Washington State University

College of Education

David Alpizar

Will defend the dissertation

Date: March 30, 2021

Time: 9:30 A.M.

Location: Zoom by request

Faculty, students and the general public are encouraged to attend

A SYSTEMATIC REVIEW AND EVALUATION OF TESTLET RESPONSE THEORY MODELS

Chair: Brian French

Testlets are clusters of items that share the same stimuli. International and national assessments are often assembled with testlets. However, the measurement of the examinee's ability might be influenced by the presence of testlets (Yen, 1993), as responses may be dependent on each other. As a result, testlet data might violate the assumption of conditional independence for item response theory (IRT; DeMars, 2006). When test items meet this assumption, responses that are locally independent are solely explained by the examinees' ability measured by the test. That is, no additional variables influence the performance of such a test (Lord & Novick, 1968; Yen, 1993). However, this might not be the case for tests containing testlets. When performance is influenced by this additional variable in such a test, it is named the testlet effect. In this situation, traditional IRT models are inaccurate in assessing responses in tests. Conventional statistics (e.g., Yen's Q3) that are proposed to detect local item dependence might not be useful in identifying this assumption for testlet data. Innovative models are needed to improve the detection of local item dependence for testlet data. Testlet response theory (TRT) models may be able to assist in evaluating interrelated responses, detecting and

accounting for local item dependence. However, the TRT models lack (a) a systematic review about simulation studies investigating its performance, and (b) information about its performance on parameter estimation, model selection, and accuracy for detecting local conditional dependence under different conditions (i.e., number of items per testlet, local item dependence, and sample size). The purposes of this work were to review systematically the TRT literature and examine the performance of a TRT model under several simulated conditions. A systematic review and Monte Carlo study were conducted to accomplish these purposes. Collectively, the findings inform the field about the current state of the TRT literature, and situations encountered in practice when a TRT model is useful. Findings, limitations, and implications for future research are discussed.