The Predictive Validity of Florida’s CLAST (College Level Academic Skills Test) for Predicting GPA: A Revisal

David Walker
Northern Illinois University
1. Determining if the predictive validity of the CLAST with GPA has, over the course of time since the preliminary study, increased, decreased, or persisted.

2. It is valuable to understand if the test has predictive validity qualities because it is a significant component of the undergraduate experience in identifying potential academic skill efficacy and/or inadequacy.

3. The degree to which we can predict group membership based upon a set of predictor variables: How accurately can we predict group membership in either “Average” GPA or “Above Average” GPA from the subjects’ four CLAST scores?
An achievement test that was first implemented in 1984 as part of Florida’s educational accountability measures.

Comprised of four subtests in mathematics, reading, writing, and essay that purport to measure students’ academic proficiency in the areas of computation and communication.

To receive an associate in arts degree from any of Florida’s 28 public community colleges or obtain admission to upper-division status in any of Florida’s 11 public, four-year institutions, a student must pass all subtests of the CLAST or receive one of many exemption options. Subtests can be taken as many times as needed.
Subtests

1. Math.
2. Reading.
3. English.

Scores range from 200 to 400 points. The present cut scores for passing is 295.


Readers rate each essay test. Scores range from 2 to 12 points. The current minimal score is a 6.
The four predictor variables were the subtests on the CLAST: mathematics, reading, English, and essay.

The criterion variable was undergraduate GPA, where 4.00 = A, 3.00 = B, and 2.00 = C (note: could not have been in the study as a recent graduate if GPA was < 2.00).
Reliability

- Kuder-Richardson 20 method, with SEM, .83 (3.03), .84 (3.02), and .86 (3.07) for mathematics; .74 (2.74), .83 (2.38), and .77 (2.37) for reading; and .71 (2.21), .67 (2.17), and .68 (2.21) for English.

- Inter-rater reliability derived from a six-point holistic scoring rubric: .86, .85, and .86 for topic 1 and .86, .87, and .83 for topic 2.
Subjects were identified who had complete data (GPA, CLAST scores, gender, ethnicity, age).

Random sample of 750 recent graduates from a regional university in Florida who were conferred with baccalaureate degrees in academic years 1997-98 (n = 250), 1998-99 (n = 250), and 1999-2000 (n = 250).
65.3% females and 34.7% males.
68% Caucasians made up 68% and 32% people of color comprised (e.g., African American = 13.1%; Latino/a = 10.9%; Asian/Pacific Islander = 3.6%; Non-Resident = 3.6%; and American Indian/Native Alaskan = .8%).
Mean age was 29 years.
Results: Descriptives

All of the subgroups (i.e., male/female and Caucasian/students of color) performed about evenly on the four subtests and on GPA.
Results: Correlations

- For the entire sample, writing and reading, as were found in the initial study, still had the highest linear relationships with GPA, where $r = 0.222$ (90% CI = 0.165, 0.227) and 0.180 (CI = 0.119, 0.239), respectively.

- When the subgroups were analyzed, the correlations between the subtests and GPA remained in the same low area as found with the overall sample.
The four predictor variables were entered into the regression model in the following order: writing, reading, essay, and mathematics.

The regression results indicated that the predictor variables were direly insignificant at explicating the variance in GPA ($R^2 = .063; 90\% CI = .034, .090$ and adjusted $R^2 = .058$).

Writing was the best predictor of GPA (.152) followed by reading (.086).
As a classification rule, equal prior probabilities external to the sample were established at .50/.50, which measured the probability of population membership in either group and equal cost of misclassification for the two populations.

The mean GPA for “Average,” with a range between a “C” (i.e., 2.00) and “B-“ (i.e., 2.99), was 2.59.

The mean GPA for “Above Average,” with a range between a “B” (i.e., 3.00) and “A” (i.e., 4.00), was 3.50.

The cut point used for the two groups was 3.00.
How well does this model apply to subjects from the population or its generalizability?

“Average” GPA = 170 hits (63.7%; 90% CI for a Binomial Parameter = .585, .686) and 97 (36.3%; CI = .314, .415) “Above Average” or misses.

“Above Average” = 223 misses (46.2%; CI = .424, .500) and 260 (53.8%; CI = .500, .576) “Above Average” or hits.

Accuracy: 63.7% correct for “Average” GPA and 53.8% correct for “Above Average” GPA.

Total precision = 57.3% accuracy.

The model correctly classified a little over half of the cases, with a total group error rate estimate of 42.7%.
To determine the percentage correctly classified exceeding chance. The “Average” group’s $ES(1) = .274$.

- The “Above Average” group’s $ES = .076$.
- The total model $= .146$.
- These effect sizes are considered small in terms of their ability to measure proportional reduction in error, meaning that the total model had roughly 15% less misclassifications than would have occurred if just classified by chance.
For future institutional decisions related to students’ academic success, the PDA model verified that using the four CLAST subtest scores to classify students as either “Average” or “Above Average” was not accurate enough across all groups, or for each group, and its hit rate, although better than chance, was not considerably so.

The four subtest scores did not appear to predict accurately group membership in either “Average” or “Above Average” GPA, nor did the scores estimate effectively academic success in terms of predicting GPA. Score use by institutions as a measure of educational accountability, specifically in the instance as a mode to estimate academic success, might need rethinking.