An Evaluation of the Validity and Reliability of the Classic Learning Test (CLT) from Phase 1 of the Evaluation Project

Brian D. Ray, Ph.D.

March 22, 2016

Copyright © 2016 by Brian D. Ray
Introduction

The purpose of this project is to provide an initial evaluation of the validity and reliability of the Classic Learning Test (CLT). This initial evaluation is Phase I of three phases that should be conducted in this evaluation.

The CLT is a standardized test used for college admissions and as to measure the mastery of key subject areas in the high school (secondary) curriculum in the United States. The CLT “… invites students to wrestle with works of the greatest minds in the history of Western thought across literary and mathematical content. Rich material reflecting both theistic and secular perspectives benefit and enrich the student in the test taking process. Among standardized college entrance exams, the CLT provides the most accurate and rigorous measure of academic formation, accomplishment, and potential” (Classic Learning Initiatives, 2016a). Some colleges and universities use the CLT as an acceptable alternative to the SAT or ACT in their admissions procedures.

The publishers of the CLT (Classic Learning Initiatives, 2016c, March 2) describe it as follows:

The exam score for the CLT can be used for a variety of assessment and advising circumstances. High schools may use it to gauge student mastery of key academic subject areas. Colleges may use the results for admissions and placement in college courses. State and local governments may use scores for assessments and measurements of student progress. Other boards may use the CLT as a qualifier for scholarships and other financial assistance.

The CLT is especially useful as a measure of students’ readiness to engage in college-level material that requires not just comprehension of a topic, but the understanding of the broader connections and history within the Western intellectual tradition. Instead of testing students on a specific curriculum, the exam measures students’ ability to use logic and to engage with rich and complicated texts. Since high school curricula vary extensively, the CLT tests not precise academic subject areas, but rather a student’s broader understanding and ability to reason, as applied to both reading texts and mathematical problems. This in turn allows schools, colleges, and universities a broader way to measure academic progress across regions and populations.

Methods

Definitions

Classic Learning Test (CLT).
The CLT is a standardized test composed of three multiple-choice sections in three different academic disciplines (i.e., reading, writing, and mathematics). Each of the three sections is comprised of 40 items (questions) worth one point each. The CLT is scored on a scale of 0 to 120. Students receive three sub-scores (Reading, Writing, and Mathematics). There are no points freely given, nor are there partial points. Students are not penalized for incorrect answers. Tests are administered under standard, rigorous test conditions.
Validity.
In testing, validity refers to “… the appropriateness, meaningfulness, and usefulness of specific inferences made from test scores” (Gall, Gall, & Borg, 2007, p. 657). Or, validity is the “… degree to which accumulated evidence and theory support specific interpretations of test scores entailed by proposed uses of a test” (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999, p. 184).

Evidence of validity may come from various sources and there are many types of validity. One source of evidence comes from test content. Content-related evidence of test validity deals with the extent to which the content of the test matches the construct that the test (i.e., its authors) claims to measure. For example, do items in the CLT measure mastery of certain subject-area content or readiness to do well in college?

Convergent evidence is of the type that deals with the degree to which “… a sample’s test scores correlate positively with their scores on other measures that are hypothesized to measure the same construct” (Gall, Gall, & Borg, 2007, p. 198). For example, to what extent do CLT scores correlate with high school grade point average (HSGPA) or SAT or ACT scores, since HSGPA, SAT, and ACT scores purportedly measure important academic achievement and thinking skills in a valid manner?

A third kind of validity is face validity. This “… involves only a casual, subjective inspection of the test items to judge whether they cover the content that the test purports to measure” (Gall, Gall, & Borg, 2007, p. 196). This is one of the weakest sources of evidence of validity.

Reliability.
Reliability is the “… degree to which test scores for a group of test takers are consistent over repeated applications of a measurement procedure and hence are inferred to be dependable, and repeatable for an individual test taker; the degree to which scores are free of errors of measurement for a given group. (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999, p. 180).

Reliability “… refers to the degree to which measurement error is absent from the scores yielded by the test (Gall, Gall, & Borg, 2007, p. 200).

There are different approaches to examining test reliability. A common one is called internal consistency which “… is an approach to estimating test score reliability that involves examination of the individual items of the test” and one method of doing this is called rational equivalence (e.g., using Kuder-Richardson formulas or Cronbach’s alpha coefficient) (Gall, Gall, & Borg, 2007, p. 202).

Reliability is essential to validity but this does not mean that test scores with good reliability always yield valid score inferences (or that that test comprised of these items is a valid test).

