MAKING THE MARK: ARE GRADES AND DEEP LEARNING RELATED?

Corbin M. Campbell
Assistant Professor
Department of Organization and Leadership
Teachers College, Columbia University
525 W. 120th St., Box 101
New York City, NY 10027
campbell2@tc.columbia.edu
212-531-5182

Alberto F. Cabrera
Professor
Department of Counseling, Higher Education, and Special Education
University of Maryland
cabrera@umd.edu

Paper to be presented at the Annual Meeting of the Association for the Study of Higher Education, November 2012 in Las Vegas, NV.

Abstract

Assessing gains in learning has received increased attention as one dimension of institutional accountability both in the USA (Arum & Roska, 2011) and abroad (OECD, 2012). Current approaches to assessing college learning are dominated by objective tests as well as student self-reported questionnaires, such as the National Survey of Student Engagement (NSSE). This study examined how the three NSSE Deep Approaches to Learning scales contribute to the narrative on academic rigor at a large, public research institution. Using Confirmatory Factor Analyses and Structural Equation Modeling, results showed that the three Deep Approaches to Learning constructs were internally valid, but deep learning was not related to GPA. Findings raised questions regarding good measurement of student learning and student reward for rigorous performance.

Keywords: Deep Learning, GPA, Academic Rigor, NSSE

Making the Mark: Are Grades and Deep Learning Related?

Introduction

National and global movements toward assessing, tracking, and reporting institutional effectiveness via student learning outcomes are at the forefront of the Higher education debate in both the USA and abroad. For example, the Organization for Economic Co-operation and Development has begun an effort to compare student learning internationally with an initiative called Assessing Higher Education Learning Outcomes (AHELO; OECD, 2012). AHELO also produced a report that details the various ways that countries assess the learning outcomes of higher education, including seven measures from the U.S., such as the Collegiate Assessment of Academic Proficiency (CAAP), the Measure of Academic Proficiency and Progress (MAPP), the Collegiate Learning Assessment (CLA), the National Survey of Student Engagement (NSSE), among others (Nusche, 2008). Meanwhile, in the USA, the Association of Public Land Grant Universities (APLU) has been touting the College Portrait as a mechanism to bring greater accountability to public higher education by reporting gains in critical thinking and written communication as appraised by such tests as the ETS Proficiency Profile, ACL and CAAS (http:// collegeportraits.org/).

While the question of how to best assess student learning has yet to be resolved, the debate has been plentiful. Some measures are direct tests of student knowledge and thinking (e.g., CLA, CAAP) while others seek to appraise the impact of college on learning via surveys asking participants about their learning outcomes (e.g. NSSE). Additionally, some are domain and field specific (e.g. Major Field Tests), while others aim to capture broad measures of learning, such as critical thinking (e.g. CLA). Some scholars and practitioners criticize standardized testing for higher education, such as the CLA, for questions of methodology and motivating students to take the test and take it seriously (Lederman, 2011). Others criticize the use of self-reported surveys, such as NSSE, due concerns of reliability, validity, theoretical underpinnings, and intercultural meanings (Campbell & Cabrera, 2011; Dowd, et al, 2011; Porter, 2011).

Evidence of the effectiveness of U.S. higher education in terms of student learning has had mixed results, at best. The most prominent and contentious example of this research is the 2010 book by Arum and Roska, *Academically Adrift*. Arum and Roska demonstrated through their expansive review of the literature, data from student surveys, transcript analysis, and from CLA data that hours spent on academic work has declined and learning gains in college are spotty both across and within institutions. Simultaneously, grades and graduation rates have increased, rather than declined. All in all, *Academically Adrift* paints a picture of U.S. higher education that is lacking rigor. This book has received enormous attention in the higher education community, as demonstrated by Jaschik (2011) from Inside Higher Education: "If the purpose of a college education is for students to learn, academe is failing, according to *Academically Adrift*." Other studies have supported and reinforced these claims e.g., Babcok & Marks, 2010). Pascarella, Blaich, Martin, and Hanson (2011), used CLA and CAAP data from 17 institutions from the Wabash National Study and nearly replicated the results from Arum and Roska regarding learning "gains" and also amount of time spent studying.

