

Testing the Factorial Equivalence of the Collegiate Learning Assessment Performance Task Diagnostic Instrument across Lower Class and Upper Class Predominantly Black College Students

Maurice Y. Mongkuo^{1*}and Meya Y. Mongkuo²

¹Department of Government and History, Fayetteville State University, USA

²Department of Psychology, East Carolina University, USA

Research Article

Received date: 03/02/2017

Accepted date: 22/02/2017

Published date: 28/02/2017

***For Correspondence**

Mongkuo MY, Department of Government & History, Fayetteville State University, U.S.A,
Tel: 910-267-5448.

E-mail: mmongkuo@uncfsu.edu

Keywords: Factorial equivalence, Collegiate learning assessment, CLA, Task diagnostic instrument, CLAPTDI, Confirmatory factor analysis, Multi-group invariance, AMOS

ABSTRACT

Objective: This study was aimed at determining the external validity of the psychometric properties of a two-factor Collegiate Learning Assessment Performance Task Diagnostic Instrument (CLAPTDI) for use in assessing learning skills among predominantly black college students. The construct validity of the two factors CLAPTDI had been established in a previous study exploratory and confirmatory factor analyses. Establishing the external validity involved conducting a multi-group test of the measurement instrument's factorial scores equivalence across panels of lower class and upper class students from a predominantly black college.

Method: The study relied on a strict test of equivalence categorization by focusing on tests for invariance across the two groups with respect to factor loadings, intercepts, and error factor loadings by estimating the difference in chi-square goodness-of-fit statistic and comparative fit index (CFI). Sets of measurement and structural parameters were put to the test in a logically ordered and increasingly restrictive manner.

Results: The analyses found that the CLAPTDI scale's factorial measurement structure was invariant across lower class and upper class PBC students.

Conclusion: The collegiate learning assessment performance task diagnostic instrument with two latent factors and five observed variables is a valid measurement scale for assessing the level of analytic reasoning and problem solving learning among predominantly black college students.

INTRODUCTION

The Collegiate Learning Assessment Performance Task Diagnostic Instrument (CLAPTDI) is an assessment tool used nationwide in the United States to measure the contribution of an educational institution to learning gained by its students [1]. The CLAPTDI measures a student's ability to perform cognitively demanding tasks from which quality of responses are scored on a 4-point scale ranging from 0 = Not Attempted to 4 Mastering [1]. The tool is considered to be better than standardized test scores, grade point averages and course test scores in assessing students' learning outcomes [2,3], as well as effective in promoting a culture of evidence-based assessment in higher education [2,4]. Unlike traditional assessment instruments that rely on multiple choice items to measure the responses of the study participants,

CLAPTDI utilizes open-ended prompts requiring constructive responses to measure high order thinking skills such as, critical thinking, analytic reasoning, written communications, and problem solving [1,4]. However, as any diagnostic instrument, the utility of the CLAPTDI as a gauge in determining student learning depends on its validity (both internal and external) in measuring student learning.

While the CLA seems quite promising in assessing student learning, some scholars have raised a number of methodological issues about this approach and the CLAPTDI's ability to effectively capture a student's learning [2,5-7]. Perhaps the most serious issue involved the validity of the psychometric properties used to measure the major constructs (i.e., critical thinking, analytic reasoning, written communications, and problem solving) the CLAPTDI. An extensive review of the literature reveals that despite its widespread use in colleges and universities across the United States, very few studies have focused attention to validating the CLAPTDI. To be sure, only one study to date has examined the psychometric properties of the CLAPTDI [8]. The study was, however, limited to using Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) to assess the internal validity or stability of the key latent constructs of the CLAPTDI. For the instrument to have a broader universal use, it is necessary that the validation process be extended to establishing the external validity of its psychometric properties. Byrne [9] suggest that to do so require testing the factorial equivalence of the CLAPTDI across groups.

