What Works Clearinghouse™

ies INSTITUTE OF EDUCATION SCIENCES

February 2015

WWC Review of the Report "The Impact of Indiana's System of Interim Assessments on Mathematics and Reading Achievement"[1]

The findings from this review do not reflect the full body of research evidence on using *Diagnostic Assessment Tools* on mathematics and reading.

What is this study about?

The study authors examined the effects of using *Diagnostic Assessment Tools (DAT)* on mathematics and reading outcomes for students in Indiana schools during the 2009–10 academic year. *DAT* consists of interim assessment tools—Wireless Generation's mCLASS for students in grades K–2 and CTB/McGraw-Hill's Acuity for students in grades 3–8—modified to align with Indiana's state assessments. The intent is for teachers to use these *DAT* interim assessment results to inform their instructional practice to meet the needs of their students.

The study is a clustered randomized controlled trial in which 59 schools serving students in grades K–8 were randomly assigned to condition: 35 to the intervention group and 24 to the delayed-treatment comparison group. This set of schools was randomly selected from a larger sample of 116 K–8 schools originally identified to be eligible for the study (and that volunteered to participate) in an effort to create a sample that would be representative of the state geographic balance.[2]

Teachers at schools in the intervention condition received the assessment tools and training on how to use them. In the comparison condition, schools did not receive the tools or the training during the study. Comparison schools were eligible to receive the intervention in future years.

The study authors assessed students' mathematics and reading achievement using the Indiana Statewide Testing for Educational Progress–Plus (ISTEP+, grades 3–8) and TerraNova (grades K–2) standardized tests. The study presented a large number of impact analyses, including full-sample analyses, grade-specific analyses, and analyses that pool subsets of grades together. The study authors conducted analyses that focus on an initially assigned sample (an intent-to-treat or ITT approach) and analyses that excluded schools that did not participate in the study for the full year (the authors refer to this as a treatment-on-treated or TOT approach). The ITT and TOT results presented in the study are based on samples that included students who joined the study classrooms after random assignment ("joiners") and students who crossed over from one study group to another ("crossovers"). The study authors also described, but did not present, quantitative results for a sample that excludes joiners and crossovers ("stayers-only"). The WWC obtained information on the stayers-only analyses from the authors. The WWC reviewed all analyses presented in the study, including those obtained from the authors. The analytic samples for the full set of grades contributing to the impact estimate included 19,167 students for mathematics (10,708 in the intervention condition and 8,459 in the comparison condition) and 19,173 students for reading (10,708 in the intervention condition and 8,465 in the comparison condition) in 57 schools.

What did the study find?

The study authors found, and the WWC confirmed, that the use of *Diagnostic Assessment Tools* did not have a statistically significant impact on general mathematics achievement or reading achievement for the full sample of students in grades K–8. However, the authors found, and the WWC confirmed, statistically significant positive effects for grades 5 and 6 in mathematics achievement and grades 3–5 in reading achievement. The authors found, and the WWC confirmed, no statistically significant impacts on mathematics achievement in grades 3 and 4 or on reading achievement in grade 6.[3]

WWC Rating

The research described in this report meets WWC group design standards without reservations

This study is a well-executed randomized controlled trial with low sample attrition. **A subset of the analyses described in the study meet WWC group design standards without reservations**. Specifically, this rating pertains to the intent-to-treat (ITT) estimates for the full (grades K–8) stayers-only student sample, which was provided in response to an author query. This rating also applies to the grade-specific ITT estimates for the stayers-only samples in grades 3–6. These stayers-only analyses are the focus of the remainder of this single study review.

The study presented a number of analyses of samples that include joiners and crossovers (under both an ITT and TOT framework). These include a full-sample analysis, grade-specific analyses, pooled results for a subset of grades (K–2, 3–6, 3–8), and results for geographic subgroups. In order for these analyses to meet WWC standards with reservations, baseline equivalence of the study groups must be established. For these analyses, there was insufficient information to determine the equivalence of the analytic samples, and therefore, none of the analyses that include joiners and crossovers meet WWC standards. The findings from these analyses are therefore not reported in this single study review.

Features of *Diagnostic Assessment Tools (DAT)*

DAT consists of two interim diagnostic assessment software packages aligned to Indiana state standards and grade-level expectations—Wireless Generation's mCLASS, which is used for students in grades K–2, and CTB/McGraw-Hill's Acuity, which is used for students in grades 3–8.

