4-13-2016

The disaggregation of value-added test scores to assess learning outcomes in economics courses

William B. Walstad
University of Nebraska-Lincoln, wwalstad1@unl.edu

Jamie Wagner
University of Nebraska at Omaha, jfwagner@unomaha.edu

THE DISAGGREGATION OF VALUE-ADDED TEST SCORES TO ASSESS LEARNING OUTCOMES IN ECONOMICS COURSES

Short title for running header: Disaggregating Value-added Test Scores

William B. Walstad and Jamie Wagner*

Abstract: This study disaggregates posttest, pretest, and value-added or difference scores in economics into four types of economic learning: positive; retained; negative; and zero. The types are derived from patterns of student responses to individual items on a multiple choice test. The micro and macro data from the *Test of Understanding in College Economic* (TUCE) are used to show how aggregate scores can be re-interpreted based on their learning components. The regression analysis shows the relative contribution from learning components to aggregate scores. A value-added or difference score has a potential problem because it is a mixture of positive and negative learning. A better alternative would be to use the positive learning scores to assess improvement in economic understanding.

Keywords: difference score, positive learning, *Test of Understanding in College Economics* (TUCE), value-added test

JEL codes: A22, I21

---

*William B. Walstad (e-mail: wwalstad1@unl.edu) is a professor of economics at the University of Nebraska-Lincoln. Jamie Wagner is an assistant professor of economics and Director, Center for Economic Education at the University of Nebraska at Omaha. Walstad is the corresponding author.

Partial support for this research came from the Calvin K. Kazanjian Economics Foundation.

Research on the teaching and learning of economics at the undergraduate level often focuses on what students attain from an economics course as measured by an achievement test or another outcome measure such as an exam grade (Siegfried and Fels 1979; Becker 1997; Allgood, Walstad, and Siegfried 2015). In essence, when these attainment measures are used as the outcome variable in regression studies in economic education, they provide an assessment of the *stock* of student economic knowledge or understanding at a point in time, such as at the end of the course in the case of a final exam score. What also should be of interest to economics instructors is the change in student economic knowledge or understanding during a course, which means that economic learning should be thought of as a *flow* measure.

In a value-added study of factors that affect economic learning students would be given an economics pretest at the beginning of the course or unit of instruction and then given the same economics posttest at the end of the course or unit of instruction. The difference between the pretest and posttest scores would provide an estimate of the value-added from the course in terms of what students learn about economics. In economic education studies, this value-added or difference score (and sometimes called a change or gain score) is used as a dependent variable in a regression with student demographic characteristics and other factors serving as control variables (e.g., Maxwell, Mergendoller, and Bellisimo 2005; Dickie 2006).

Our study shows that when multiple choice data are available for pretest and posttest scores in economics, these value-added scores can be disaggregated based on the pattern of pretest and posttest responses to individual test items. The response patterns can then be used to create four other scores that measure different learning outcomes: (1) gaining economic understanding or *positive learning*; (2) maintaining economic understanding or *retained learning*; (3) losing economic understanding or *negative learning*; and, (4) continuing economic *mis*-understanding or *zero learning*. The second section of the study provides a detailed explanation and justification for this disaggregation of the multiple choice pretest and posttest scores into the four types of economic learning.

The third section describes the source for the pretest and posttest scores in decomposition analysis. The test data comes from the national norming of the *Test of Understanding of College Economics* (TUCE) (Walstad, Watts, and Rebeck 2007). This standardized test has been used for measuring student achievement or learning in principles of economics courses through four editions. The TUCE contains two different multiple choice exams for principles of economics courses, one for a microeconomics (micro TUCE) and one for macroeconomics (macro TUCE). The national norming data set provides two large samples of students with matched pretest and posttest scores for the test analysis, one for micro (n = 3,255) and one for macro (n = 2,789). Also included in the data set are demographic and other background variables related to each student.

The fourth section presents the regression model and the set of control variables for student characteristics that are included across all equations. It discusses the expectations from the regression analysis based on how the four different learning components are likely to affect the coefficients for the aggregate scores (posttest, pretest, and difference). The primary focus in this discussion is on positive learning, which is expected to be a better measure of value-added from economics courses than when a difference score is used as a dependent variable.

In the fifth section, the results from the regression analysis of the aggregate test scores and the four learning components are presented using the micro TUCE and macro TUCE data sets. The learning components are then used to demonstrate empirically the relative contribution of different types of learning that comprise the posttest, pretest, and difference scores. One of the main findings from the regression analysis is that the difference score may be subject to misinterpretation because it is a construct that contains measures of both positive and negative learning. A preferable alternative would be to use the positive learning score as the dependent variable. The regression results also show that factors influencing positive learning sometimes have different effects than the same factors have on a difference score.

DISAGGREGATING POSTTEST, PRETEST, AND DIFFERENCE SCORES TO ASSESS LEARNING

Assume that economics students are given a multiple choice test as a pretest at the beginning of an economics course and the same test at the completion of the course. In this case, a "difference" score for either a student or for the class is a constructed measure that is obtained by subtracting the pretest score from the posttest score. It has three possible outcomes: (1) posttest is greater than pretest; (2) posttest is less than pretest; and (3) posttest equals pretest. A difference score also has been called a "change" score, which is a substitute term because it too covers all three possibilities. A "gain" score is another term used to describe a change from pretest to posttest, but as the name implies there is less ambivalence about the direction of the change because with most instruction there is likely to be an improvement in scores from pretest to posttest. These scores are also called a "value-added" score to reflect the change from some type of educational intervention. This study uses the terms "difference" or "value-added" score interchangeably because difference scores are studied in educational measurement (Petscher and Schatschneider 2011) while value-added modeling is a focus of research in economics (Koedel, Mihaly, and Rockoff 2015).

