Numeracy Counts Assessments

Selection Criteria for the KDE Approved List

The Kentucky Department of Education (KDE) has adopted the following minimum criteria for the selection of valid and reliable universal screening and diagnostic tools guided by the National Center on Intensive Intervention at American Institutes for Research Screening Tools Chart Rating System and the Institute for Education Sciences:

Districts are advised to perform internal reviews of the tests considered for selection to make assessment decisions that are an appropriate fit for their local context. These tools are intended to be administered as part of a comprehensive screening and assessment system within a multi-tiered system of supports

Technical Adequacy (Based on statistical analysis)	➢ Classification Accuracy – The area under the curve (AUC) must meet or exceed .70 or sensitivity must meet or exceed .70 and specificity must meet or exceed .70. ➢ Reliability – Internal consistency (often reported as Cronbach's alpha) or alternate form where applicable. Must meet or exceed .70. ➢ Validity – Predictive for universal screeners. Predictive and/or content validity for diagnostics. Validity must meet or exceed .60. ➢ Representative Sample – The sample used for psychometric analysis closely matches the characteristics of its population as a whole.
Administration Usability and Support	➢ **Brief** – Administration time 60 minutes or less; ➢ **Scoring criteria** (includes a minimum of the following): ○ Provides a percentile rank that compares the students' results to a nationally representative group; ○ Includes a pre-determined (external) benchmark score that represents levels of proficiency; ○ Provides percentile ranks and benchmark scores in narrowly defined skills (e.g., counting and cardinality, number and operations, etc.); ➢ **Able to be administered more than once per year** – Screening three times per year is necessary to evaluate effectiveness of the program; and ➢ **Vendor training and support.**

Definitions:

Benchmark/Cut-Score means a score on a screening measure that divides students who are considered potentially at risk from those who are not at risk.

Classification Accuracy means how well the scores on a screening assessment correctly identify students at risk versus those not at risk. Area under the curve (AUC) values closer to 1 indicate the screening measure reliably distinguishes students at risk from those not at risk.

Content validity broadly defined means how well the assessment measures what it is supposed to measure or how well a set of scores reflect the intended construct or domain being measured.

Mathematics diagnostic assessment means an assessment that identifies a student at risk of failure in mathematics or a student with major deficits in numeracy and other mathematical concepts and skills**.**

Predictive validity shows how well a score predicts a student's future performance on a later criterion test or benchmark.

Reliability means the consistency of a set of scores that are designed to measure the same thing within a single test and/or across multiple forms.

Representative Sample means the group used for psychometric analysis closely matches the characteristics of its population as a whole.

Sensitivity means the probability of correctly identifying a problem. Sensitivity rates help determine which tests can achieve a high rate of accurate classification.

Specificity means the probability of correctly identifying that there is not a problem. Specificity rates help determine which tests can achieve a high rate of accurate classification.

Universal screener for mathematics means a brief assessment given to all students within a grade level to assess the students' performance on the foundations of numeracy.