



REL Appalachia Ask-A-REL Response

Data Use, Early Childhood, Literacy
May 2017

Question:

What does the research say about the equivalency of scales on early grades (PK-4) literacy universal screening and progress monitoring measures (specifically: AIMSweb, easyCBM, STAR, DIBELS, iReady, NWEA MAP)?

Response:

Thank you for your request to our REL Reference Desk regarding evidence-based information about universal screening and progress monitoring measures in literacy. Ask-A-REL is a collaborative reference desk service provided by the ten regional educational laboratories (REL) that, by design, functions much in the same way as a technical reference library. It provides references, referrals, and brief responses in the form of citations in response to questions about available education research.

Following an established REL Appalachia research protocol, we conducted a search for research reports and descriptive study articles on the equivalency of scales on early grades literacy universal screening and progress monitoring measures, and important considerations when trying to equate those assessments. The sources searched included ERIC and other federally funded databases and organizations, research institutions, academic research databases, and general Internet search engines. For more details, please see the methods section at the end of this document.

We have not evaluated the quality of the resources provided in this response; we offer them only for your reference. Also, the search included commonly used research databases and search engines to provide the references in this document, but the references are not necessarily comprehensive, and other relevant references and resources may exist.

Research References

Dorans, N.J. (2008). The practice of comparing scores on different tests. *R&D Connections*, 6, 1-5. Retrieved from https://www.ets.org/Media/Research/pdf/RD_Connections6.pdf.

From the introduction: "Appropriate comparisons require careful data collection, analysis, and interpretation. To what extent can we use scores to compare performance on one test with performance on a different test?"

Iowa Department of Education. (2013). *A summary report of Iowa's review of PreK-6 reading assessments for universal screening and progress*. Des Moines, IA: Iowa Department of Education. Retrieved from <http://eric.ed.gov/?id=ED544319>.

From the abstract: "This document contains summary information for the Iowa Department of Education's review of PreK-6th grade reading assessments for the purposes of Universal Screening and Progress Monitoring. It is intended to provide general information to help inform decisions about selecting assessments for use as a part of Iowa's Response to Intervention work. Brief descriptions of what each column of information means in the summaries of the reviews of the universal screening and progress monitoring assessments are provided. Each description represents one column of information, starting with the column title in bold. This document is built around talking points for the reviews of assessments for universal screening and progress monitoring. There are other more technical documents that go into greater detail about the reviews. Universal screening is about assessing all students three times a year to identify those on track for success in reading, and those that might need something "more" in order to get back on track for success. Monitoring progress is about assessing students who are getting something "more" in order to make sure that the students are improving."

Lai, C., Alonzo, J., & Tindal, G. (2013). *easyCBM Reading criterion related validity evidence: Grades 2-5. Technical Report #1310*. Eugene, OR: Behavioral Research and Teaching, University of Oregon. Retrieved from <http://eric.ed.gov/?id=ED545272>.

From the abstract: "In this technical report, we present the results of a study to gather criterion-related evidence for Grade 2-5 easyCBM[®] reading measures. We used correlations to examine the relation between the easyCBM[®] measures and other published measures with known reliability and validity evidence, including the Gates-MacGinitie Reading Tests and the Dynamic Indicators of Basic Early Literacy Skills (DIBELS). Across grades, the correlation between easyCBM[®] vocabulary and comprehension-based measures and comparator measures ranged from low to moderate ($r_s = 0.39-0.76$), and the correlation between the easyCBM[®] fluency-based measures and DIBELS ORF was consistently strong ($r > 0.80$)."

Lazer, S., Mazzeo, J., Way, W.D., Twing, J.S., Camara, W., & Sweeney, K. (2010). *Thoughts on linking and comparing assessments of Common Core Standards* (ETS, Pearson, & the College Board white paper). Princeton, NJ: Educational Testing Service. Retrieved from http://images.pearsonassessments.com/images/tmrs/tmrs_rg/LinkingandComparingCCAssessments.pdf.

From the introduction: "It is likely that even given the federal interest in developing assessments of common standards, a single national test will not emerge. The purpose of this paper is to discuss the types of comparisons that can and cannot be made among students who take different assessments supposedly developed to measure a single set of standards."

Livingston, S.A. (2014). *Equating test scores (without IRT). Second edition*. Princeton, NJ: Educational Testing Service. Retrieved from <https://eric.ed.gov/?id=ED560972>.