Yet, *Academically Adrift* and the affront on academic rigor in higher education have its critics. Astin (2011) noted problems with the statistical techniques employed in the book and raised methodological issues embedded in the CLA test. Sternberg (2011) described concerns that the claims of lack of academic rigor are narrowly focused on specific types of thinking and ignore other important learning outcomes, such as creative thinking and ethical thinking. While

Pascarella and associates found similar results to Arum and Roska, they also criticized the broad reaching conclusions: namely, the Arum and Roska's study lacks a comparison group of those who did not attend college to help situate their findings.

Because of the broad reaching debate on academic rigor, measures of academic challenge warrant further scrutiny. The National Survey of Student Engagement is a survey of more than 1,400 institutions regarding student engagement with effective educational practices within their institutions. The survey has several measures that are relevant to the debate on academic rigor. Of particular interest are the Level Academic Challenge benchmark and the three Deep Approaches to Learning (DAL) constructs (integrative, higher order, and reflective). Previous research had demonstrated that certain NSSE benchmarks, including the Level of Academic Challenge benchmark, may not be stable or reliable at the institutional level (Campbell & Cabrera, 2011; Gordon, Ludlum, & Hoey, 2008; Kuh, 2009; LaNasa, Cabrera, & Transgrud, 2009). However, little research outside the NSSE-affiliated scholarship has examined the construct validity of the Deep Approaches to Learning constructs as operationalized by NSSE at an institutional level. As such, this study focuses on the NSSE DAL scales.

Certain scholars have validated the NSSE Deep Approaches to Learning constructs on a national level in terms of both internal and predictive validity (Nelson Laird, Shoup, & Kuh, 2006; Nelson Laird, Garver, Niskode-Dossett, & Banks, 2008; Mayhew, Seifert, Pascarella, Nelson Laird, & Blaich, 2012). Evidence largely supports the internal validity of the NSSE DAL scales using multi-institutional samples. For example, Nelson Laird, Shoup, & Kuh (2006) used Confirmatory Factor Analysis with NSSE multi-institutional data and found excellent internal validity of the NSSE three DAL scales (Integrative, Reflective, and Higher Order) as well as a second-order omnibus DAL construct. Studies on the predictive validity of DAL scales have largely found modest, yet significant gains with several outcomes measures. Nelson Laird and associates (2008) studied 383 students at three institutions and found that the NSSE omnibus scale for Deep Approaches to Learning was moderately positively related to critical thinking disposition, but not significantly related to critical thinking skills. They also found that the effects of deep learning on reflective thinking skills were conditional based on prior academic ability. Similarly, Reason, Cox, McIntosh, and Terenzini (2010) found that the DAL scales were associated with self-reported learning outcomes, but not with the CAAP critical thinking test.

Other scholars have considered the connection between the NSSE Deep Approaches to Learning scales and other kinds of student outcomes outside the realm of critical thinking. Overall, these studies also find a modest, yet significant relationship. For example, Mayhew and associates studied 1,457 first year students across 19 institutions with a longitudinal design to determine whether the DAL scales were related to moral reasoning outcomes. They found that that the Overall Deep Learning Scale (a composite of three scales: Integrative Learning, Reflective Learning and Higher Order Learning) had a significant (liberal *p*-value, .10), yet marginal influence on end of the year Moral Development Scores (Effect Size=.08). Of the three individual DAL scales, only the Integrative Learning scale had a small, but significant influence on moral reasoning (Effect Size=0.10).

While the topic of whether NSSE Deep Approaches to Learning scales are associated with important outcomes has received attention at the multi-institutional level, little research has explored the DAL scales in depth on an institutional level. Yet, individual institutions are using the DAL scales to guide intervention strategies and institutional policy. For example, Rhode Island University tracks progress on the DAL scales as a formative assessment tool to understand institutional effectiveness for student learning (OIR, 2011).

The purpose of this study was twofold: to investigate the construct validity of the Deep Approaches to Learning scales from the National Survey of Student Engagement (NSSE) at one research extensive institution and, if the constructs were valid measures of deep learning, to determine whether they predict GPA. All in all, we wonder how the NSSE deep learning scales can add to the narrative on academic rigor at the institutional level. The research questions were:

1) Do the NSSE Deep Approaches to Learning constructs hold as three separate, though inter-related, constructs?
2) Is there an omnibus Deep Approaches to Learning construct? If so, which of the three facets of deep learning appear to be most important for this omnibus construct?
3) Are the three Deep Approaches to Learning constructs related to GPA?

