

Psychometric Measurement Characteristics 2017-2018

To ensure appropriate item difficulty the assessment shall use the following classical test theory model.

- (1) 15% - 25% of test items with a P-Value (item difficulty) between .10 - .30
- (2) 20% - 30% of test items with a P-Value between .31 - .50
- (3) 20% - 30% of test items with a P-Value between .51 - .70
- (4) 10% - 20% of test items with a P-Value between .71 - .90
- (5) Each assessment should have a minimum Cronbach's coefficient Alpha \geq .85

The following psychometric measurement characteristics shall be provided as evidence of test validity:

- (1) Each test item and form shall be demonstrably related to the skills and competencies in the Texas Essential Knowledge and Skills required for the determination of established performance levels for State accountability.
- (2) Overall reliability for each assessment shall be .85 (Cronbach's Alpha or KR20) or higher given the proposed uses of the test.
- (3) Overall reliability for each CR assessment shall be \geq .70 (70.0%) inter-rater exact score agreement given the proposed uses of the test.
- (4) Overall validity for each assessment shall address content, one factor/construct, consequential impact, and accommodations; provide evidence of validity in each of these areas.
- (5) Incorporate differential item functioning (DIF) analyses in test development and eliminate items with DIF that is statistically significant based upon psychometric calculations. Examine DIF for subgroups including gender, race/ethnicity and socio-economic status.
- (6) Assessments shall have a high degree of precision (conditional standard error of measurement) around the passing (proficient) cut-score.
- (7) The assessments shall have the capability of producing linear scales continuous across all levels of student attainment.