# Validation Report

October 2010

## Next Generation Student Assessment of Technological Proficiency: Psychometric Validation of Beta Version


Prepared by Hypothesi LLC

Educational Research, Evaluation & Consulting
www.hypothesi.com

Juliane Blazevski, PhD

# Introduction

This report describes the results of an external evaluation of the psychometric validity of the Next Generation Online Assessment of Student Technological Proficiency (beta version) conducted by Hypothesi LLC. In this paper, we describe the beta-testing process used to assess the validity of the assessment instruments, results of the psychometric validation, and next steps in the assessment development process.

# Next Generation Assessments (NGA)

The NGA student assessment is a criterion-referenced test, which reports the extent to which students have mastered skills and knowledge aligned to the 2007 ISTE Student NETS Standards, AASL Standards, and 21st Century Skills. As a criterion-referenced test, the NGA student assessment determines whether a minimum level of competency for technological knowledge as defined by these standards has been achieved. The beta version of the assessment contains 30 multiple choice items scored as correct or incorrect with each item receiving equal weight in the total score. Although scores on six sub-standards are also reported, the assessment is considered uni-dimensional in that each item is reflective of a general competency in technological literacy.

# Methodology

## Item Development.

To ensure content validity (face validity), items were written to align with curriculum standards set forth by nationally recognized entities in the field of technology literacy. A crosswalk mapping NGA student assessment on to these standards is available at www.nextgenerationassessments.com. To further ensure that the assessment has face validity in the schools, the items were scrutinized by a team of experts in the field (school technology directors, technology curriculum

specialists, and classroom teachers who actively use technology in their instruction) for relevance to technology instruction and student technology use at the middle school level.

Beta-Testing. The NGA student assessment was beta tested among a diverse sample of nine schools and approximately 1,400 eighth grade students in a Midwestern school district (16% African American, 19% Hispanic, 3% Asian, 62% White). The sample was fairly evenly split by gender (47% female). Standardized administration procedures were used during the online testing period at each school. Examination of the validity and reliability of the assessment tool as a measure technological proficiency was carried out by the external evaluation team (Hypothesi LLC). The validation process involved exploratory factor analysis and reliability analysis (using SPSS Version 17.0 software) and item response modeling, also referred to as Rasch modeling or IRT analysis (using ACER ConQuest Version 2.0 software).

## Results

Exploratory Factor Analysis. Exploratory factor analysis was used to determine the extent to which the items measure a uni-dimensional construct, specifically, technological proficiency, as opposed to a set of largely unrelated skills (uni-dimensionality is a necessary prerequisite for IRT). Analyses confirmed the one factor structure with the largest percentage of variance being explained by a single factor (with an eigenvalue of 4.83).

Reliability Analysis. Classical tests of reliability produced a Cronbach alpha of .75. This suggests that the internal consistency of the assessment is good—a student's scores on each of the items are highly correlated/consistent.

IRT Analyses. IRT analyses were used to examine the relative difficulty level of the items and their ability to discriminate between students of varying levels of technological proficiency. Results indicated that the assessment as a whole was effective at discriminating among students, with a separation reliability of .997, meaning that the test is capable of identifying a range of technological proficiency levels if they exist in the sample. The mapping of item difficulties to person ability estimates indicated that the assessment achieved adequate coverage of a wide range of proficiency levels, with items spread relatively evenly across the bell curve of ability estimates. That is, the assessment includes items that are most likely to be answered correctly by only those who are "high" in technological proficiency, as well as items that are likely to be answered correctly by those who are "average" or above in technological proficiency, and so on.

Only three of the thirty items on the assessment were identified as being potentially problematic (through an examination of fit statistics and item characteristic curves) in that the likelihood of answering these particular items correctly was not consistently related to a students overall technology proficiency score (ability estimate). These items with substantial "misfit" were reviewed to identify possible

issues with the wording of the questions and were revised for future administrations.

IRT analyses were also used to examine differential item functioning by gender and ethnicity. Ideally, items should perform similarly for both male and female students and for students of all ethnic backgrounds. For example, both male and female students with the same overall technology ability estimate should also have the same likelihood of getting a particular item correct. Results indicated that while there was differential functioning by gender and/or minority status for a small subset of items, in most cases, the difference was not very pronounced and did not occur across the full ability spectrum. Differential functioning typically occurred on items that were already under revision for other reasons (identified as "misfitting" in the general IRT analysis).

## Next Steps

Results of these and other preliminary analyses will be used to inform further development efforts. Next steps in the assessment development process involve the construction and pilot testing of parallel items for the student assessment. Items with poor fit or low discrimination power will be dropped the potential item pool. A set of overlapping items will be included in the pilot tests of post-test items so as to allow for equating of the test administrations and the creation of a common scale for item difficulties. Items with similar difficulty levels to the pre-test will be selected across the six ISTE standards to ensure that the post-test is well matched in terms of the overall

difficulty level and content so that changes in scores represent true gains/losses in technological proficiency rather than an artifact of test scaling or item content. Once the post-test form has been created, test–retest reliability analysis (administering the same assessment to the same set of students within a short timeframe) will also be conducted to further establish the validity of the assessment as an adequate tool for measuring gains in technological proficiency.