The mCLASS software package consists of diagnostic assessments in reading (mCLASS: Reading3D) and math (mCLASS: Math). For mCLASS Reading3D, students work individually with teachers to complete brief DIEBELS probes, administered via a personal digital assistant. The probes can be administered periodically at the teacher's discretion, and teachers can use the results to identify problem areas and track student progress over time. mCLASS: Math is administered via pencil and paper and later entered into a computer, allowing the teacher to view reports and run queries. The items for the mathematics assessments are linked to expectations based on Indiana state standards and the timing of the assessment window.

The Acuity software package consists of seven online multiple-choice tests that are offered in either reading or mathematics. Acuity includes four diagnostic assessments, which are administered throughout the year to identify the specific needs of students. In addition, there are three predictive assessments, which are used to predict student performance on the Indiana Statewide Testing for Educational Progress–Plus (ISTEP+).

Both assessments are accompanied by progress monitoring tools, instructional tools, and online support systems that allow teachers to assess student performance on state standards and objectives and identify students performing at different achievement levels. Teachers at intervention schools received training to use these tools and support systems and were instructed to use them to monitor student progress and adjust instruction to student needs with the ultimate goal of improving student achievement.

Appendix A: Study details

Konstantopoulos, S., Miller, S. R., & van der Ploeg, A. (2013). The impact of Indiana's system of interim assessments on mathematics and reading achievement. *Educational Evaluation and Policy Analysis, 35*(4), 481–499.

Setting

The study was conducted in K–8 grade schools in Indiana during the 2009–10 academic year.

Study sample

Study schools were randomly selected from 116 schools that volunteered and were initially identified to be eligible for the study. A total of 59 K–8 grade schools were randomly assigned to the intervention (35 schools) and comparison (24 schools) groups, with 20,428 students in mathematics and 20,436 students in reading across both groups. The final analytic sample, after attrition, included 19,167 students for mathematics (10,708 in the intervention group and 8,459 in the comparison group) and 19,173 students for reading (10,708 in the intervention group and 8,465 in the comparison group) in 57 schools. The WWC review focused on the ITT estimates for the stayers-only sample which excluded joiners and crossovers; estimates for this sample were provided in response to an author query.

Intervention group

Schools in the intervention group received *DAT*, a set of two interim assessments aligned to Indiana state standards and grade-level expectations. *DAT* is intended to help teachers monitor student learning and adjust their instruction to student needs during the school year. Teachers in the intervention group were provided with the assessment tools—Wireless Generation's mCLASS for students in grades K–2 and CTB/McGraw-Hill's Acuity for students in grades 3–8—and training on how to use them. Both assessments are accompanied by progress monitoring tools, instructional tools, and online support systems that allow teachers to assess student performance on state standards and objectives and identify students performing at a different achievement levels.

Comparison group

Schools in the comparison group did not receive the assessment tools or associated training during the year in which the study occurred. In comparison schools, 88% of reading and mathematics teachers reported using assessment data to monitor student progress, and 75% reported customizing their instruction based on the monitoring results. Comparison schools were eligible to receive the intervention in the following school year.

Outcomes and measurement

The study authors assessed students in participating schools on the mathematics and reading achievement outcome measures of ISTEP+ (grades 3–8) and TerraNova (grades K–2) standardized tests. For a more detailed description of these outcome measures, see Appendix B.

Support for implementation

The Indiana Department of Education (IDOE) covered schools' costs of implementing the new diagnostic assessment tools. Training was delivered to teachers at intervention schools using a train-the-trainer model. Between one and four volunteer teachers from participating schools received 2–3 days of summer training on the tools from the state of Indiana and the assessment tool vendors. The assessment vendors conducted another training in the fall after the first testing window. The volunteer teacher trainers were then given training materials and asked to deliver two or three training sessions within 6 months to teachers at their schools.

Reason for review

This study was identified for review by the WWC because it was supported by a grant to Learning Point Associates (Co-Principal Investigators: Spyros Konstantopoulos and Shazia Miller) from the National Center for Education Research (NCER) at the Institute of Education Sciences (IES).