From the abstract: “This booklet grew out of a half-day class on equating that author Samuel Livingston teaches for new statistical staff at Educational Testing Service (ETS). The class is a nonmathematical introduction to the topic, emphasizing conceptual understanding and practical applications. The class consists of illustrated lectures, interspersed with self-tests for the participants. Livingston has included the self-tests in this booklet, at roughly the same points as they occur in the class. The answers are in a separate section at the end of the booklet. The topics in this second edition include raw and scaled scores, linear and equipercentile equating, data collection designs for equating, selection of anchor items, equating constructed-response tests (and other tests that include constructed-response questions), and methods of anchor equating. Livingston begins by assuming that the participants do not even know what equating is. By the end of the class, Livingston is explaining the logic of the Tucker method of equating and what conditions cause it to be biased.”

Mullis, I.V.S., & Martin, M.O. (2016). Dependable trend measurement is not just IRT scaling: Commentary on “Linking Large-Scale Reading Assessments: Measuring International Trends over 40 Years.” *Measurement: Interdisciplinary Research and Perspectives*, 14(1), 30-31. Retrieved from <https://eric.ed.gov/?id=EJ1091728>.

From the abstract: “Linking IEA's international reading assessments across 40 years is an interesting endeavor from several perspectives. Being able to examine trends in reading achievement at the 4th grade over such a long period and relate these to policy changes during that time span is an attractive idea. However, this work brings to the fore many thorny issues that should be considered in linking together several disparate assessments, and the technical complexities involved in using IRT scaling to analyze trends in large-scale international assessments.”

Renaissance Learning. (2014). *Converting Measures of Academic Progress (MAP) reading, language usage, and math RIT scores to STAR Reading and STAR Math scaled scores*. Wisconsin Rapids, WI: Renaissance Learning. Retrieved from <http://doc.renlearn.com/kmnet/r0057878bf88c975.pdf>.

From the introduction: “The purpose of this project is to statistically link the Northwest Evaluation Association (NWEA) Measures of Academic Progress (MAP) and STAR Assessments scales in order to facilitate the conversion of MAP RIT scores to STAR scaled scores. Linkages were completed between MAP Reading to STAR Reading, MAP Language Usage to STAR Reading, and MAP Math to STAR Math. The resulting conversion table makes it possible for present or future STAR users to translate their MAP RIT scores to STAR scaled scores.”

Reschly, A.L., Busch, T.W., Betts, J., Deno, S.L., & Long, J.D. (2009). Curriculum-based measurement oral reading as an indicator of reading achievement: A meta-analysis of the correlational evidence. *Journal of School Psychology, 47*(6), 427-469. Retrieved from <https://eric.ed.gov/?id=EJ869756>.

From the abstract: "This meta-analysis summarized the correlational evidence of the association between the CBM Oral Reading measure (R-CBM) and other standardized measures of reading achievement for students in grades 1-6. Potential moderating variables were also examined (source of criterion test, administration format, grade level, length of time, and type of reading subtest score). Results indicated a significant, strong overall correlation among R-CBM and other standardized tests of reading achievement and differences in correlations as a function of source of test, administration format, and reading subtest type. No differences in the magnitude of correlations were found across grade levels. In addition, there was minimal evidence of publication bias. Results are discussed in terms of existing literature and directions for future research."

Shapiro, E.S., & Gibbs, D.P. (2014). *Comparison of progress monitoring with computer adaptive tests and curriculum based measures*. Bethlehem, PA: Center for Promoting Research to Practice, Lehigh University. Retrieved from <http://doc.renlearn.com/KMNet/R0057324CE9DD5FD.pdf>.

From the abstract: "The purpose of this study was to compare both rates of reading achievement growth and predictive power of two widely-used assessments representing two different approaches to measurement – a computer adaptive assessment called STAR Reading and a curriculum based measurement called AIMSweb. A total of 117 students from a school district in Tennessee were included in the sample. Data collection spanned two school years, and included students who were progress monitored (taking a minimum of 4 tests per year) in grades 1 through 4 in one year, and in grades 2 through 5 the subsequent year. Across the two years, interventions for both groups of students were consistent. The results of this study indicate that both measures were able to detect incremental change, and provide further support that both computer adaptive measures such as STAR Reading and CBMs such as AIMSweb R-CBM are acceptable for progress monitoring. Of the two measures, only STAR Reading achieved a significant correlation with the state reading assessment."