Theoretical Perspectives

The National Survey of Student Engagement created the Deep Approaches to Learning constructs to identify college environments and institutions that promote student understanding and thinking about the underlying meaning of information in comparison to surface learning, such as rote memorization (Nelson Laird, Shoup, & Kuh, 2006). NSSE scholars based their DAL constructs on previous landmark work by Marton and Saljo (1976). Marton and Saljo studied Swedish students' responses to prose, finding that there were differences in the ways that students made meaning of the information in the prose. They found that some students made meaning at a surface level and others in a deeper level that emphasized connections, integrating knowledge, and metacognition. NSSE also relied on scholars such as Biggs (2003), Ramsden (2003), and Tagg (2003) to define specific patterns of deep learning behaviors, such as integrating knowledge, meaning-making, and creating connections memorization (Nelson Laird, Shoup, & Kuh, 2006).

Previous studies have demonstrated a connection between deep approaches to learning and important student outcomes, such as earning higher grades, persistence, and better ability to process information (Entwistle & Ramsden, 1983; Nelson Laird, Garver, Niskode-Dossett, & Banks, 2008; Prosser & Millar, 1989). In a comprehensive, longitudinal study of more than a thousand students across 19 institutions, Mahew, and associates (2012) found that the three NSSE DAL scales together had a small influence on moral reasoning, but only one of the three deep learning scales, the integrative learning scale, had a small effect on its own.

Methods

This study sought to examine the relationship between the NSSE Deep Approaches to Learning scales and GPA, based on a large, public research institution. A single institution was chosen for this study in order to contextualize the findings within the university practices concerning student learning and assessment. Additionally, previous studies of NSSE at single institution sites (e.g. Campbell & Cabrera, 2011; Carle, Jaffe, Vaughan, & Eder, 2009; LaNasa, Cabrera, & Transgrud, 2009; Pike, 2006) have found that certain NSSE scales have questionable reliability at the institutional level, even if these scales hold in national samples. As such, it is particularly important to focus on a single institution to understand the connection between the NSSE DAL scales and GPA.

Data Sources

This study is based on a sample of 5,117 senior students who completed the NSSE survey in the spring of 2009. The survey was administered by NSSE via the web, yielding a response rate of 28%. Only non-transfer seniors were utilized, yielding an analytical sample of 1,026. Non-transfer seniors were chosen because transfers may have had a different experience with learning at a previous institution, and this study is situated in one institution. Institutional variables such as college cumulative GPA, high school GPA, and SAT math scores were added to the dataset.

Data Screening

Following Muller and Hancock's (2008) best practices for conducting structural equation modeling, we examined the extent to which our data were normal prior to testing our CFA and SEM models. In examining departures of normality, we relied on SPSS 18.0, Stata 12.1[1] and EQS 6.1[2]. Each of the 12 higher learning items was significantly skewed and non-normal. Moreover multivariate tests of skewness (Mardia Skewness = 91.2, p <.001) and kurtosis (Mardia mKurtosis = 389.5, p <.001) indicated that the 12 Deep Approaches to Learning items as a group significantly departed from normality. During the data screening process, we noticed that 8 items displayed low variability across their anchors components, leading us to recode them[3]. Having ruled out Maximum Likelihood procedures for normality-distributed data, we decided to use EQS robust methods and the MPlus robust weighted least squares (WLSMV) estimator which is ideally suited to handle categorical, ordinal and continuous variables (Brown, 2006; Finney & DiStefano, 2006) . All models were tested first using MPlus 6.1 software and then confirmed using EQS 6.1 software.

Model Testing

We tested a confirmatory factor model assuming that three dimensions account for the intercorrelations underlying the 12 items that comprise a higher order Deep Approaches to Learning construct. Following recommendations from the Structural Equation Modeling literature (e.g., Bollen, 1989; Brown, 20006; Byrne, 2006; Kline, 2011), we adopted a two-step

---

[1] We used Stata's mvtest normality command to estimate both univariate, bivariate and multivariate tests of departure of normality.

[2] We relied on the EQS's mardia test

[3] Those items are: analyze, synthetize, applying, occide, integr, intede, othrvi and chngui.

strategy in answering our three research questions. First, we used Confirmatory Factor Analysis (CFA) to answer the first two research questions, regarding whether there are three stable and distinct Deep Approaches to Learning constructs (higher order, integrative, and reflective) and whether the three constructs tap a higher order DAL construct which Mayhew and associates termed ‘Deep Learning.’ We set the variance of each latent construct associated to each benchmark to one, allowing us to ascertain the extent to which the items indeed loaded in their corresponding conceptual latent factor (Brown, 2006; Kline, 2011). We also investigated the thresholds of each item to determine whether the item was better for discriminating the upper or the lower categories of responses. Brown (2006) notes that item thresholds estimated by Mplus for CFA with categorical data are analogous to the Item Response Theory’s (IRT) item difficulty parameters.