Appendix B: Outcome measures for each domain

Mathematics achievement	
TerraNova mathematics achievement (grades K–2)	The TerraNova is a nationally-normed standardized achievement assessment. No additional detail on the level(s) or form(s) of the TerraNova used is provided in the study. The authors translated the students' scores on the TerraNova to *z*-scores by standardizing student scores within each grade level using sample mean and standard deviation.
Indiana Statewide Testing for Educational Progress–Plus (ISTEP+) mathematics (grades 3–8)	ISTEP+ is the Indiana state assessment for grades 3–8. Documentation on the IDOE's website indicates that the spring 2010 ISTEP+ mathematics test contained multiple-choice and open-ended questions that focused on the following areas: number sense, computation, algebra and functions, geometry, measurement, data analysis and probability, and problem solving skills. The authors translated the students' scores on the ISTEP+ to *z*-scores by standardizing student scores within each grade level using sample mean and standard deviation.
Reading achievement	
TerraNova reading achievement (grades K–2)	The TerraNova is a nationally-normed standardized achievement assessment. The authors do not provide any detail on the level or form of the TerraNova used for this study. The authors translated the students' scores on the TerraNova to *z*-scores that standardized student scores within each grade level.
ISTEP+ English/language arts (grades 3–8)	ISTEP+ is the Indiana state assessment for grades 3–8. Documentation on the IDOE's website indicates the spring 2010 ISTEP+ English/language arts test contained multiple-choice, open-ended, and gridded-response questions that focused on the following areas: vocabulary, nonfiction/informational text, literary text, writing process, application, and language conventions. The authors translated the students' scores on the ISTEP+ to *z*-scores by standardizing student scores within each grade level using sample mean and standard deviation.

Appendix C: Study findings for each domain

Domain and outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			*p*-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Mathematics achievement								
TerraNova mathematics achievement (grades K–2) and Indiana Statewide Testing for Educational Progress–Plus (ISTEP+) mathematics achievement (grades 3–8)	Grades K–8	57 schools/ 19,167 students	nr	nr	0.13	0.13	+5	.07
Domain average for mathematics achievement						**0.13**	**+5**	**Not statistically significant**
Reading achievement								
TerraNova reading achievement (grades K–2) and ISTEP+ English/ language arts (grades 3–8)	Grades K–8	57 schools/ 19,173 students	nr	nr	0.08	0.08	+3	.14
Domain average for reading achievement						**0.08**	**+3**	**Not statistically significant**

Table Notes: For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on individual outcomes, representing the average change expected for all individuals who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average individual's percentile rank that can be expected if the individual is given the intervention. The statistical significance of the study's domain average was determined by the WWC. nr = not reported.

Study Notes: No corrections for clustering or multiple comparisons and no difference-in-differences adjustment were needed. In the paper, the authors reported results based on a variety of analytic strategies and samples. In this table, the effect sizes and improvement indices are based on the regression coefficients and *p*-values from the two-level model ITT analyses provided by the authors in response to an author query. The estimates are from the model that includes student and school characteristics and grade dummy variables, and are based on a sample which excludes students who joined the study classrooms after random assignment and students who crossed over from one study group to another. The WWC focused on the ITT estimates because they are the only estimates eligible for the highest WWC rating of *meets WWC group design standards without reservations*. We present the estimates for the full K–8 sample as the primary estimate here because it is the most comprehensive estimate of the intervention effect. We also present the grade-specific estimates in Appendix D. This study is characterized as having neither statistically significant nor substantively important effects for mathematics or reading achievement. For more information, please refer to the WWC Standards and Procedures Handbook (version 3.0), p. 26.

Appendix D: Supplemental findings by domain

Domain and outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			*p*-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Mathematics achievement								
Indiana Statewide Testing for Educational Progress–Plus (ISTEP+) mathematics achievement	Grade 3	57 schools/ 3,432 students	nr	nr	0.13	0.13	+5	.12
ISTEP+ mathematics achievement	Grade 4	57 schools/ 3,431 students	nr	nr	0.13	0.13	+5	.11
ISTEP+ mathematics achievement	Grade 5	56 schools/ 3,267 students	nr	nr	0.30	0.31	+12	< .01
ISTEP+ mathematics achievement	Grade 6	26 schools/ 1,473 students	nr	nr	0.31	0.31	+12	.02
Reading achievement								
ISTEP+ English/language arts	Grade 3	57 schools/ 3,429 students	nr	nr	0.15	0.15	+6	.02
ISTEP+ English/language arts	Grade 4	57 schools/ 3,422 students	nr	nr	0.13	0.13	+5	.01
ISTEP+ English/language arts	Grade 5	56 schools/ 3,260 students	nr	nr	0.14	0.14	+5	.03
ISTEP+ English/language arts	Grade 6	26 schools/ 1,470 students	nr	nr	–0.02	–0.02	–1	.85

Table Notes: The supplemental findings presented in this table are additional findings that do not factor into the determination of the evidence rating. For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on individual outcomes, representing the average change expected for all individuals who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average individual's percentile rank that can be expected if the individual is given the intervention. nr = not reported.