The purpose of this study was to address this external validity void by extending the CLAPTDI validation process to performing a multi-group factorial equivalence test of the CLAPTDI. In particular, this study extends Mongkuo and colleagues [8] CLATDI validation process to determining the extent to which the instrument is equivalent across lower class and upper class predominantly black college students. To do so, the study addressed the following research question: Is the factorial structure of the CLAPTDI scale equivalent across lower class and upper class predominantly black college student? Providing empirically-grounded answer to this question generally involves testing hypotheses related to multi-group invariance of a single measurement scale across two different panels of PBC students. According to Jöreskog [10] this test for equivalence begins with a global test of the equality of covariance structures across the groups of interest. The null hypothesis (H_0) for the test is $\Sigma_1 = \Sigma_2 = \dots = \Sigma_G$, where Σ is the population variance-covariance matrix, and G is the number of groups. Rejection of the null hypothesis argues for the non-equivalence of the groups and, thus, for the subsequent testing of increasingly restrictive hypotheses in order to identify the source of non-equivalence. On the other hand, if H_0 cannot be rejected, the groups are considered to have equivalent measurement and covariance structures and, thus, tests for invariance are not needed.

METHODOLOGY

Research Design

The study uses a pre-experimental one-shot case study design [11]. Schematic representation of the design is as follows:



Figure 1. Pre-experimental one-shot case study design

Where X, is exposure of a predominantly black student to high school and/or college core curriculum courses. O₂ is the level of a student's learning abilities (that is, critical thinking/analytic reasoning, problem solving, persuasive writing, and writing mechanics).

Participants and Procedure

Participants in the study included a purposive, convenience sample of students attending a predominantly black college in south-eastern United States. The college has a population of 5,567 students enrolled. A breakdown of the population by race/ethnicity shows that approximately 70% is black or African American, 17% is Caucasian, 4% is Hispanic, 1% is Native American and 4% is other racial/ethnic groups.

The age distribution of the student population consists of 55% in the age range of 17-25 years old, 31% aged 26-40 years, and 14% is over 40 years. Most of the students (68%) are females, while 32% is males. The distribution of the population by academic class shows that 19% is freshmen, 15% is sophomore, 18% is junior, 32% is senior, and 11% is post-bachelors. Most of the students (66%) attending the university are enrolled as full-time students, while 34% are part-time. The distribution of the student population by academic class shows that 43% are lower class (Freshman and Sophomore) students and 61% is upper class (juniors, seniors and graduates) students.

However, the CLA conducted at this institution does not focus on the level of student learning by demographics beyond academic class.

Instead, the institution requires freshman; rising junior, and senior students to take the CLA as an integral part of the overall university strategic plan for determining the level of student learning at each academic level. In particular, all incoming freshmen are required to take the CLA as a baseline measure of learning ability upon entering the university.

Those same students are tested again as rising juniors to assess any increase in skill levels and ability. Finally, that same group of students is tested as graduating seniors so that the test scores at all levels can be compared to ensure that program learning outcomes are being met.

The data generated from the CLA is used by university administrators to identify areas of learning strengths or deficiencies in order to designing effective corrective action plan to improve or maintain acceptable retention and graduation rates. Based on this requirement, the population for this study was delimited to students who had taken to CLA during their freshman, junior and senior years only. University records show that in academic year 2013-2014, a total 764 students had taken the CLA in all three years. The participants in this study consisted of a random sample of 320 students obtained from the University's CLA Data File who took the CLA Performance Diagnostic Task during their freshman, sophomore, junior, and senior years. After data screening and deletion of cases with excessive missing values, the actual sample used for the study was 253 students, representing 79% participation rate.

CLA Measures

The CLA Performance Task Diagnostic Instrument (CLAPTDI) consisted of eight items aimed at measuring four interrelated higher order thinking abilities or skills – critical thinking/ analytic reasoning, problem solving, persuasive writing, and writing mechanics [1].