Yu, C.H., & Popp, S.E.O. (2005). Test equating by common items and common subjects: Concepts and applications. *Practical Assessment, Research, & Evaluation, 10*(4), 1-19. Retrieved from <http://pareonline.net/pdf/v10n4.pdf>.

From the introduction: "Since the invention of z-scores (standardized scores), comparison among different tests has been widely conducted by test developers, instructors, educational researchers, and psychometricians. Equating, calibration, and moderation are terms used to describe broad levels of possible comparison among educational assessments (Dorans, 2004; Feuer, Holland, Green, Bertenthal, & Hemphill, 1999; Linn, 1993; Mislevy, 1992). Equating is at one end of the linking continuum, involving the most stringent requirements of equivalence among the assessments and examinee populations to be

linked, and compares tests that measure the same construct and have been designed to be equivalent. Less equivalent conditions involve calibration, which compares tests that measure the same construct but vary in design or difficulty, and moderation, which compares tests that measure different constructs. Psychometric approaches to linking assessments include linear equating, equipercentile equating, and item response theory (IRT). This article is a practical guide to conducting IRT test equating in two different scenarios.”

Additional Organizations to Consult

National Center on Intensive Intervention: <http://www.intensiveintervention.org/>

From the website: “Our mission is to build capacity of state and local education agencies, universities, practitioners, and other stakeholders to support implementation of intensive intervention in reading, mathematics, and behavior for students with severe and persistent learning and/or behavioral needs. In order to accomplish our mission, the NCII will establish technical review committees, provide intensive implementation support, and conduct a summative evaluation.”

Center on Response to Intervention, Progress Monitoring and Screening Tools Charts: <http://www.rti4success.org/resources/tools-charts>

From the website: “The Center on Response to Intervention, in collaboration with the National Center on Intensive Intervention, has established a standard process to evaluate the scientific rigor of commercially available tools and interventions that can be used in an MTSS/RTI context. Together, these two Centers conduct annual reviews of tools and interventions in the following three domains: Screening, Progress Monitoring, and Academic Intervention Programs.”

Methods

Keywords and Search Strings

The following keywords and search strings were used to search the reference databases and other sources:

- “Universal screening assessment” AND compar* OR link*
- “Progress monitoring assessment” AND compar* OR link*
- AIMSweb OR easyCBM OR STAR OR DIBELS OR iReady OR NWEA MAP AND compar* OR link*
- Linking assessments
- Item response theory

Databases and Resources

We searched ERIC, a free online library of over 1.6 million citations of education research sponsored by the Institute of Education Sciences, for relevant resources. Additionally, we

searched the academic database ProQuest, Google Scholar, and the commercial search engine Google.

Reference Search and Selection Criteria

When Reference Desk researchers review resources, they consider—among other things—four factors:

- **Date of the publication:** Searches include the most current information (i.e., within the last five years), except in the case of nationally known seminal resources.
- **Search priorities of reference sources:** Search priorities include IES, nationally funded, and certain other vetted sources known for strict attention to research protocols. Applicable resources must be publically available online and in English.
- **Methodology:** The following methodological priorities/considerations guide the review and selection of the references: (a) study types – randomized control trials, quasi experiments, surveys, descriptive data analyses, literature reviews, policy briefs, etc., generally in this order; (b) target population, samples (representativeness of the target population, sample size, volunteered or randomly selected, etc.), study duration, etc.; (c) limitations, generalizability of the findings and conclusions, etc.
- **Existing knowledge base:** Vetted resources (e.g., peer-reviewed research journals) are the primary focus; however, the research base is occasionally slim or nonexistent. In these cases, the best resources available may include, for example, reports, white papers, guides, reviews in non-peer reviewed journals, newspaper articles, interviews with content specialists, and organization websites.

Resources included in this document were last accessed on May 10, 2017. URLs, descriptions, and content included in this document were current at that time.

This memorandum is one in a series of quick-turnaround responses to specific questions posed by educational stakeholders in the Appalachian Region (Kentucky, Tennessee, Virginia, and West Virginia), which is served by the Regional Educational Laboratory Appalachia (REL AP) at SRI International. This Ask-A-REL response was developed by REL AP under Contract ED-IES-17-C-0004 from the U.S. Department of Education, Institute of Education Sciences, administered by SRI International. The content 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.