

January 2017

Thank you for your request to our REL Reference Desk regarding **research on the impact of diagnostics and interims on instruction**. Ask REL Southwest is part of a collaborative Ask-A-REL reference desk service provided by the 10 regional educational laboratories (RELs). By design, this service functions much in the same way as a technical reference library, by providing references, referrals, and brief responses in the form of citations for research-based education questions.

Please note that REL Southwest has not done an evaluation of the resources themselves but offers this list to you for your information only.

BACKGROUND

“An **interim assessment** is a form of assessment that educators use to (1) evaluate where students are in their learning progress and (2) determine whether they are on track to performing well on future assessments, such as standardized tests or end-of-course exams. Interim assessments are usually administered periodically during a course or school year (for example, every six or eight weeks) and separately from the process of instructing students.

Defining *interim assessment* is complicated by the fact that educators use a variety of terms for these forms of assessment—such as *benchmark assessment* or *predictive assessment*—and the terms may or may not be used synonymously. It should also be noted that there is often confusion and debate about the distinctions between formative assessments and interim assessments, and educators may define the terms differently from school to school or state to state.

Generally speaking, interim assessments fall between formative and summative assessment, and understanding their intended purpose requires an understanding of the basic distinctions between these two assessment strategies.”¹

Following an established REL Southwest protocol, we conducted a search for research reports, websites, as well as descriptive briefs on the impact of diagnostics and interims on instruction. **We searched the references in the response from the most commonly used resources of research, but they are not comprehensive and other relevant references and resources may exist.** The sources included federally funded organizations, additional research institutions, educational databases, and general Internet searches using Google and Bing. See the methods section at the end of this Ask A REL for additional information on how we identified the following sources.

¹ The glossary of education reform: For Journalists, parents, and community members. <http://edglossary.org/interim-assessment/>. (Last updated 10.30.2013)

QUESTION

How do diagnostics and interims impact instruction? (Does having strong diagnostics and interims support student achievement?)

SOURCES

Blanc, S., Christman, J. B., Liu, R., Mitchell, C., Travers, E., & Bulkley, K. E. (2010). Learning to learn from data: benchmarks and instructional communities. *Peabody Journal of Education*, v85 n2 p205-225. <https://eric.ed.gov/?id=EJ883193>.

From the ERIC abstract: “This article examines the use of interim assessments in elementary schools in the School District of Philadelphia. The article reports on the qualitative component of a multimethod study about the use of interim assessments in Philadelphia. The study used an organizational learning framework to explore how schools can best develop the capacity to utilize the potential benefits of interim assessments. The qualitative analysis draws on data from intensive fieldwork in 10 elementary schools and interviews with district staff and others who worked with the schools, as well as further in-depth case study analysis of 5 schools. This article examines how school leaders and grade groups made sense of data provided through interim assessments and how they were able to use these data to rethink instructional practice. We found substantial evidence that interim assessments have the potential to contribute to instructional coherence and instructional improvement if they are embedded in a robust feedback system. Such feedback systems were not the norm in the schools in our study, and their development requires skill, knowledge, and concerted attention on the part of school leaders. (Contains 1 figure.)”

Chan, P. E., Graham-Day, K. J.; Ressa, V. A.; Peters, M. T. & Konrad, M. (2014). Beyond involvement: Promoting student ownership of learning in classrooms. *Intervention in School and Clinic*, v50 n2 p105-113. <https://eric.ed.gov/?id=EJ1042118>.

From the ERIC abstract: “With many states adopting new standards and evaluation systems, teachers must adopt effective instructional strategies and assessment methods aligned to the rigor of new standards and assessments. One way to improve student achievement is through supporting student ownership of learning, a core component of formative instructional practices. Teaching students to take an active role in their learning can benefit students by promoting student goal setting, self-assessment, and self-determination. As students become meaningfully engaged in their learning, they gain a better understanding of learning targets, how to collect and document evidence of their learning, and how to evaluate and clarify additional learning needs, leading to the ultimate goal of improving student achievement. This article (a) describes how promoting student ownership benefits students, (b) identifies some evidence-based practices that promote student ownership of learning, and (c) illustrates the important role student ownership plays in formative instructional practices.”

Chojnacki, G., Eno, J., Liu, F., Meyers, C., Konstantopoulos, S., Miller, S., & van der Ploeg, A. (2013). Do interim assessments influence instructional practice in year

one? Evidence from Indiana elementary school teachers. *Society for Research on Educational Effectiveness*. <https://eric.ed.gov/?id=ED563059>.