Critical thinking/Analytic reasoning: Critical thinking/ analytic reasoning skill was measured by the following two items scored on a 4- point scale ranging from 0=not attempted to 4=mastery: (a) How well does the student assess the quality and relevance of evidence in terms of determining what information is or is not pertinent to the task at hand, distinguishing between rational claims and emotional ones, facts from unsupported opinion, recognizing the ways in which the evidence might be limited or compromised; sporting deception and holes in the argument of others, and considering all sources of evidence; and (b) How well does the student analyse and synthesize data and information, including; presenting his/her own analysis of the data or information rather than “as is”; recognizing and avoiding logical flaws such as distinguishing correlation from causation; breaking down the evidence into its component parts; drawing connections between discrete sources of data and information; and attending to contradictory, inadequate or ambiguous information.

Problem solving: Problem solving skill was measure by two items scored on a 4-point scale ranging from 0=not attempted to 4=mastery:

- (a) How well does the student form a conclusion from his/her analysis, including, constructing cogent arguments rooted in data/information rather than speculation/opinion, selecting the strongest and most relevant set of supporting data, avoiding overstated or understated conclusions, and identifying holes in the evidence and subsequently suggesting additional information that might resolve the issue;
- (b) How well does the student consider other options and acknowledge that his/her answer is not the only perspective, including, recognizing that the problem is complex with no clear answer, proposing other options and weighing them in the decision, considering all stakeholders or affected parties in suggesting a course of action, and qualifying responses and acknowledging the need for additional information in making an absolute determination.

Persuasive writing: Persuasive writing was measured by two items scored on a 4-point scale ranging from 0=not attempted to 4=mastery: (a) How effective is the writing structure in terms of logical and cohesive organization of the argument, avoidance of extraneous elements in the argument's development, and presentation of evidence in an order that contributes to a persuasive and coherent argument;

- (b) How well does the student defend the argument in terms of effective presentation of the evidence in support of the argument, drawing thoroughly and extensively from available range of evidence, analysis of the evidence in addition to simply presenting it, and considering counter-arguments and addressing weaknesses in his/her own argument.

Writing mechanics: Writing mechanics was measured by two items scored on a 4-point scale ranging from 0=not attempted to 4=mastery: (a) How clear and concise is the argument in terms of clear articulation of the argument and the context for the argument, correct and precise use of evidence to defend the argument, comprehensible and coherent presentation of evidence, and citation of sources correctly and consistently; (b) What is the quality of the student's writing in terms of using vocabulary and punctuation correctly and effectively, demonstrating a strong understanding of grammar, using sentence structure that is basic or more complex and creative, using proper transition, and structuring paragraphs logically and effectively.

Data Analysis: The statistical test for factorial and structural invariance or equivalence involved a series of hierarchical analyses using AMOS 24.0 [12]. Following Joreskog [10] guidelines, the test began with a determination of the CLAPTDI baseline model (with no between-group constraints) for each group of PBC students separately. The model is one that best fit the data in terms of parsimony and substantive meaningfulness [9].

Generating this best fitting model was accomplished by performing a first-order confirmatory factor analysis (CFA) of the four-factor CLAPTDI. Following completion of this preliminary task, test for the equivalence of parameters were conducted across the two groups of students at each of several increasingly stringent levels beginning with the scrutiny of the measurement model.

In particular, patterns of factor loadings for each observed measure was tested for its equivalence across the groups. Once it was known which measure was group-invariant, these parameters were constrained equal while subsequent tests of the structural parameters were conducted. As subsequent new sets of parameters were tested, those known to be group-invariant were cumulatively constrained equal. Thus, the process of determining non-equivalence of measurement and structural parameters of the CLAPTDI parameters across groups involved the testing of a series of increasingly restrictive hypotheses.

If the model fit the data well for both groups of young adults, it will be maintained as the hypothesized model in the test for equivalence across the two groups of young adults. If the model exhibit a poor fit to the data for each group of young adults, it will be modified accordingly and become the hypothesized multi-group model under test.

