EdInstruments

Considerations for Use of Middle Grades Mathematics Assessments[1]

As part of the larger library of instruments, EdInstruments catalogues middle grades mathematics assessments for users to compare options for use in research and practice. These instruments are organized within Academic Knowledge and Skills > Math on our website. To guide potential users, we briefly describe middle grades mathematics content, highlight general psychometric properties to consider, and underscore specific considerations for use of such instruments by school practitioners.

Middle Grades Mathematics Content

The middle grades are a crucial period for mathematics learning, where students build on the foundational knowledge from elementary school and are exposed to new content domains en route to more advanced high school mathematics domains such as algebra.

The Common Core State Standards offer one potential learning progression for mathematics in the middle grades. Specifically, the standards state that Grade 5 students should build on their elementary mathematics education and continue learning about operations and algebraic thinking; numbers and operations in base ten; fractions; measurement and data; and geometry. In Grades 6 and 7, students should be exposed to more advanced mathematics domains including ratios and proportional relationships; the number system; expressions and equations; and statistics and probability. In Grade 8, students should then be exposed to functions and increasingly more depth into the domains from prior grades.

Middle grades mathematics assessments can serve a range of purposes. These uses can include individual diagnostics, universal screening, progress monitoring, accountability, research, and evaluation. As such, content coverage of a given assessment will vary depending on its intended purpose and use. Users should carefully consider their needs and ensure this aligns with the intended purpose of the assessment and the appropriate grade-level learning objectives.

[1] We thank Leanne Ketterlin Geller, Yue (Helena) Jia, Jim Soland, Jon R. Star, and Ming-Te Wang for their helpful feedback and expertise. Their participation in this work does not signify endorsement of any individual instrument/assessment. Any errors or omissions are our own.

General Psychometric Considerations

Users need to consider the technical adequacy and psychometric properties of mathematics assessments. We highlight key points here. Users can also consult the National Center on Intensive Intervention screening and progress monitoring tools charts for specific information on technical adequacy of some specific assessments at specific grade levels for specific uses.

Validity

Validity is the extent to which theory and evidence support the intended interpretations of scores for proposed uses (Standards for Educational and Psychological Testing, 2014). At a high level, instruments can only be supported by validity evidence for particular interpretations and uses with particular populations. When evaluating validity evidence of instruments under consideration, the relevant evidence should directly support your interpretations and uses.

Intended Interpretations and Uses

Users must consider the intended purpose of the instrument (i.e., for whom the instrument is intended to be used, the proposed purposes, and contexts in which the instrument is used) and whether this matches the intended use of the eventual user. For example, what is the intended age range? Is there an intended (sub-)population under consideration?

Content

An important source of evidence concerns the content of the questions comprising each measure. The greatest threats to validity include weak or limited content or construct coverage and construct-irrelevant variance. Is the instrument content supported by theory and prior research? Was there expert review to support the content coverage? Is there evidence that instrument items were reviewed by experts for sensitivity to gender, race/ethnicity, linguistic complexity, culture, or other characteristics relevant to your intended uses?

Consistency across Groups

Is there evidence that the measurement model parameters are consistent across groups, i.e., that there is no bias across gender, English learner (EL) versus non-EL students, students of different race or ethnicity? Evidence should, at a minimum, include differential item functioning (DIF) analyses, or better yet, measurement invariance or multigroup confirmatory factor analyses.

Consistency over Time

If the measure is intended for use longitudinally, such as in growth models, pre/post interventions, or value-added models, is there evidence that the measurement model parameters are consistent over time (longitudinal invariance)? Is there evidence of test-retest reliability?

Additional Considerations for Use by School Practitioners

Middle math practitioners use assessments for three purposes:

- *Screening* all students to identify those who may need additional support;

- *Diagnosing* specific skill strengths and deficits to inform instructional approach, either for a full class or for students needing supplemental and more intensive support; and
- *Progress monitoring* to assess rate of progress and make decisions about when a change in instructional approach is needed, for students receiving supplemental and intensive support.

Screening Assessments Are Not Diagnostic

As schools seek to reduce the amount of time students spend being assessed in favor of instruction, there is often a desire to select and use efficiently administered assessments for multiple purposes. Most commonly this occurs when screening assessments are adopted because of their purported alignment with curriculum and use for diagnostic purposes. However, screening and diagnostic assessments, by definition are constructed in opposition to one another. It is important for practitioners to use assessments aligned with their intended purpose.

Usability and Cost

Middle grades math assessments should be easy for educators to use. Automated score reports that are easy to understand for both teachers and parents are ideal in this case. While most assessments have software packages that support data entry, scoring and reporting, the cost for these additional supports varies, which is also of interest to school practitioners.

Decision-making "Stakes"

Because of the limited availability of assessments that have been rigorously validated for all subgroups of the population, among other concerns, it is important to consider the "stakes" of the decisions being derived from assessment data. High stakes decisions, such as determining whether a student is eligible for special education, require assessments that have robust technical adequacy evidence for the target population. Low stakes decisions, such as determining whether a student needs more practice with a particular skill, can be carried out using assessments with somewhat weaker technical adequacy evidence.

Accessibility and Accommodations

School practitioners may need to consider issues of accessibility and accommodations for students being assessed. Some students may require additional time, for example. For timed mathematics tests, however, applying an extra-time accommodation may compromise the integrity of the scores. Similarly, for mathematics tests in which the tested construct is students' ability to perform computations, calculators may not be appropriate. Educators will need information about accommodations when determining which accommodations can be provided to still maintain the integrity of the results.

Some assessments specify the accessibility features available (e.g., items are read aloud by the assessor and the examinee may respond with gestures such as pointing, which can permit scoring of responses from individuals who may have limited expressive abilities). When assessing students with low-incidence disabilities (e.g., visual impairment, hearing loss, significant cognitive impairment) and moderate to severe display, extensive modifications may be needed to permit established assessments to be accessible to these students, which may render assessment scores of little value.

Concluding Remarks

School practitioners strive to use reliable and valid middle grades mathematics assessments to provide information to guide instruction; monitor student learning; evaluate instruction; and identify academically vulnerable students, among others. However, while EdInstruments intends to provide descriptive information for the benefit of the user, few test publishers provide full access to information about their tests (e.g., technical reports). Currently, a prospective user of a middle grades math instrument needs to contact many of the publishers directly. Insufficient reporting and testing of psychometric properties may affect the accuracy of findings, leading to questionable study results and/or decisions in practice.