From the ERIC abstract: “Recent work that has examined the impact of what are variously called periodic, interim, benchmark, or diagnostic assessments, typically administered three or four times during a school year, has produced mixed findings. For instance, one study reported small significant effects in mathematics in grades 3-8, but not in reading (Carlson et al., 2011). Other research however, has reported significant effects on both mathematics and reading (Slavin et al., 2011). Finally, a very recent study found no effects on reading achievement in grades 4-5 (Cordray et al., 2012). This study compares instructional practices of teachers in schools that were randomly assigned to receive an interim assessment tool with those of teachers in schools that did not receive the tool. Using rich data collected at 16 time points during the school year, the authors study teachers' self-reported instructional practices to determine whether teachers with access to an interim assessment tool alter each of three facets of instructional practice--scope and sequence of content coverage, instructional level, and instructional grouping--more than those without the tool. The research questions are: (1) Do teachers with access to the interim assessment change the scope and sequence of content, and/or vary instructional difficulty level and grouping methods more than those without? (2) Do variations in these teacher practices respond to variations in student Acuity performance? Researchers employ treatment vs. control comparisons to explore whether teachers with the interim assessment intervention engage in expected instructional practices more than those without it. Results are reported from rich data on teacher instructional practices generated at sixteen intervals by teachers with and without access to a specific interim assessment tool. Estimates provide no strong evidence that teachers change the instructional practices measured here in response to Acuity performance data. One possible reason for these findings is that Acuity is not a unique intervention, and a significant number of control teachers reported using other interim assessment tools. Another possible explanation for these results is that the relatively small sample of teachers completing checklists harms the study's power. Finally, these results pertain to the first year of the intervention, when teachers are likely still learning how to use the assessment tool and integrate it into their instructional practice. Future research should explore the hypothesis that impacts on teacher practice grow over time as teachers learn to use the assessment tool. Tables and figures are appended.”

Klug, J., Bruder, S. & Schmitz, B. (2016). Which variables predict teachers diagnostic competence when diagnosing students' learning behavior at different stages of a teacher's career? *Teachers and Teaching: Theory and Practice*, v22 n4 p461-484. <https://eric.ed.gov/?id=EJ1093563>.

From the ERIC abstract: “Diagnosing is one of teachers' key competences. Lately, a new model of teachers' diagnostic competence and therewith a new measurement method focusing on diagnosing students' learning behavior in an interdisciplinary approach have been developed. They can build a basis for promoting teachers' diagnostic competence. However, some questions relevant to promoting teachers' diagnostic competence concerning students' learning behavior still remain unanswered:

Which values do teachers-on-the-job in contrast to preservice teachers with or without teaching experience gain through experience and which are the relevant variables that lead to a high level of diagnostic competence focused on learning behavior? Thus, this study aims at enlightening these issues. In a sample of 293 participants (93 German grammar school teachers, 107 German teacher students in their second phase of teacher education, and 93 German preservice students in their first phase of teacher education), teachers' diagnostic competence concerning learning behavior as well as teachers' motivation, attitude, self-efficacy, knowledge, and reflection on experience concerning diagnosing was measured. Motivation, attitude, and knowledge were found to be substantial positive predictors of diagnostic competence concerning learning behavior, with differentiated predictions for teachers' experience. Reflection on experience and self-efficacy unexpectedly was not found to be relevant. Teachers-on-the-job were generally better at diagnosing their students' learning behavior than were preservice teachers without teaching experience. However, the diagnostic work of preservice teachers with teaching experience was as credible as that of the in-service teachers."

Konstantopoulos, S., Li, W., Miller, S. R., & van der Ploeg, A. (2016). Effects of interim assessments across the achievement distribution: Evidence from an experiment. *Educational and Psychological Measurement*, v76 n4 p587-608.
<https://eric.ed.gov/?id=EJ1106300>.

From the ERIC abstract: "We use data from a large-scale experiment conducted in Indiana in 2009-2010 to examine the impact of two interim assessment programs (mCLASS and Acuity) across the mathematics and reading achievement distributions. Specifically, we focus on whether the use of interim assessments has a particularly strong effect on improving outcomes for low achievers. Quantile regression is used to estimate treatment effects across the entire achievement distribution (i.e., provide estimates in the lower, middle, or upper tails). Results indicate that in Grades 3 to 8 (particularly third, fifth, and sixth) lower achievers seem to benefit more from interim assessments than higher achieving students."

Konstantopoulos, S., Miller, S. R.; van der Ploeg, A., & Li, W. (2016). Effects of interim assessments on student achievement: Evidence from a large-scale experiment. *Journal of Research on Educational Effectiveness*, v9 suppl 1 p188-208.
<https://eric.ed.gov/?id=EJ1115334>.

