Building a validity argument for a high-stakes test

For quite some time now language test researchers have accepted that test validity is not divisible into separate components but is a unitary, multi-faceted concept. A major concern that remains is how to build a validity argument. Messick (1989, 1994, 1996) identifies several sources of evidence that can be collected and used in a validity argument; these include, inter alia, content analysis, empirical item or task analysis, investigation of relationships between test scores and external criterion measures, research into differential performance across groups of test takers and test consequences, each of which can provide evidence of construct-irrelevant variance. Unfortunately Messick doesn’t tell us how we can actually take sources of construct-irrelevant variance into consideration when building a validity argument thereby enabling the generalizability of inferences about ability derived from language test scores to a broader target-language use domain. Bachman (2003) has suggested broadening the argument-based approach to validation to include test use, suggesting a two-part framework that includes both assessment validation and assessment use arguments. Bachman (2005, 2010) extends this to propose a specific two-part framework consisting of warrants and rebuttals, based on Toulmin’s (2003) argument structure, for building a validity argument.

In this talk I will begin by discussing both Messick’s and Bachman’s frameworks; I will then use data from previous research to demonstrate how they can be used to develop assessment use arguments for several high-stakes tests.