Measurement invariance is a fundamental assumption in item response theory models, where the relationship between a latent construct (ability) and observed item responses are of interest. Violation of this assumption would render the scale misinterpreted or cause systematic bias against certain groups of people. While a number of methods have been proposed to detect measurement invariance violations, they all require definition of problematic model parameters and respondent grouping information in advance. However, these “locating” pieces of information are typically unknown in practice.

As an alternative, this dissertation focuses on a family of recently-proposed tests based on stochastic processes of casewise derivatives of the likelihood function (i.e., scores). These score-based tests only require estimation of the null model (when measurement invariance assumption holds), with problematic subgroups of respondents and model parameters being identified in a factor-analytic, continuous data context. In this dissertation, I aim to generalize these tests to item response theory models for categorical data. The tests’ theoretical background and implementation are detailed. The tests’ ability to identify problematic subgroups and model parameters is studied via simulation. An empirical example involving the tests is also provided. In the end, potential applications and future development are discussed.