## **The Independent Evaluation of the Validity and Reliability of the Grades 3-8**

## **State of Texas Assessments of Academic Readiness (STAAR)**

## **Part I: Grades 5 and 8 Reading and Mathematics, and**

## **Grades 4 and 7 Writing**

**Executive Summary**

To comply with House Bill (HB) 743, the Texas Education Agency (TEA) contracted with the Human Resources Research Organization (HumRRO) to provide an independent evaluation of the validity and reliability of the State of Texas Assessments of Academic Readiness (STAAR) in grades 3-8. Established in 1951, HumRRO is an independent, nonprofit corporation engaged in research, development, and evaluation in the behavioral and social sciences, including education research and evaluation. HumRRO has past experience in conducting validity and reliability studies for other testing programs in other states.

The assessments subject to evaluation are grades 3-8 reading and mathematics, grades 4 and 7 writing, grades 5 and 8 science, and grade 8 social studies. Part I of the report contains the evaluation of grades 5 and 8 and reading and mathematics, and grades 4 and 7 writing. Part II of the report, to be posted in May 2016, will include the remaining grades 3-8 assessments.

HumRRO’s evaluation consists of several tasks: 1) provide empirical evidence for the validity of the STAAR scores (Task 1); and 2) provide empirical evidence for the projected reliability of the assessment (Task 2). For Task 3, HumRRO evaluated the procedures to build the STAAR assessments and whether they support the creation of valid and reliable assessments.

HumRRO finds support for the validity and reliability of the 2016 STAAR assessments.

* For Task 1, HumRRO identified evidence of the content validity of the assessments. Overall, the content of the 2016 forms aligned with blueprints and the vast majority of items were aligned with the TEKS expectations for grades 5 and 8 mathematics and reading, and grades 4 and 7 writing.
* Task 2 provided empirical evidence of the projected reliability and standard error of measurement for the 2016 forms as the conditional standard error of measurement (CSEM) estimates were all acceptable.
* For Task 3, the processes used to construct the 2016 tests and the proposed methods for scoring the 2016 test are consistent with industry standards and support the development of tests that measure the knowledge and skills outlined in the content standards and test blueprint. The processes allow for the development of tests that yield valid and reliable assessment scores.