092
Free/Reduced Lunch	0.510	0.500	0.519	0.500	-0.018
Limited English Proficiency	0.395	0.489	0.412	0.493	-0.035
	After Matching				
	Non-Disrupted		Disrupted		
	Mean	*SD*	Mean	*SD*	*d*
Predicted Probability	0.017	0.005	0.017	0.005	0.000
2012 Math Theta	0.250	0.973	0.253	1.025	-0.003
2012 Reading Theta	0.199	1.072	0.186	1.058	0.013
School Free/Reduced Lunch	0.437	0.244	0.436	0.219	0.007
School Achievement	0.132	0.416	0.112	0.391	0.050
Indian	1.739	0.783	1.721	0.769	0.024
Asian	0.040	0.196	0.037	0.189	0.016
Hispanic	0.091	0.288	0.100	0.301	-0.031
Black	0.060	0.238	0.051	0.220	0.040
White	0.108	0.311	0.108	0.311	0.000
Male	0.701	0.458	0.704	0.457	-0.007
Free/Reduced Lunch	0.520	0.500	0.520	0.500	0.000
Limited English Proficiency	0.381	0.486	0.401	0.491	-0.041

Note. Bolded values indicate Cohen's *d* values greater than .10 or -.10. Sample sizes before matching ranges from for 40,302 – 42,612 for the non-disrupted group and 657 – 769 for the disrupted group. Sample size is 648 after matching.

Table A.12. Mean Covariate Differences for Math Grade 7

	Before Matching				
	Non-Disrupted		Disrupted		
	Mean	*SD*	Mean	*SD*	*d*
Predicted Probability	0.017	0.006	0.020	0.007	-0.367
2012 Math Theta	0.155	0.966	0.277	0.962	**-0.126**
2012 Reading Theta	0.203	1.113	0.296	1.094	-0.085
School Free/Reduced Lunch	0.379	0.217	0.413	0.216	**-0.159**
School Achievement	0.191	0.389	0.153	0.407	0.095
Indian	1.658	0.751	1.640	0.787	0.024
Asian	0.022	0.148	0.021	0.142	0.011
Hispanic	0.071	0.257	0.065	0.246	0.026
Black	0.062	0.241	0.047	0.211	0.069
White	0.098	0.298	0.102	0.303	-0.013
Male	0.701	0.458	0.732	0.443	-0.070
Free/Reduced Lunch	0.502	0.500	0.527	0.500	-0.049
Limited English Proficiency	0.382	0.486	0.360	0.480	0.045
	After Matching				
	Non-Disrupted		Disrupted		
	Mean	*SD*	Mean	*SD*	*d*
Predicted Probability	0.020	0.007	0.020	0.007	0.000
2012 Math Theta	0.279	0.939	0.299	0.953	-0.021
2012 Reading Theta	0.290	1.099	0.315	1.092	-0.023
School Free/Reduced Lunch	0.399	0.212	0.407	0.214	-0.038
School Achievement	0.160	0.394	0.154	0.408	0.016
Indian	1.598	0.780	1.629	0.779	-0.040
Asian	0.004	0.065	0.013	0.112	-0.096
Hispanic	0.056	0.231	0.071	0.256	-0.058
Black	0.038	0.192	0.048	0.214	-0.049
White	0.121	0.327	0.110	0.313	0.035
Male	0.780	0.415	0.759	0.428	0.050
Free/Reduced Lunch	0.513	0.500	0.526	0.500	-0.025
Limited English Proficiency	0.346	0.476	0.351	0.478	-0.012

Note. Bolded values indicate Cohen's *d* values greater than .10 or -.10. Sample sizes before matching ranges from for 40,832 – 43,418 for the non-disrupted group and 713 – 773 for the disrupted group. Sample size is 709 after matching.

Table A.13. Mean Covariate Differences for Math Grade 8

	Before Matching				
	Non-Disrupted		Disrupted		
	Mean	*SD*	Mean	*SD*	*d*
Predicted Probability	0.019	0.008	0.023	0.011	-0.411
2012 Math Theta	0.131	0.994	0.068	0.963	0.065
2012 Reading Theta	0.206	1.011	0.214	0.970	-0.008
School Free/Reduced Lunch	0.362	0.205	0.360	0.200	0.010
School Achievement	0.184	0.350	0.153	0.365	0.088
Indian	1.638	0.748	1.506	0.673	**0.186**
Asian	0.020	0.141	0.013	0.115	0.054
Hispanic	0.062	0.241	0.114	0.318	**-0.187**
Black	0.062	0.241	0.054	0.225	0.037
White	0.096	0.295	0.091	0.288	0.018
Male	0.726	0.446	0.703	0.457	0.051
Free/Reduced Lunch	0.501	0.500	0.550	0.498	-0.097
Limited English Proficiency	0.363	0.481	0.373	0.484	-0.020
	After Matching				
	Non-Disrupted		Disrupted		
	Mean	*SD*	Mean	*SD*	*d*
Predicted Probability	0.023	0.011	0.023	0.011	0.000
2012 Math Theta	0.084	1.039	0.072	0.964	0.012
2012 Reading Theta	0.168	0.979	0.226	0.978	-0.060
School Free/Reduced Lunch	0.372	0.221	0.357	0.196	0.072
School Achievement	0.137	0.394	0.152	0.365	-0.038
Indian	1.487	0.644	1.472	0.639	0.023
Asian	0.011	0.103	0.011	0.103	0.000
Hispanic	0.167	0.373	0.125	0.331	**0.119**
Black	0.065	0.246	0.059	0.236	0.022
White	0.089	0.285	0.097	0.296	-0.028
Male	0.668	0.471	0.708	0.455	-0.084
Free/Reduced Lunch	0.523	0.500	0.553	0.498	-0.059
Limited English Proficiency	0.394	0.489	0.363	0.481	0.064

Note. Bolded values indicate Cohen's *d* values greater than .10 or -.10. Sample sizes before matching ranges from for 37,661 – 40,131 for the non-disrupted group and 747 – 822 for the disrupted group. Sample size is 742 after matching.