From the ERIC abstract: "We use data from a large-scale, school-level randomized experiment conducted in 2010-2011 in public schools in Indiana. Our sample includes more than 30,000 students in 70 schools. We examine the impact of two interim assessment programs (i.e., mCLASS in Grades K-2 and Acuity in Grades 3--8) on mathematics and reading achievement. Two-level models were used to capture the nesting in the data. Results indicate that the treatment effect is insignificant in Grades 3-8, and thus students in treatment schools perform as well as students in control schools. In contrast, the treatment effect is negative and significant in Grades K-2 (i.e., kindergarten and second grade), indicating that students in control schools perform higher than students in treatment schools."

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*, v35 n4 p481-499. <https://eric.ed.gov/?id=EJ1019178>.

From the ERIC abstract: “Interim assessments are increasingly common in U.S. schools. We use high-quality data from a large-scale school-level cluster randomized experiment to examine the impact of two well-known commercial interim assessment programs on mathematics and reading achievement in Indiana. Results indicate that the treatment effects are positive but not consistently significant. The treatment effects are smaller in lower grades (i.e., kindergarten to second grade) and larger in upper grades (i.e., third to eighth grade). Significant treatment effects are detected in Grades 3 to 8, especially in third- and fourth-grade reading and in fifth- and sixth-grade mathematics.”

Koon, S.& Petscher, Y. (2015). Can scores on an interim high school reading assessment accurately predict low performance on college readiness exams? REL 2016-124. *Regional Educational Laboratory Southeast*. <https://ies.ed.gov/pubsearch/pubsinfo.asp?pubid=REL2016124>.

From the IES description: “The purpose of this study was to examine the relationship between measures of reading comprehension, decoding, and language with college-ready performance. This research was motivated by leaders in two Florida school districts interested in the extent to which performance on Florida’s interim reading assessment could be used to identify students who may not perform well on the Preliminary SAT/National Merit Scholarship Qualifying Test (PSAT/NMSQT) and ACT Plan. One of the districts primarily administers the PSAT/NMSQT and the other primarily administers the ACT Plan. Data included the 2013/14 PSAT/NMSQT or ACT Plan results for students in grade 10 from these districts, as well as their grade 9 results on the Florida Assessments for Instruction in Reading – Florida Standards (FAIR-FS). Classification and regression tree (CART) analyses formed the framework for an early warning system of risk for each PSAT/NMSQT and ACT Plan subject-area assessment. PSAT/NMSQT Critical Reading performance is best predicted in the study sample by a student’s reading comprehension skills, while PSAT/NMSQT Mathematics and Writing performance is best predicted by a student’s syntactic knowledge. Syntactic knowledge is the most important predictor of ACT Plan English, Reading, and Science in the study sample, whereas reading comprehension skills were found to best predict ACT Plan Mathematics results. Sensitivity rates (the percentage of students correctly identified as at risk) ranged from 81 percent to 89 percent correct across all of the CART models. These results provide preliminary evidence that FAIR-FS scores could be used to create an early warning system for performance on both the PSAT/NMSQT and ACT Plan. The potential success of using FAIR-FS scores as an early warning system could enable districts to identify at-risk students without adding additional testing burden, time away from instruction, or additional cost. The analyses should be replicated statewide to verify the stability of the models and the generalizability of the results to the larger Florida student population.”

Li, H., Hunter, C. V., & Lei, P. (2016). The Selection of Cognitive Diagnostic Models for a Reading Comprehension Test. *Language Testing*, v33 n3 p391-409.
<https://eric.ed.gov/?id=EJ1104447>.

From the ERIC abstract: “Cognitive diagnostic models (CDMs) have great promise for providing diagnostic information to aid learning and instruction, and a large number of CDMs have been proposed. However, the assumptions and performances of different CDMs and their applications in regard to reading comprehension tests are not fully understood. In the present study, we compared the performance of a saturated model (G-DINA), two compensatory models (DINO, ACDM), and two non-compensatory models (DINA, RRUM) with the Michigan English Language Assessment Battery (MELAB) reading test. Compared to the saturated G-DINA model, the ACDM showed comparable model fit and similar skill classification results. The RRUM was slightly worse than the ACDM and G-DINA in terms of model fit and classification results, whereas the more restrictive DINA and DINO performed much worse than the other three models. The findings of this study highlighted the process and considerations pertinent to model selection in applications of CDMs with reading tests. [This paper was presented at the 2014 Annual Meeting of Northeastern Educational Research Association (NERA), Trumbull, CT.]”