SAT.
The SAT is a standardized test that is used for college admissions in the United States. It was first introduced in 1926. Its name has been changed several times over the years (e.g., Scholastic Aptitude Test, the Scholastic Assessment Test, SAT I: Reasoning Test, SAT Reasoning Test, and now simply the SAT).

ACT
The ACT test is a standardized test used for college admissions in the United States. It was first introduced in 1959. The ACT was originally an abbreviation of American College Testing.

Population and Sample

The CLT is designed to be used with high school (secondary school) and early college students in the United States. The sample of test takers (students) for this phase of the evaluation was drawn from a high school in a west coast state and another high school in an east coast state. The students come from households of varying socioeconomic status (i.e., low, middle, and upper) and both males and females are in the sample.

Statistical Analysis

IBM SPSS Statistics (version 22) is used for all statistical analyses (IBM Corporation, 2013). Alpha is set at .05 for all statistical tests, and significance levels are reported in the findings.

Limitations

This is Phase I of a three-phase evaluation of validity and reliability, as I have recommended to the test publisher. This evaluation is based on a limited sample size with limited information on the demographics of the test takers and on only two sources of validity. Readers should be careful about generalizing the findings.
Findings

Validity

There is convergent evidence of validity for the CLT. Table 1 presents the correlations between the CLT and SAT, SAT subscale scores, and CLT subscale scores. The correlation \( r = .86 \) between CLT and SAT scores is statistically significant, and it is a strong\(^1\) correlation.

**Table 1.**

*Pearson Correlation Coefficients between the CLT and other Variables*

<table>
<thead>
<tr>
<th></th>
<th>SAT</th>
<th>SAT reading</th>
<th>SAT math</th>
<th>SAT writing</th>
<th>CLT</th>
<th>CLT reading</th>
<th>CLT math</th>
<th>CLT writing</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLT</td>
<td>.855**</td>
<td>.747**</td>
<td>.761**</td>
<td>.753**</td>
<td>1</td>
<td>.750**</td>
<td>.860**</td>
<td>.800**</td>
</tr>
<tr>
<td>Correlation Sig. (2-tailed)</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>N</td>
<td>63</td>
<td>63</td>
<td>63</td>
<td>63</td>
<td>65</td>
<td>64</td>
<td>64</td>
<td>64</td>
</tr>
</tbody>
</table>

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

**Table 2** shows the correlations between the CLT and high school GPA and the SAT and high school GPA among these students. The correlation between GPA and the CLT is \( r = .69 \) and between the SAT and GPA \( r = .78 \) and both are statistically significant; both are strong correlations. The reader should note, again, that correlations above .50 are considered to be strong correlations in social science research (Cohen & Cohen, 1983; Shanahan & Lonigan, 2010).

1 Qualitative terms referring to the strength of correlations are weak \( r < .30 \), moderate/medium \( r = .30 \) to .50), and strong \( r > .50 \); Shanahan & Lonigan, 2010. See also, Cohen & Cohen, 1983.
Table 2.
Spearman Correlation Coefficients between the CLT and SAT and High School GPA

<table>
<thead>
<tr>
<th></th>
<th>gpa</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Spearman's rho</strong></td>
<td></td>
</tr>
<tr>
<td>sat Correlation Coefficient</td>
<td>.781**</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.000</td>
</tr>
<tr>
<td>N</td>
<td>61</td>
</tr>
<tr>
<td>clt Correlation Coefficient</td>
<td>.691**</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.000</td>
</tr>
<tr>
<td>N</td>
<td>61</td>
</tr>
</tbody>
</table>

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

There is also some content-related evidence of test validity for the CLT. The publishers of CLT hired experienced persons to construct a table of specifications along the lines of the subject areas of reading, math, and writing. They vouch that they did so.

Finally, there is evidence of face validity regarding the CLT. Following are comments by persons who have knowledge and expertise regarding what a test such as that which the CLT purports to be should contain or be able to do.

"The Classic Learning Test is a fairer test of college readiness than tests with passages mainly on science or contemporary political questions (for comprehension or editing). I use "fairer" in the sense that the texts and textbooks students will read in their college freshman year are likely written for readers with high school-level or adult reading skills addressing all the humanities, broadly speaking. I would anticipate the results of this kind of test to correlate better with (or to better predict) success in authentic college coursework than current college admissions tests." — Sandra Stotsky - Professor emerita at the University of Arkansas