Study Notes: A correction for multiple comparisons was needed but did not affect whether any of the contrasts were found to be statistically significant. In the paper, the authors reported results based on a variety of analytic strategies and samples. In this table, the effect sizes and improvement indices are based on the regression coefficients and *p*-values from the two-level model ITT analyses provided by the authors in response to an author query. The estimates are from the model that includes student and school characteristics and are based on a sample which excludes students who joined the study schools after random assignment and students who crossed over from one study group to another. The WWC focused on the ITT estimates because they are the only estimates eligible for the highest WWC rating of *meets WWC group design standards without reservations*. ITT estimates were not available for grades K–2 because the TerraNova was not administered to schools that attrited from the study (in response to an author query, the authors indicated that the ITT estimates for grades K–2 reported in Table 6 of the study were reported in error). The authors did not present grade-specific ITT estimates for grades 7 and 8 because of concerns about data availability.

Endnotes

[1] Single study reviews examine evidence published in a study (supplemented, if necessary, by information obtained directly from the authors) to assess whether the study design meets WWC group design standards. The review reports the WWC's assessment of whether the study meets WWC group design standards and summarizes the study findings following WWC conventions for reporting evidence on effectiveness. This study was reviewed using the single study review protocol, version 2.0. The Primary Math (version 3.0), Secondary Math (version 3.0), Beginning Reading (version 2.1), and Adolescent Literacy (version 2.1) review protocols were consulted to determine the correct domain for the outcomes presented in this review.

[2] Originally, 70 schools were randomly selected from a sample of 116 K–8 schools deemed eligible to participate in this study (from an initial pool of 264 schools who volunteered to participate in mCLASS and 421 who volunteered to participate in Acuity). The 116 schools in the initial sample were deemed eligible because they (a) agreed to use both mCLASS and Acuity, (b) had not previously used either tool or a similar tool in the previous year, and (c) were not participating in Indiana's No Child Left Behind differentiated accountability pilot in 2009–10 academic year (which required the use of mCLASS and Acuity). Random selection for participation in the study was stratified by census locale (urban, suburban, small town, and rural) to ensure the sample was representative of the state geographic balance. After the 70 schools were selected for participation, eleven schools were dropped from the study prior to random assignment, either because the software vendors indicated the schools had used one of their products in the previous year (10 schools) or the school had closed (one school).

[3] Grade-specific results for grades 7 and 8 were not reported. In response to an author query, the authors indicated that the ITT estimates for grades K–2 reported in Table 6 of the study were reported in error.

[4] These estimates were provided in response to an author query.

Recommended Citation

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2015, February). *WWC review of the report: The impact of Indiana's system of interim assessments on mathematics and reading achievement.* Retrieved from http://whatworks.ed.gov

Glossary of Terms

Attrition Attrition occurs when an outcome variable is not available for all participants initially assigned to the intervention and comparison groups. The WWC considers the total attrition rate and the difference in attrition rates across groups within a study.

Clustering adjustment If intervention assignment is made at a cluster level and the analysis is conducted at the student level, the WWC will adjust the statistical significance to account for this mismatch, if necessary.

Confounding factor A confounding factor is a component of a study that is completely aligned with one of the study conditions, making it impossible to separate how much of the observed effect was due to the intervention and how much was due to the factor.

Design The design of a study is the method by which intervention and comparison groups were assigned.

Domain A domain is a group of closely related outcomes.

Effect size The effect size is a measure of the magnitude of an effect. The WWC uses a standardized measure to facilitate comparisons across studies and outcomes.

Eligibility A study is eligible for review if it falls within the scope of the review protocol and uses either an experimental or matched comparison group design.

Equivalence A demonstration that the analytic sample groups are similar on observed characteristics defined in the review area protocol.

Improvement index Along a percentile distribution of individuals, the improvement index represents the gain or loss of the average individual due to the intervention. As the average individual starts at the 50th percentile, the measure ranges from –50 to +50.

Multiple comparison adjustment When a study includes multiple outcomes or comparison groups, the WWC will adjust the statistical significance to account for the multiple comparisons, if necessary.

Quasi-experimental design (QED) A quasi-experimental design (QED) is a research design in which study participants are assigned to intervention and comparison groups through a process that is not random.

Randomized controlled trial (RCT) A randomized controlled trial (RCT) is an experiment in which eligible study participants are randomly assigned to intervention and comparison groups.

Single-case design (SCD) A research approach in which an outcome variable is measured repeatedly within and across different conditions that are defined by the presence or absence of an intervention.

Standard deviation The standard deviation of a measure shows how much variation exists across observations in the sample. A low standard deviation indicates that the observations in the sample tend to be very close to the mean; a high standard deviation indicates that the observations in the sample are spread out over a large range of values.

Statistical significance Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The WWC labels a finding statistically significant if the likelihood that the difference is due to chance is less than 5% ($p < .05$).

Substantively important A substantively important finding is one that has an effect size of 0.25 or greater, regardless of statistical significance.

Please see the WWC Procedures and Standards Handbook (version 3.0) for additional details.