Li, Y., Marion, S., Perie, M., & Gong, B. (2010). An Approach for Evaluating the Technical Quality of Interim Assessments. *Peabody Journal of Education*, v85 n2 p163-185. <https://eric.ed.gov/?id=EJ883187>.

From the ERIC abstract: “Increasing numbers of schools and districts have expressed interest in interim assessment systems to prepare for summative assessments and to improve teaching and learning. However, with so many commercial interim assessments available, schools and districts are struggling to determine which interim assessment is most appropriate to their needs. Unfortunately, there is little research-based guidance to help schools and districts to make the right choice about how to spend their money. Because we realize the urgency of developing criteria that can describe or evaluate the quality of interim assessments, this article presents the results of an initial attempt to create an instrument that school and district educators could use to evaluate the quality and usefulness of the interim assessment. The instrument is designed for use by state and district leaders to help them select an appropriate interim assessment system for their needs, but it could also be used by test vendors looking to evaluate and improve their own systems and by researchers engaged in studies of interim assessment use. (Contains 14 tables and 4 footnotes.)”

Pereira, M., & Tienken, C. (2012). An evaluation of the influence of interim assessments on grade 8 student achievement in mathematics and language arts. *International Journal of Educational Leadership Preparation*, v7 n3.
<https://eric.ed.gov/?id=EJ997471>.

From the ERIC abstract: “A review of the literature pertaining to the effect and influence that interim assessments have on student achievement lacks quantitative data to determine the efficiency of their use in the classroom as a school reform tool. This study examined the strength and the direction of the relationships between interim pre and

posttest assessments in language arts and mathematics in Grade 8 and student achievement on the New Jersey Grade 8 state standardized tests in those subjects. Analyses were conducted using simultaneous multiple regression models. All student data explored in this study pertained to 670 students in Grade 8 enrolled in four middle schools located in a suburban/urban central New Jersey community during the 2009-2010 academic school year. The results of the study revealed each school produced a combination of site specific results and the interim pretests accounted for the same or almost the same amount of variance in state test scores as the interim posttests. (Contains 1 table, 1 figure, and 3 footnotes.)”

Wang, Y. & Gushta, M. (2013). Improving student outcomes with mCLASS: Math, a technology-enhanced CBM and diagnostic interview assessment. *Society for Research on Educational Effectiveness*. <https://eric.ed.gov/?id=ED563114>.

From the ERIC abstract: “The No Child Left Behind Act resulted in increased school-level implementation of assessment-based school interventions that aim to improve student performance. Diagnostic assessments are included among these interventions, designed to help teachers use evidence about student performance to modify and differentiate instruction and improve student outcomes. The mCLASS: Math software (Ginsburg, Cannon, Eisenband, & Pappas, 2006) is comprised of screening/progress monitoring curriculum-based measures (CBMs) and Diagnostic Interviews to help teachers identify students' skill levels. mCLASS: Math enables teachers to target instruction to each student's needs and monitor each student's progress toward mastery. Educators are expected to use the constantly updated diagnostic information to improve ongoing instruction and increase student achievement. Schools have found that the use of mCLASS: Math helps support curriculum, instruction, and assessment (Ginsburg et al, 2006). This study provides evidence about the effectiveness of mCLASS: Math in improving student outcomes on a statewide math achievement test, examining whether or not using the mCLASS: Math CBM impacts students' achievement test scores. Among the students who were administered the CBM, we investigate whether or not being administered the mCLASS: Math Diagnostic Interviews further influences subsequent test scores. This study was conducted in a large mid-western state which provided data from 175 districts, 606 schools, and 1856 teachers for this study. All the schools and teachers had elected to implement mCLASS: Math; however, there are instances in which individuals students may be exempted from the assessment. CBM measures quickly identify students who are at risk for poor mathematical performance while Diagnostic Interviews identify students' mathematical strengths and weaknesses which can be used to inform instruction. Overall, the administration of CBM measures provides broad, useful information about students' skill levels and identifies those students in need of further assessments and development. Tables are appended.”

What Works Clearinghouse (2015). WWC Review of the Report "The Impact of Indiana's System of Interim Assessments on Mathematics and Reading." What Works Clearinghouse Single Study Review. What Works Clearinghouse. <https://eric.ed.gov/?id=ED553423>.

