Abstract
Background: Large-scale educational assessments in the U.S. are used as policy tools for improving instruction and evaluating educational programs and the effectiveness of educators. Because of the high-stakes nature of their use, it is imperative that validity evidence based on testing consequences is obtained to support their multiple purposes. Method: A comprehensive review of the literature related to validity evidence for test use was conducted. Results: A theory of action for a testing program should be delineated. A theory of action reflects an interpretative and use argument and a validity argument for assessment programs and delineates the purposes and uses of the system as well as the outcomes of the system. The articulation of a validity argument allows for the accumulation of evidence not only for, but also against, intended score interpretations and uses. Conclusions: As described in the paper, for assessment and accountability systems that are intended to have an effect on both instruction and student learning, the consequences, both positive and negative, of the systems need to be evaluated.

Keywords
Validity, standards, evidence of testing consequences, test use