Next, we employed Structural Equation Modeling to answer our third research question regarding whether the three Deep Approaches to Learning constructs predict college cumulative GPA. The model testing the relationship between Deep Approaches to Learning and GPA included a measure of precollege academic ability that was a composite factor of high school GPA and SAT math score. High school GPA and SAT math score have been consistently cited in the literature as predictors of college academic success and persistence. (see, for example, Allen, Robbins, Casillas, & Oh, 2008; Pascarella & Terenzini, 2005; Tinto, 1993).

We relied on three robust measures of fit to judge the CFA model and the SEM models. These indices include: (a) the Comparative Fit Index (CFI), (b) the Tucker-Lewis Index (TLI), and (d) the Root Mean Square Error of Approximation (RMSEA). We guided our selection of goodness-of-fit values based on recommendations from the SEM literature (Byrne, 2006; Hu & Bentler, 1999). Accordingly, we sought CFI and TLI values of 0.95 or higher to signify an excellent fit, but we also considered values greater than .90 to be appropriate (Möller, Retelsdorf, Köller & Marsh, 2011). In terms of RMSEA, we judged values ranging from 0 to .05 excellent, but we also considered RMSEA values less than .08 to be suitable. In addition, we estimated 90% (CI_{90}) confidence intervals to check that RMSEA values did not fall beyond the cut off value of .10, signifying the rejection of the model (Byrne, 2006).

Limitations

There are several possible noteworthy, but not insurmountable, limitations to this study. First, this study utilizes data from one single institution, rather than using a multi-institutional sample. This was an intentional decision because NSSE data needs to be validated on an institutional level and contextualized by institutional policies and practices. Caution should be used when generalizing to other institutions, particularly those that are less similar to the focus institution (i.e. a large, public, urban, research institution). Further studies should explore whether the findings of this study could be replicated in other kinds of institutions, such as liberal arts institutions. Secondly, the response rate for this survey (28%) is not ideal, but also not uncommon for a response rate for large, research universities that participate in NSSE. It is possible that there is a respondent bias that might be related to learning patterns (i.e. are those students who are more motivated to take surveys also more motivated in the classroom). Yet, even if this were true, findings would still illuminate whether or not a connection exists between Deep Approaches to Learning and GPA for students who are more engaged in learning.

Results

First, we present results from the CFA of the three, intercorrelated Deep Approaches to Learning constructs. The model demonstrated excellent fit (RMSEA = 0.042; CI_{90} = [.034, .051]; CFI = 0.978; TLI = 0.972). In addition, the loadings of individual items ranged from .494

to .854, with all but one item having loadings <.5 (Table 1). Additional evidence of the solid psychometric properties of the three DAL constructs can be seen with the reliability measure, the Coefficeint-*H*. All three reliabilities were above .750, which is excellent. Further, the intercorrelations among constructs demonstrate that the three DAL constructs are related moderatly to strongly, but are distinct (Table 2). We also found that all but two items were able to discriminate at both the upper and lower category responses (Table 3)[4]. However, we also noted that 7 out of the 12 Deep Approaches to Learning items are better at discriminating lower category responses than they are for upper category responses[5].

[INSERT TABLES 1, 2, & 3 ABOUT HERE]

Next, we proceeded to test the hypothesis that the three consructs tap a second order Deep Approaches to Learning construct. Once again, results demonstrated excellent fit (RMSEA = 0.042; CI_{90} = [.034, .051]; CFI = 0.978; TLI = 0.972) All three first order constructs had strong loadings from the second order DAL construct (Figure 1). Of particular note is the integrative learning construct for whom 98% of its variance can by explained by the latent second order DAL factor instead of measurement error.

[INSERT FIGURE 1 ABOUT HERE]

Next we present the results from the model that uses the three Deep Approaches to Learning constructs to predict GPA, while controlling for High School GPA and SAT Math score. Once again, excellent fit in the model (CFI =.99; RMSEA = .027 [CI_{90} = .017, .037]). However, interestingly, the only significant contributor to GPA was pre-college academic ability. None of the three DAL constructs contributed significantly to GPA (Figure 2).