From the ERIC abstract: "The study, "The Impact of Indiana's System of Interim Assessments on Mathematics and Reading," examined the effects of using Diagnostic Assessment Tools (DAT) on mathematics and reading outcomes for students in 59 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 goal is for teachers to use the assessment results to tailor instruction to students' needs. After random assignment, schools in the intervention group received DAT, and schools in the comparison group did not receive the assessment tools or associated training. The study is a well-executed randomized controlled trial with low sample attrition. A subset of the analyses described in the study meets WWC group design standards without reservations. The study authors found, and the WWC confirmed, that the use of DAT did not have a statistically significant impact on general mathematics achievement or reading achievement for the full sample of students in grades K-8, but that the use of DAT did have statistically significant positive effects for grades 5 and 6 in mathematics achievement and grades 3-5 in reading achievement.[The following study is the focus of this "Single Study Review:"
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. Appended are: (1) Study details; (2) Outcome measures for each domain; (3) Study findings for each domain; and (4) Supplemental findings by domain. A glossary of terms is included."

ADDITIONAL RESOURCES

Ananda, S., & Rabinowitz, S. (2001). High stakes and assessment innovation: A negative correlation? San Francisco: WestEd.

http://www.wested.org/online_pubs/highstakes.pdf

From the abstract: "Thus far, there has been an inverse correlation between innovation and accountability in statewide assessment systems: the higher the stakes attached to assessment results, the more conservative the assessment methodology ultimately used." This research report examines the impact of high stakes accountability systems on assessment innovation, particularly as the use of high stakes has affected the introduction of performance-based assessments. Included are two case studies of state programs (Kentucky and California) whose experiences illustrate the increasing and often overwhelming demands for accountability throughout the education system and the inadequacy of existing assessment delivery infrastructure and methodology to readily accommodate innovation. The report includes a series of "lessons learned" that may be helpful to states now revisiting their assessment systems." NOTE: This source was not peer reviewed.

Marshall, K. (2006). Interim assessments: Keys to successful implementation. Interim Assessment Project New Leaders for New Schools.

<http://www.marshallmemo.com/articles/Interim%20Assmt%20Report%20Apr.%202012,%202006.pdf>

From the introduction: “How can educators raise student achievement to high levels and close the racial/economic gap? What is the “secret sauce” in the much-admired 90/90/90 schools – those with 90 percent students of color, 90 percent qualifying for free and reduced-price meals, and 90 percent meeting rigorous state standards? One of the key ingredients, the research tells us, is the effective use of interim assessments. Virtually all low-SES schools with impressive student achievement – as well as higher-SES schools with a shrinking achievement gap – systematically analyze assessments during the school year and put the data to work in classrooms.” This paper also provides quotes about interim assessments.” NOTE: This source was not peer reviewed.

Pinkus, L. (2008). Using early-warning data to improve graduation rates: Closing cracks in the education system. Washington, DC: Alliance for Excellent Education.
<https://eric.ed.gov/?id=ED510882>

From the ERIC abstract: “With a national high school graduation rate, hovering around 70 percent, far too many of the nation's students are falling through the cracks of the education system and leaving high school without the skills necessary for success in college, work, and life. By preventing students from falling through the cracks and ensuring that they receive the appropriate level of attention, instruction, engagement, and support needed to succeed in their classes, educators can give every student the chance to graduate from high school prepared for college, the modern workforce, and life. This brief explores the predictive power of early-warning data, offers examples of current efforts to use such data to guide secondary school interventions across the country, and discusses the policies that can support these efforts. (Contains 1 footnote and 43 endnotes.)” NOTE: This source was not peer reviewed.

METHODS

Keywords and Search Strings Used in the Searches:

Diagnostics; Diagnostic testing; Educational + Interims; Interim assessment; Benchmarks

Search of Databases and Websites

- Institute of Education Sciences (IES) website (<http://www.ies.ed.gov>) and IES sources: Regional Educational Laboratory (REL) Program, National Center for Education Statistics (NCES), National Center for Education Research (NCER), What Works Clearinghouse (WWC)
- ERIC database (www.eric.ed.gov)
- Google Scholar (scholar.google.com)
- Google (www.google.com)
- Bing (www.bing.com)

Criteria for Inclusion

REL Southwest selected resources that provide research on the impact of diagnostics and interims on instruction. When REL Southwest staff reviewed resources, we considered – among other things – three factors:

1. **Date of Publication:** The most current information (primarily published from 2010 to the present) is included.
2. **Source and Funder of the Report/Brief/Article:** Priority was given to publications written in relevant, peer-reviewed journals or reports or produced by well-known research organizations.
3. **Methodology:** sources include reported studies, literature reviews and policy reports.

Ask A REL is a service provided by a collaborative of the Regional Educational Laboratory (REL) Program, funded by the U.S. Department of Education's Institute of Education Sciences (IES). This response was prepared by REL Southwest under contract ED-IES-12-C-0012 with IES. The content of this document does not necessarily reflect the views or policies of IES or the U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.