

The Reliability and Consequential Validity of Two Teacher-Administered Student Mathematics Diagnostic Assessments

Teachers need to assess their students' current level of mathematical understanding to provide appropriate interventions for students who are struggling. Several school districts in Georgia currently use two assessments for this purpose—the Global Strategy Stage (GloSS)¹ and the Individual Knowledge Assessment of Number (IKAN).² The IKAN is available in two formats: the IKAN Counting Interview, which is for students performing at lower levels on the GloSS, and the IKAN Written Assessment, which is for students performing at higher levels on the GloSS.

The Regional Educational Laboratory Southeast conducted an exploratory study to analyze Georgia-specific grade 1 and grade 3 data on interassessor reliability and consequential validity of the GloSS and IKAN assessments. The analysis was designed to inform the state's recommendations on whether and how districts could use these assessments. Interassessor reliability indicates whether two teachers using the same assessment to assess the same student on two occasions within a short period of time reach the same Stage Score, and in this study consequential validity is a gauge of how useful the teachers found the assessments.

Key findings

- *Interassessor reliability was adequate for the GloSS assessment.* Interassessor reliability for the GloSS was 92 percent across grades 1 and 3, when calculated using the plus-or-minus one method. It was 91 percent for grade 1 students and 93 percent for grade 3 students. A reliability above 90 percent is considered adequate when using the plus-or-minus-one agreement method, in which the two assessors must reach either the same Stage Score or a Stage Score one above or one below the Stage Score reached by the other.³
- *Interassessor reliability was adequate for the IKAN Counting Interview but not for the IKAN Written Assessment.* Interassessor reliability was 71 percent for the IKAN Counting Interview and 58 percent for the IKAN Written Assessment, when calculated using the exact agreement method. A reliability above 70 percent is considered adequate when using the exact agreement method, in which the two assessors must reach the same Stage Score.⁴
- *The consequential validity of the GloSS and IKAN assessments was mixed—teachers considered the data useful but also identified several concerns with training, administration, and scoring.* Teachers indicated that they found the data from the GloSS and IKAN assessments more useful than data from other school assessments for screening students and determining which students require intervention. Many teachers stated that the assessments helped them determine how to group their students for instruction. Teachers also identified areas of concern and remarked that the IKAN was inadequate for determining students' knowledge and that the GloSS was time-consuming to administer.

1. New Zealand Ministry of Education. (2012). *Global Strategy Stage (GloSS)*.

2. New Zealand Ministry of Education. (2011). *Individual Knowledge Assessment for Numeracy (IKAN)*.

3. Stemler, S. E. (2004). A comparison of consensus, consistency, and assessment approaches to estimating interrater reliability. *Practical Assessment, Research & Evaluation*, 9(4), 1–11. <https://doi.org/10.7275/96jp-xz07>.

4. Stemler, 2004.