[4] Those items are 'analyze' and 'integrate' corresponding to Higher Order Learning and Integrative Learning respectively.

[5] Four correspond to Higher Order Learning (analyze, synthesis, evaluate and apply), two belong to Integrative Learning (divclass and integr) and one is an indicator of Reflective Learning (Ownview).

[INSERT FIGURE 2 ABOUT HERE]

Discussion

This study makes three important contributions to the literature. First, it confirms that the national studies of the NSSE Deep Approaches to Learning constructs do, in fact, hold for a single research extensive institution. This may indicate that the three DAL constructs could be useful for formative assessment for institutions to track progress across time and possibly across institutions. Secondly, we found support to Mayhew and associates' (2012) implicit hypothesis that Higher Order Learning, Integrative Learning and Reflective Learning are themselves manifestations of a higher order learning factor; namely, Deep Learning. Of the three DAL constructs the Integrative Learning construct appears to be the most potent considering it is a nearly perfect indicator of the second order Deep Learning construct. Institutions might consider the items in this construct to be particularly important for interventions that target students' deep learning: discussing ideas from coursework with others and with faculty, integrating different ideas into a paper, integrating ideas across courses, and including diverse perspectives in assignments. We call for continued research in the development of DAL measures considering the finding that 58% of the NSSE DAL items are better discriminators for those respondents who were more likely to report low deep learning rather than high deep learning.

Finally, the finding that the three Deep Approaches to Learning constructs do not contribute significantly to college GPA deserves further scrutiny. We assert two possible explanations to this finding. In essence, we have a chicken or the egg question—is the lack of relationship due to a problem with the NSSE measurement of deep learning or an indication that there is a problem with GPA? We explore these two possibilities in the context of the literature review, below.

The first explanation would assume that deep learning and GPA should be related. In this case, the finding of no relationship between deep learning and GPA in this study could indicate that there may be predictive validity concerns with the NSSE scale for this particular institution. Under this explanation, it may be that NSSE's DAL scales, while internally valid, do not measure deep learning as defined by Bigg's 1987 framework. This is contextualized by prior research on multi-institutional studies that investigated relationships between other kinds of deep learning measures (not NSSE) and GPA (Hall, Bolen, & Gupton, 1995; Zhang, 2000). The results of these studies appear mixed. Zhang conducted an international comparative study that used the Student Process Questionnaire "Deep Approach" Scale, which was based on Bigg's 1987 theory of student learning approaches. Zhang found that the Deep Approach was positively and significantly related to student achievement, measured by cumulative GPA, as in the present study. By contrast, Hall Bolen, and Gupton found that the achievement approach was more influential on GPA than the other two approaches (including deep learning) in the Student Process Questionnaire. If other forms of deep learning scales, such as the Deep Approach in the Student Process Questionnaire, are associated with GPA (Zhang, 2000), this unearths a question regarding why the NSSE DAL scales do not demonstrate this association. Is it possible that the NSSE scales are tapping something other than deep learning?

By contrast, it is equally possible, that the deep learning constructs are, in fact, measuring deep learning—but at this public, research institution, deep learning is not a necessary condition for a high GPA. The NSSE DAL scales have strong construct validity at this institution and have also been validated in several national studies by several learning and engagement experts for content and predictive validity (Nelson Laird, Shoup, & Kuh, 2006; Nelson Laird, Garver, Niskode-Dossett, & Banks, 2008; Mahew, Seifert, Pascarella, Nelson, and Blaich, 2012; Reason, Cox, McIntosh, and Terenzini, 2010). If it is true that these three NSSE DAL constructs are a strong measure of deep learning and they do not predict GPA, this appears to support the contention that at this one research extensive institution certain aspects of academic rigor, such as integrative learning, synthesizing and analyzing information, and reflective learning, do not contribute adequately to student achievement.

If certain kinds of skills and modes of thinking and learning associated with "deep learning" are important to the modern workforce, it would be important for the reward for student work (i.e. grades) to reflect these forms of learning. The finding of no-relationship between deep learning and GPA raises potential questions for future research: If GPA does not reflect deep learning, what does it reflect? Would students be motivated to think more deeply if GPA were based more on higher order, integrative, and reflective learning? Did grade creep play a role in the findings of this study? Future research should continue to investigate and tease apart whether and how student achievements, such as GPA and degree attainment, are related to deep learning.

