

Developing and Validating an Instrument to Measure College Students' Inferential  
Reasoning in Statistics: An Argument-Based Approach to Validation

A DISSERTATION

SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL

OF THE UNIVERSITY OF MINNESOTA

BY

Jiyoon Park

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Robert delMas, Adviser

Joan Garfield, Co-adviser

June 2012

## Abstract

The purpose of this study was to develop and validate an assessment to measure college students' inferential reasoning in statistics. This proposed assessment aims to help statistics educators guide and monitor students' developing ideas of statistical inference.

Within the two-stage cycle, the formative and summative stages, this study first built arguments for the use of assessment and score interpretations, and verified inferences made from those arguments. The five claims were used to examine the plausibility of the validity arguments: 1) The test measures students' level of statistical inferential reasoning in two aspects—informal statistical inference and formal statistical inference; 2) The test measures statistical inferential reasoning in the representative test domains; 3) The test produces scores with sufficient precision to be meaningfully reported; 4) The test is functional for the purposes of formative assessment; and 5) The test provides information about students' level of statistical inferential reasoning in the realms of informal and formal statistical inference.

Using a mixed-methods study design, different types of validity evidence were gathered and investigated. Three content experts provided their evaluation of the test blueprint and assessment, based on their qualitative reviews. For the revised assessment resulting from the experts' feedback, cognitive interviews were conducted with nine college students using think-aloud protocols, whereby the students verbalized their reasoning as they reached an answer. A pilot-test administered in a classroom provided preliminary information of the psychometric properties of the assessment. The final version of the assessment was administered to 2,056 students in 39 higher education institutions across the United States. For the data obtained from this large-scale

assessment, a unidimensional model in confirmatory factor analysis and the Graded Response Model in item response theory were employed to examine the arguments regarding the internal structure and item properties. The results suggest that the AIRS is unidimensional with appropriate levels of item difficulty and information. The pedagogical implications for the use of the AIRS test are discussed with regard to the areas where students showed difficulties in the domain of statistical inference.