

Salem, Jr., Joseph A., Ph.D., December 2014

Evaluation and Measurement

THE DEVELOPMENT AND VALIDATION OF ALL FOUR TRAILS (TOOL FOR REAL-TIME ASSESSMENT OF INFORMATION LITERACY SKILLS) TESTS FOR K-12 STUDENTS (348 pp.)

Co-directors of Dissertation: Jason Schenker, Ph.D.
Shawn Fitzgerald, Ph.D.

This study sought to determine whether the items that were in use for the TRAILS K-12 information literacy classroom assessments at the 3rd, 6th, 9th, and 12th grade levels could be used to develop a valid and reliable test for each grade level. To determine whether the assessments could be reengineered into valid tests of information literacy, the entire item banks for each grade level were administered to students recruited across the United States. The data gathered were then used to answer three research questions.

First, the study sought to determine whether the current item banks could be used to create an efficient and reliable test at each grade level. The Rasch model of Item Response Theory was used to develop each test based on overall scale reliability, item fit to scale, distractor function, lack of bias based on differential item functioning analysis, item difficulty spread, and content coverage. It was determined that reliable tests could be created at each grade level (3rd $N = 20$; 6th $N = 25$; 9th $N = 30$; 12th $N = 35$) with generally strong psychometric properties at the item level across all four tests.

The study then sought to gather evidence of construct validity through two methods. First, content experts at each grade level were presented each item in the draft test and were asked to rate the degree to which it measured its associated TRAILS objective. The results of this study identified items on each test for further examination

but found general endorsement at the item and scale level for objective measurement.

Additionally, the amount of reading and item difficulty were examined through a correlation study. No relationship was found on any of the tests, reducing the likelihood that reading affects item difficulty.

Finally, the study sought to determine the most generally agreed upon proficiency score for each test by using a modified bookmarking standard-setting procedure. This process utilized expertise at each grade level throughout the United States as well as the TRAILS project team and established proficiency scores for all four tests.