Conclusion

We find that the three constructs of the NSSE Deep Approaches to Learning scales as well as an omnibus construct of Deep Learning to be internally valid for one public, research institution. Integrative learning appears to be a particularly important aspect of Deep Learning. Yet, something is amiss in the absence of a significant relationship between the DAL scales and GPA. There are two possibilities that explain this phenomenon, and unfortunately, both are problematic in different ways. Either the NSSE DAL constructs are internally valid, but do not hold predictive validity at this research institution, or the constructs are tapping deep learning, but students who assume a deep learning approach do not necessarily earn better grades. Further research should explore these conclusions, as both have important implications for institutions and the discourse on academic rigor in higher education. For the focus institution, we would recommend continued exploration to determine whether deep learning and other forms of academic rigor are rewarded by the institution in terms of student achievement. For researchers, this manuscript might suggest that continued work should be done to understand multiple approaches to measuring student learning and academic achievement across and within institutions.

References

Allen, J., Robbins, S. B., Casillas, A., & Oh, In-Sue. (2008). Third-year college retention and transfer: Effects of academic performance, motivation, and social connectedness. *Research in Higher Education, 49*, 647–664.

Arum, R., & Roksa, J. (2011). *Academically adrift: Limited learning on college campuses*. University of Chicago Press.

Astin, A. (2011). In 'Academically Adrift,' Data Don't Back Up Sweeping Claim. *Chronicle of Higher Education*. Retrieved at: http://chronicle.com/article/Academically-Adrift-a/12637

Babcok, P. & Marks, M. (2010). *Leisure college USA: The decline in study time*. Report # 7. Washington, DC.: American Enterprise Institute for Public Research.

Biggs, J. B. (2003). *Teaching for quality learning at university*. Buckingham: Open University Press.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley.

Brown, T. A. (2006). *Confirmatory Factor Analysis for Applied Research*. New York, NY: Guilford Press.

Byrne, B. M. (2006). *Structural equation modeling with EQS* (2nd ed.). Newbury Park, CA: Sage.

Campbell, C. M., & Cabrera, A. F. (2011) How Sound Is NSSE?: Investigating the Psychometric Properties of NSSE at a Public, Research-Extensive Institution. *The Review of Higher Education, 35*(1), 77-103.

Carle, A. C., Jaffe, D., Vaughn, N. W., & Eder, D. (2009). Psychometric properties of three new national survey of student engagement based engagement scales: An item response theory analysis. *Research in Higher Education, 50*, 775– 794.

Dowd, A. C., Sawatzsky, M., & Korn, R.(2011). Theoretical foundations and a research agenda to validate measures of intercultural effort. *Review of Higher Education, 35*, 17-44.

Entwistle, N. J. & Ramsden, P. (1983). *Understanding student learning*. London: Croom Helm.

Finney, S. J. & DiStefano, C. (2006). Non-normal and categorical data in structural equation modeling. In G.R. Hancock & R. O. Muller (Eds.). *Structural Equation Modeling: A second course* (pp. 269-314). Greenwich, Connecticut: IAP.

Hall, C. W., Bolen, L. M., & Gupton, R. H. (1995). Predictive validity of the Study Process Questionnaire for undergraduate students. *College Student Journal, 29*, 234-239.

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55.

Kline, R. B. (2011). *Principles and practices of structural equation modeling*. New York, NY: Guilford Press.

Jaschik, S. (2011, January 18). 'Academically Adrift.' *Inside Higher Education.* Retrieved from http://www.insidehighered.com/news/2011/01/18/study_finds_large_numbers_of_college_students_don_t_learn_much

Lederman, D. (2011, November 14). CLA as Catalyst for Change. *Inside Higher Education.* Retrieved from http://www.insidehighered.com/news/2011/11/14/cla-experiment-focused-private-colleges-attention-assessment

LaNasa, S. & Cabrera, A. F., Tangsrud, H. (2009). The construct validity of student engagement: A confirmatory factor analysis. *Research in Higher Education*, 50(4), 313-352.

Marton, F. & Säljö, R. (1976). On qualitative differences in learning I: Outcome and process. *British Journal of Educational Psychology, 46*, 4-11.

Mayhew, M., Seifert, T. A., Pascarella, E. T., Nelson Laird, T.F., & Blaich, C F. (2012). Going deep into mechanisms for moral reasoning growth: How deep learning approaches affect moral reasoning development for First-year students. *Research in Higher Education*, 53, 26-46.