Because the estimation of the baseline model involves no between-group constraints, the data was analysed separately for each group. However, in testing for invariance, equality constraints was imposed on particular parameters and, thus allowing for the data for the two groups to be analyzed simultaneously to obtain efficient estimates. In essence, the model being tested here, commonly termed the configural model [13], is a multi-group representation of the baseline models because it contained the baseline models of lower class and upper class students within the same file. Hence, we tested for configural invariance. Because no equality constraints were imposed on any parameters in the model, no determination of group differences related to either the items or the factor covariance could be made. Such claim was derived from subsequent tests for invariance. In testing for invariance, the fit of the configural model provided the baseline value against which all subsequently specified invariance models were compared.

Given that this model comprised the final best-fitting baseline model for each group, it was expected that results will be indicative of a well-fitting model. However, Byrnes [9] notes that despite evidence of good fit to the multi-sample data, the only information that we have at this point of the test is that the factor structure is similar, but not necessarily equivalent across groups. Given that no equality constraints are imposed on any parameters in the model, no determination of group differences related to either the items or the factor covariances could be made. Despite the multi-group structure of this and subsequent models, analyses yield only one set of fit statistics for overall model fit. Using ML estimation, the χ^2 statistics were summative and, thus, the overall χ^2 value for the multi-group model was equal to the sum of the χ^2 values obtained when the baseline model was tested separately for each group of students [9].

A number of indices were used to evaluate the goodness of fit of the two-factor orthogonal CLATPDI configural model. The models absolute fit was assessed using chi-square (χ^2) statistics, with low χ^2 considered good fit [10]. Incremental fit was evaluated using the Root Mean Square Errors of Approximation (RMSEAs) with a value less than 0.06 indicating a relatively good fit, along with Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) with values of .95 or greater considered desirable [10,14-17]. Assessing invariance involved comparing the goodness-of-Fit for the configural model to the constrained measurement and structural model, with evidence of non-invariance claimed if the χ^2 difference ($\Delta\chi^2$) value is statistically significant [9,10] and/or the CFI difference (ΔCFI) is less than 0.01 [10,19]. Assessing multi- group invariance involved comparing the goodness-of-Fit for the configural model to the constrained measurement and structural model, with evidence of non-invariance claimed if the χ^2 difference ($\Delta\chi^2$) value is statistically significant [10,11] and/or the CFI difference (ΔCFI) is less than 0.01 [19].

Normality of the distribution of the variables in the model was assessed by Mardia's [20,21] normalized estimate of multivariate kurtosis with a value of 5 or less reflexive of normal distribution.

Multivariate outliers were detected by computation of the squared Mahalanobis distance (D2) for each case with D2 values standngs distinctively apart from all the other D2 values as indicative of an outlier [22,23].

RESULTS AND DISCUSSION

Preliminary first-order CFA of the CLATPDI identified a hypothesized model with two latent constructs: analytic reasoning/ critical thinking with three observed variables and problem solving with two observed variables, respectively (Figure 2). Table 1 displays the goodness-of-fit test results for the CLATPDI multi-group invariance.

The results of the multi-group model testing for the configurable invariance reveal the χ^2 value to be 19.199 with 8 degrees of freedom.

The CFI and RMSEA values are 0.991 and 0.022, respectively. From this information, we conclude that the hypothesized multi-group configurable model of the CLATPDI structure is well fitting across lower class and upper class PBC students. The result of the goodness-of-fit statistics for the measurement model shows the fit to be fairly consistent with the configurable model. (CFI=0.982; RMSEA=0.022).

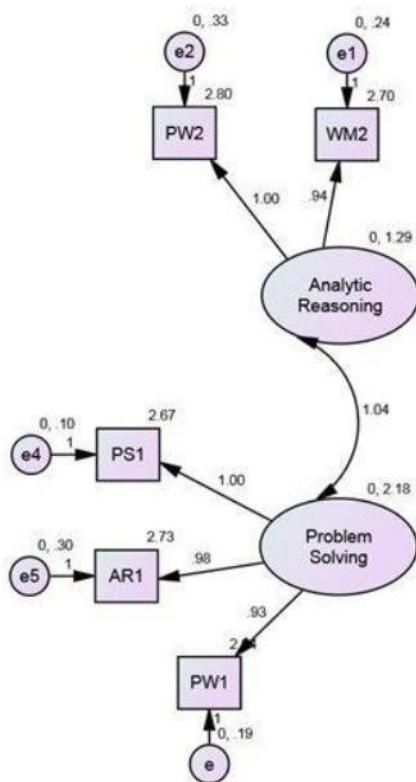


Figure 2. Hypothesized model of 5-item CLAPTDI structure for lower class and upper class students.