"I teach freshman and sophomore English courses at a selective research institution. Within a few weeks, the students sort out into those who have read Great Books and possess literary-historical knowledge and those who don't. The first group handles assignments well, the second not so well. Many fall into the second group. Few of them make the first one, but they set a rigorous tone for the classroom. Long novels don't overwhelm them and Modernist poems don't baffle them. They've read enough of those kinds of works before to approach them without anxiety, and they have the background knowledge that saves them the trouble of those works' historical, social, and religious contexts. The Classic Learning Test is a fantastic preparation for the smaller group. Students who study for it will enter college with the right formation, and their teachers will notice it immediately. I am certain that if, ten years from now, we could collect data on the performance of CLT students after one year of college, we would find a stark demonstration of superior achievement." — Mark Bauerlein, Ph.D., is an English professor at
Emory University and the author of 2008 book, *The Dumbest Generation: How the Digital Age Stupefies Young Americans and Jeopardizes Our Future*

"The Classic Learning Test offers students and parents an opportunity to be tested on what they should be learning -- how to read and understand the best that has been written in our Western tradition. They will no longer have to face the dilemma of teaching what is important versus teaching to a secular test. The Institute for Catholic Liberal Education is grateful for the work Jeremy Tate and the CLI team have already done, and very confident that the final results will satisfy colleges, parents and students." --- Andrew T. Seeley, Ph.D., Executive Director, Institute for Catholic Liberal Education

"As the president of a college where more than half our students were home-schooled--and as an HSLDA member and home-schooling father myself--I wholeheartedly endorse the CLT Exam. This test has tremendous potential to restore virtues-based testing and traditional values to American education." --- Kevin Roberts, Ph.D., President of Wyoming Catholic College

“In addition to being more user friendly and convenient, I especially appreciated that the reading passages (in the CLT) prompted deep analysis, enabling students to more fully use their full range of critical thinking skills and insight.” --- Caroline Arnold, a senior at a high school academy in Maryland in 2016, who scored in the top 2% nationally on both the SAT and ACT

Reliability

Cronbach’s alpha on the 120-item CLT was calculated at .895. This is statistically significant (Table 3) and a strong or sound alpha relative to most standardized tests used in the United States.

**Table 3.**

<table>
<thead>
<tr>
<th>ANOVA Table for Cronbach’s Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Between People</td>
</tr>
<tr>
<td>Within People</td>
</tr>
<tr>
<td>People</td>
</tr>
<tr>
<td>Total</td>
</tr>
<tr>
<td>Grand Mean = .69</td>
</tr>
</tbody>
</table>

The split-half approach to calculating reliability was also used. Cronbach’s alpha for part 1 was .800 and for part 2 was .857; the Spearman-Brown coefficients for equal length and unequal length were .754 and .754; the Guttman Split-Half Coefficient was .741. All of these are statistically significant and strong coefficients (Table 4).
Table 4.
ANOVA Table for Split-Half Analysis

<table>
<thead>
<tr>
<th></th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between People</td>
<td>82.508</td>
<td>54</td>
<td>1.528</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within People</td>
<td>283.461</td>
<td>119</td>
<td>2.382</td>
<td>14.81</td>
<td>.000</td>
</tr>
<tr>
<td>Between Items</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>People</td>
<td>1033.455</td>
<td>6426</td>
<td>.161</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>1033.455</td>
<td>6426</td>
<td>.161</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1316.917</td>
<td>6545</td>
<td>.201</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1399.425</td>
<td>6599</td>
<td>.212</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Grand Mean = .69

This evaluator was also asked to review the SAT scores of the students who scored the highest on the CLT. It could be seen that students who scored 96 or above on the CLT scored 1230 or above on the SAT. It could also be seen that 69% of students who scored 96 or higher on the CLT had SAT scores that were the same or higher than their CLT scores would project according to the CLT publishers (Classic Learning Initiatives, 2016b). That is, the CLT projected SAT scores are fairly close to reality or conservative.

Conclusions and Recommendations

This evaluation has revealed preliminary solid evidence that the CLT is a valid and reliable measure of high school (secondary school) mastery of key academic subject areas and of college readiness and course placement.

The first-phase and preliminary analyses conducted in this project find firm evidence of convergent validity. The correlations between CLT scores and both SAT scores and high school GPA are both statistically significant and very high and strong.

This evaluation also found some evidence of content-related validity and plenty of evidence of face validity.

This appraisal also found preliminary solid evidence that the CLT test items have reliability. The measure of reliability, internal consistency, is statistically significant and strong and sound compared to those associated with most standardized tests used in the United States (e.g., SAT, ACT, standardized academic achievement tests for grades K to 12).

In conclusion, this evaluation has found solid and promising evidence of both validity and reliability regarding the CLT. This is the first phase of a proposed three-phase evaluation so all readers should be cautious about generalizing the findings of this project. More analyses and
evaluation with more data should be done to produce more complete and robust information about the usefulness, validity, and reliability of the CLT.

References


# # # end of validity and reliability report, Phase I, by Brian D. Ray # # #