Möller, J., Retelsdorf, J., Köller, O. & Marsh, H. W. (2011). The reciprocal internal/external frame of reference model: An integration of models of relations between academic achievement and self-concept. *American Educational Research Journal, 48*(6), 1315-1346

Mueller, R. O., & Hancock, G. R. (2008). Best practices in structural equation modeling. In J. W. Osborne (Ed.), *Best practices in quantitative methods* (pp. 488-508). Thousand Oaks, CA: Sage Publications, Inc.

Nelson Laird, T., Shoup, R., & Kuh, G. (2006, May). *Measuring deep approaches to learning using the National Survey of Student Engagement*. Paper presented at the Annual Forum of the Association for Institutional Research, Chicago, IL.

Nelson Laird, T.F, Garver, A. K., Niskode-Dossett, A. S., & Banks, J. V. (2008, Nov). *The Predictive Validity of a Measure of Deep Approaches to Learning*. Paper presented at the Annual Meeting of the Association for the Study of Higher Education, Jacksonville, FL.

Nusche, D. (2008). *Assessment of Learning Outcomes in Higher Education: A comparative review of selected practices*. Organization for Economic Co-operation and Development.

Office of Institutional Research (OIR; 2011). *NSSE Survey Deep Learning Items: Comparison of first-year and senior student responses*. The University of Rhode Island.

Organization for Economic Co-operation and Development (OECD; 2012). *Testing student and university performance globally: OECD's AHELO*. Retrieved from: http://www.oecd.org/document/22/0,3746,en_2649_39263238_40624662_1_1_1_1,00.html

Pascarella, E.T., Blaich, C., Martin, G. L., & Hanson, J. M. (2011). How robust are the findings of Academically Adrift? Evidence from the Wabash National Study. *Change.*

Pascarella, E. T., & Terenzini, P. T. (2005). *How college affects students: A third decade* *of research.* San Francisco: Jossey-Bass.

Pike, G. R. (2006). The convergent and discriminant validity of NSSE scalelet scores. *Journal of College Student Development, 47*, 550–563.

Porter, S. R. (2011). Do college student surveys have any validity. *Review of Higher Education, 35*, 45-76.

Prosser, M. & Millar, R. (1989). The "how" and "why" of learning physics. *European Journal of Psychology of Education, 4*, 513-528.

Ramsden, P. (2003). *Learning to teach in higher education*. London: RoutledgeFalmer.

Reason, R. D., Cox, B. E., McIntosh, K., & Terenzini, P. T.(2010, May). *Deep learning as an individual, conditional, and contextual influence on first-year student outcomes*. Paper presented at the Annual Forum of the Association for Institutional Research, Chicago, IL.

Sternberg, R. J. (2011, February 8). Who Is Really Adrift? *Inside Higher Education*. Retrieved from http://www.insidehighered.com/views/2011/02/08/a_critique_of_academically_adrift_and_the_test_behind_many_of_the_findings

Tagg, J. (2003). *The learning paradigm college.* Boston, MA: Anker.

Tinto, V. (1993). *Leaving college: Rethinking the causes and cures of student attrition* (2nd ed.). Chicago: University of Chicago Press.

Zhang, L. (2000). University students' learning approaches in three cultures: An investigation of Biggs's 3P model. *The Journal of Psychology, 134*(1), 37-55.

Table 1. Loadings and variance accounted for in each indicator of Deep Learning

Deep Learning	*Measure*	*Loading*	*Variance*		*Reliability of the Deep Learning (Coefficient-H)*
			Explained	*Error*	
1. Higher Order	R_ANALYZE	0.819	0.672	0.328	0.872
	R_SYNTHESZ	0.854	0.730	0.270	
	EVALUATE	0.713	0.509	0.491	
	R_APPLYI	0.710	0.504	0.496	
2. Integrative Learning	R_OCCIDE	0.721	0.520	0.480	0.752
	FACIDEAS	0.518	0.268	0.732	
	DIVCLASS	0.494	0.244	0.756	
	R_INTEG	0.581	0.459	0.541	
	R_INTIDE	0.656	0.338	0.662	
3. Reflective Learning	OWNVIEW	0.677	0.430	0.570	0.833
	R_OTHRVI	0.792	0.628	0.372	
	R_CHNGVI	0.844	0.712	0.288	