The test for factor loadings invariance reveals a non-significant χ^2 difference between the configurable model and the measurement model ($\Delta\chi^2(8)=19.899$, $p<0.01$), and a CFI difference of 0.001. Thus, these results provide evidence of factor invariance between lower class and upper class PBC students for the measurement model of CLATPDI scale. The results of the test for structural invariance shows the factor covariance to be equivalent across lower class and upper class HBCU students ($\Delta\chi^2(6)=21.236$, $p<0.01$), ΔCFI 0.01).

Table 1. Summary of goodness-of-fit statistics for tests of CFA multi-group invariance.

Model Description	Comparative Model	χ^2	df	$\Delta\chi^2$	Δdf	Sig	CFA	ΔCFA
Phase I: Baseline model fit for each academic class student								
Lower Class Students		24.836	4			0.001	0.984	
Upper Class Students		8.102	4			0.088	0.975	
Phase II: Factorial invariance across student academic class groups								
1.Configural Model:		19.899	8			NS	0.991	

All factor Loadings								
Constrained equal	2 versus 1	22.610	11	2.711	3	NS	0.992	0.001
3. Structural Model:								
Model B with								
Covariance among								
AR and PS								
Constrained equal	3 versus 1	41.137	14	21.238	6	S	0.980	0.01
Notes: $\Delta\chi^2$ = different value between models; Δdf = difference in number of degrees of freedom between models	ΔCFI = difference in CFI values between models; AR = Analytic reasoning; PS = Problem Solving							

Root Mean Square Error Approximation (RMSEA): Lower class baseline model: 0.035, Upper class baseline model: 0.027, Model 1 (Configural): 0.022, Model 2 (Measurement): 0.024, Model 3 (Structural): 0.023.

As reported in Table 1, the multi-group test of the CLATPDI yielded evidence of factorial invariance of the measurement model (Measurement Model: $\Delta\chi^2 (3)=2.711$, $p>0.01$, ΔCFI of 0.001); and the Structural model ($\Delta\chi^2 (6)=238$, $p>0.01$, ΔCFI of 0.01).

CONCLUSION

This study was aimed at assessing the factorial invariance of the psychometric measurement and structure of the Collegiate Learning Assessment Performance Task Diagnostic Instrument (CLAPTDI) across lower class lower class and upper class students attending a predominantly black college (PBC). The study was second in a series of studies aimed at developing a valid measurement scale for assessing the contribution of the college curriculum to students' analytic reasoning, critical thinking and problem solving skills. In a first step in establishing the external validity of the CLAPTDI involved conducting a multi-group test the equivalence of the factorial structure of the measurement scale across to panels of students: lower class and upper class students. In testing for invariance across the groups, sets of parameters were put to the test in a logically ordered and increasingly restrictive manner. The study relied on Meredith's [24] strict test of equivalence by focusing on tests for invariance across the groups with respect to factor loadings, intercepts, and error factor loadings. The invariance of these parameters across the two groups of students was tested by estimating the chi-square goodness-of-fit statistic and comparative fit index (CFI). The analyses found that the CLAPTDI scale's factorial measurement structure were invariant across lower class and upper class predominantly black college students, thus, confirming the external validity of the scale.

