

QUALITY CONTROL OF PSYCHOMETRIC ANALYSIS

Previous chapters in this technical report have presented the results of our psychometric analysis of the CLT exam at the item level, the subscore level, the subject level, and the overall test level. We performed all these analyses to collect validity evidence, assess the reliability of the test scores, evaluate and assure fairness, and ultimately to confirm that the CLT test measures what it is intended to measure, and with adequate precision.

Psychometrics is important in the whole process of test development. It plays an important role in the process of test development, beginning with creating test specifications and designing field tests, then in reviewing and analyzing test data. Further, test form construction needs input from psychometrics about the overall psychometric quality of the pulled new test forms. Finally, psychometrics contributes greatly to item analysis, differential item functioning analysis, scoring, calibration, equating, scaling, standard setting, and technical report writing. High quality psychometric work ensures high quality tests. Thus, quality control should be addressed in all these aspects of psychometric work.

The first step in quality control for psychometric work is to develop a process flowchart to standardize the procedure and the steps to follow. The second key element in quality control is to replicate the analysis with different teams. The third key step is to document what has been done, what has been found, and what action has been taken to ensure quality and take correction steps. The CLT has faithfully taken all three measures in the development of the CLT exam.

First, CLT identified all psychometric steps, procedures, and analyses at the very initial stage and during the entire process of CLT test development. This aspect was further reinforced when planning the contents of this technical report. We have had multiple rounds of discussions with the CLT staff to ensure that the right and defensible psychometric steps and procedures would be followed in developing the CLT. Further, through these discussions, we reached agreement as to the type of psychometric analyses that need to be conducted, how a specific psychometric analysis should be carried out, who is responsible for the task, how quality control should be performed on the output from each psychometric analysis, and who is responsible for quality control.

Second, to ensure that the outputs from psychometric analysis for CLT were error-free, our team of psychometricians conducted each analysis independently and compared the results afterwards. We discussed any discrepancy identified in the comparison of our results and explored further to identify the sources of any discrepancy.

This section documents what we did to perform quality control on the psychometric analysis results reported in this document, what we found, and what actions and/or correction steps we and CLT have taken to ensure quality.

Data cleaning is an important step in all psychometric analysis. When CLT delivered data for different CLT test forms to us, we ran basic analyses, such as descriptive statistics, on all variables, including the frequency of values or symbols a variable may take. We ran distributions of all variables to identify any outliers in each variable and in each student data record. The frequency table helped to identify any values or symbols which did not look reasonable. This reasonableness check of each variable and case laid a fundamental base for quality psychometric analysis. We further cleaned the data when needed. We included multiple sources of information in our quality control to ensure the cleanness

and the integrity of the data to be used for psychometric analyses. These include previous analysis data and results, previous technical reports, and other reports for special studies. We identified any potential issues and reported them to the CLT staff. Before beginning our analysis, we as a psychometric team held multiple rounds of discussions to ensure the quality of the data.

During the psychometric analysis stage, one of us led the psychometric task and ran the analysis, summarized the results, put them in the written document, and then shared it with the other, who was responsible for quality control. Two layers of checking were implemented, a within-person check and a between-person level of quality control. For the former layer of defense, each of us was responsible for running quality control of our own work before the other psychometric team member ran a quality control check.

Further, the quality control party independently conducted each analysis. We checked the reported values, and if the results matched, they were used in the final reporting. If a discrepancy was found, the two of us met and discussed potential sources for the discrepancy. We each did one or multiple rounds of screening of what was implemented in our own analysis. We first independently verified that the same standardized steps and procedures had been followed in the analysis. When we had followed standard procedures, we further explored other potential sources for discrepancy. These included different approaches to the exclusion rules applied in data cleaning, variable recoding, ways of dealing with missing values, changes that needed to be made when the analyses were conducted on different data sets, and any other special issues specific to an individual form. This process continued until the exact match was obtained. This within-person and between-person quality control policy established two layers of quality assurance.

As elaborated above, we communicated frequently and followed closely the standardization of the psychometric analysis procedure. We carefully checked and evaluated the reasonableness of the analysis results. Further, some psychometric analyses were automated to avoid human errors. Some analyses were conducted for the total and subgroups of the student population across different parallel forms. Automation helped to remove the errors in manual implementation of such analyses.

In general, to ensure quality control and to increase the efficiency and accuracy in the process of quality control, each step for conducting each psychometric analysis should be standardized and well documented so that replication by independent parties is feasible. Further, the action flowchart for the psychometric analysis, the steps to follow to ensure quality control of the analysis results, and the specifics that need to be controlled should all be streamlined, standardized, and documented as well.

Following these guidelines, we documented quality control results in writing. This helped track potential sources of discrepancy. Such documents are available upon request for relevant stakeholders of CLT who sign a confidentiality agreement. The psychometric analysis in Chapters 9-12 of this report is accompanied by tables with empirical data about the test items and reported scores, and the appendix presents further detailed information.

Hong Jiao, *Ph.D., University of Maryland, College Park*

Liru Zhang, *Ph.D., Independent Consultant*