Table 2. Structural correlations among Deeep Learning Constructs

Deeep Learning Construct	*1*	*2*	*3*
1. Higher order Learning	1		
2. Integrative Learning	.517	1	
3. Reflective Learning	.336	.621	1

Goodness of Fit Indicators

<u>MPlus</u>

RMSEA = 0.042; CI_{90} = [.034, .051]; CFI = 0.978; TLI = 0.972

Table 3. Threshold Analysis for the 12 items comprising Deep Learning by Factor

Latent Factor:	Item	Threshold	Estimate	S.E.	*t*-value	*p*-value
Higher Order Learning	Analyze	1	-1.175	0.052	-22.403	0.000
		2	0.041	0.040	1.001	0.317
	Synthesis	1	-0.822	0.046	-17.863	0.000
		2	0.284	0.041	6.891	0.000
	Evaluate	1	-1.586	0.066	-24.129	0.000
		2	-0.608	0.043	-14.018	0.000
		3	0.396	0.042	9.510	0.000
	Apply	1	-0.912	0.047	-19.298	0.000
		2	0.085	0.041	2.101	0.036
Integrative Learning	Occide	1	-0.257	0.041	-6.300	0.000
		2	0.753	0.045	16.823	0.000
	FACIDEAS	1	-0.401	0.041	-9.737	0.000
		2	0.893	0.046	18.432	0.000
		3	1.608	0.066	24.432	0.000
	DIVCLASS	1	-1.406	0.057	-24.535	0.000
		2	-0.246	0.040	-6.178	0.000
		3	0.661	0.043	15.514	0.000
	Intide	1	-0.152	0.042	-12.193	0.000

Table 3. Threshold Analysis for the 12 items comprising Deep Learning by Factor						
		2	0.654	0.043	15.113	0.000
	Integr	1	-1.150	0.050	-22.896	0.000
		2	0.045	0.039	1.157	0.247
Reflective Learning	Ownview	1	-1.488	0.063	-23.784	0.000
		2	-0.099	0.041	-2.418	0.016
		3	0.921	0.048	19.213	0.000
	Chngvi	1	-0.393	0.042	-9.376	0.000
		2	0.737	0.045	16.348	0.000
	Othrvi	1	-0.333	0.042	-7.992	0.000
		2	0.783	0.046	17.122	0.000

Figure 1. Three factors as manifestations of Deep Learning (second order factor)

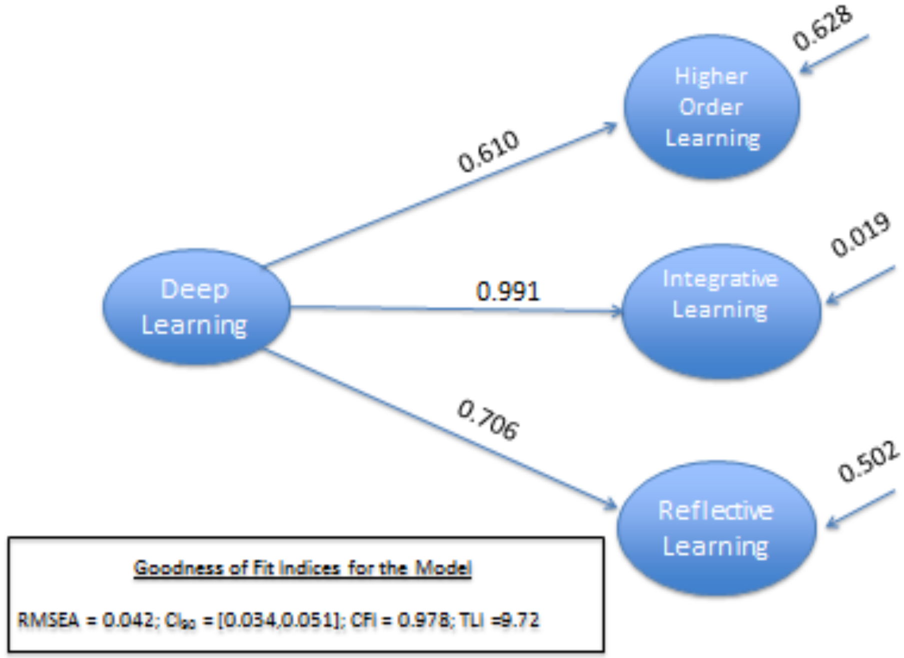

Figure 2. Deep Learning Factors Predicting GPA