This study had a limitation that should be noted. The study did not cross-validate the CLATPDI by replicating the factorial structure of the CLATPDI across independent samples drawn from the same predominantly black college student population. Future studies should extend the factorial invariance test to cross-validation of independent samples of predominantly black college students. With regards to contribution to future research, it is important to note that while this study has established the validity of the CLATPDI for use in assessing student learning in a predominantly black college setting, preliminary confirmatory factor analysis reduced the number of constructs of the original CLATPDI and their corresponding observed variables from five latent constructs to two valid latent constructs. We named the first latent construct "analytic reasoning/problem solving" measured by three observed variables (drawing conclusions, evaluating evidence, and persuasive writing), and named the second construct "critical thinking" measured by two observed variables (written mechanics and persuasive writing). All the observed variables or items of the 2-factor CLATPDI were scored on a 4-point scale ranging from 0=not attempted to 4=mastery. Hence, we recommend that the future assessment of learning among predominantly black college students using the CLATPDI should be delimited to determining the level of critical thinking/analytic reasoning and problem solving.

REFERENCES

1. AAC&U. Liberal education outcomes. Association of American Colleges and Universities, Washington, DC. 2005.
2. AASCU. Value-added Assessment Perspectives. American Association of State Colleges and Universities, Washington, DC. 2006.
3. Arbuckle JL. AMOS: Analysis of moment structures. *The American Statistician*. 1989;43:62-66.
4. Arum RJ and Velez M. Learning to reason and communicate in college: Initial report of findings from the CLA longitudinal study. The Science Research Council, New York. 2008.
5. Banta TW and Pike GR. Revisiting the blind alley of value-added. *Assessment Update*. National Survey of Student Engagement, Bloomington, IN. 2007.
6. Niels J. Introduction to Structural Equation Modeling Using SPSS and AMOS. Sage Publication, Thousand Oak, CA. 2010.
7. Timothy A. Confirmatory Factor Analysis for Applied Research. Guildford Press, New York. 2006.
8. Cheung GW and Renwsvolt RB. Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*. 2002;9:233-255.
9. No Authors Listed. Diagnostic Scoring Faculty Handbook. Collegiate Learning Assessment, New York. 2008.
10. Hair JF and Black WC. Multivariate Data Analysis. Pearson Prentice Hall, Upper Saddle River, NJ. 2010.
11. Hornes JL and Mason R. When is Invariance not Invariant: A Practical Scientist's Look at the Ethereal Concept of Factor Invariance? *Psychologist*. 1983;1:179-188.
12. Hu L and Bentler PM. Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional criteria versus New Alternatives. *Structural Equation Modeling*. 1999;6:1-55.
13. Stephen M and William B. Handbook in Research and Evaluation. EdITS Publishers, California. 1997;4:71-72.
14. Klein S and Benjamin R. The Collegiate Learning Assessment: Facts or Fantasies. *Evaluation Review*. 2007;31:415-439.
15. Leedy PD and Ormod JE. Practical Research: Planning and Design. Pearson Publishers, New Jersey. 2010.
16. Mardia KV. Measures of Multivariate Skewness and Kurtosis with Applications. *Bikometrika*. 1970;57:519-530.
17. Mardia KV. Application of Some Measures of Multivariate Skewness and Kurtosis in Testing Normality and Robustness Studies.
18. Marsh HW and Wen Z. In Search of Golden Rule: Comments and Hypothesis Testing Approaches to Setting Cutoff Values for Fit Indexes and Dangers of Overgeneralizing Hu and Bentler's Findings. *Structural Equation Modeling*. 2004;11:320-341.
19. Meredith W. Latent Curve Analysis. *Psychometrika*. 1993;55:107-122.
20. Mongkuo MY and Walsh K. Initial Validation of Collegiate Learning Assessment Performance Task Diagnostic Instrument for Historically Black Colleges and Universities. *British Journal of Education, Society, and Behavioral Science*. 2013;3:282-299.
21. Shavelson RJ. Assessing Student Learning Responsibly: From history to an Audacious Proposal. *Change*. 2007;8:26-33.
22. U.S. Department of Education. A Test of Leadership: Charting the Future of U.S. Higher Education. Washington, DC. 2006.
23. Tabachnick BG and Fidell LS. Using Multivariate Statistics. Boston, Allyn MA & Bacon Publishers. 2007.
24. Meredith W. "Latent Curve Analysis" *Psychometrika*. 1993;55:107-122.