What is more important, however, than an explanation of terms is an understanding of the learning components that underlie the pretest and the posttest that are used to construct the value-added or difference scores. The initial idea for such an analysis in economic education came from Saunders and Power (1995), who first used pre and post responses to test items to study how content coverage in microeconomics affected student learning. Our current study adopted similar methods to classify economics learning into four types, but for a completely different purpose, which was to disaggregate and re-interpret posttest, pretest, and difference scores. Given the available item data from a multiple choice pretest and posttest, the first response pattern is for students to give an incorrect response on the pretest and a correct response on the posttest. The value-added or difference score for that test item reflects *positive learning* (PL) because it provides evidence that there was an increase in understanding as measured by that test item. The second type is *retained learning* (RL), where students give a correct response to the item on the pretest and again on the posttest, demonstrating that the students have retained their economic understanding over time. The first two patterns are generally considered good learning outcomes from an economics course.

The other two response patterns provide information about what is *not* learned. Students can give a correct response to an item on the pretest and an incorrect response on the posttest. This change would suggest that there is *dis*-learning or *negative learning* (NL) because the students appeared to know the correct answer previously, but were not able to retain or demonstrate that learning at a later time. The other combination is for students to supply a wrong answer on the pretest and the posttest. This outcome is labeled as *zero learning* (ZL) because there is no evidence of economic understanding about that item. Of course further analysis can be conducted with ZL to study the consistency and inconsistency in patterns of incorrect responses, but it is a minor distinction so the focus of this study will be just on ZL in general.

For each test item, the pretest and posttest response pattern can be scored a 1 if it falls into one of the four learning categories (PL, RL, NL, or ZL) and 0 otherwise. The four item scores can then be totaled across all test items to create a unique PL, RL, NL, and ZL test score for each student. For example, on a 30-item test such as the TUCE, a student might have answered 10 items scored as PL, 6 items showing RL, 3 items indicating NL, and 11 items classified as ZL. The individual student scores for each type of learning also can be summed to create mean scores for a class or group of students that reflect economic understanding within each learning type. For example, if the scores listed above were representative of class means by learning type, the results would be interpreted as showing that students in a class, on average, demonstrate positive learning on 33 percent of items, retained learning on 20 percent of items, negative learning on 10 percent of items, and *mis*-understanding on 37 percent of items.

PL, RL, NL, and ZL scores also can be used to construct the posttest, pretest, and difference scores. The posttest score for a student is the total of PL and RL scores (PL + RL). The pretest score for a student is the sum of RL and NL scores (RL + NL). A difference score in equation form, therefore, would be: (PL + RL) − (NL + RL) that reduces to (PL − NL) because RL can be eliminated from both sides. Using the previous example of the set of individual scores for PL (10), RL (6), NL (3) and ZL (11), the posttest score (PL + RL) is 16, the pretest score (RL + NL) is 9, and the difference score (PL − NL) is 7. The learning components also can be used to describe the number wrong on the posttest (ZL + NL), which would be 15, and the number wrong on the pretest (ZL + PL), which would be 20.

What too is shown by the disaggregation of the posttest, pretest, and difference scores is that each one depends on the size of the other learning components. The posttest is the total of RL + PL, so if RL is smaller, then PL is larger. The size of RL also affects the pretest score, but in a different way because the pretest is the total of RL + NL. In this case, a pretest with less RL increases NL. A difference score (PL − NL) is directly two learning components. What the combinations indicate is that there is a trade-off in the design of tests for measuring value-added because RL moderates PL in the posttest and NL in the pretest. If the testing goal is to increase the amount of PL measured, then RL needs to be reduced, but this change will increase the amount of NL, which in turn will affect the size of the difference score. It also is not easy to interpret a difference score because it is not known whether the change occurred primarily because PL was high or NL was low. Compounding the measurement of the posttest, pretest, and is ZL. If ZL is large it limits the size of RL and PL on the posttest, RL and NL on the pretest, and PL and NL on a difference score.

When pretest and posttest multiple choice data are available for students, a better alternative to using the aggregate test scores for assessing student learning in economics would be to disaggregate the test data and analyze each of the four learning components, or the one of most interest. Using the posttest, pretest, and difference score may produce misleading results because each aggregate score is influenced by the four learning components and their unknown interactions. If the purpose of the value-added analysis is to find out what students learn about economics from economics instruction, then a focus on PL would be most important because it will show the degree of change from what students do not understand to what they do understand. There also may be some interest in knowing if students maintain that economic understanding during the course, in which case RL would be the target. NL would be studied to find out what likely influences the economic understanding students appeared to know at the beginning of the course, but did not at the end of the course. ZL could be investigated to discover what accounts for students showing no change in economic understanding during a course. With the completion of this theoretical explanation of aggregate test scores and learning components, the exposition now turns to a description of the test data to be used for the empirical analysis.

TUCE DATA

For the national norming of the fourth edition of the TUCE, 5,480 students took the micro TUCE test and 5,517 took the macro TUCE test (Walstad, Watts, and Rebeck 2007). Some students took the TUCE only as a pretest (1,621 micro; 2,022 macro) because instructors ran out of class time to give the posttest or because students dropped a course. In other cases, students took only the posttest (604 micro; 706 macro) because some instructors did not administer the pretest, but decided to administer the posttest. For test norming purposes, the core group of students was the "matched" sample of students who took the TUCE test as a pretest at the beginning and as a posttest at the end of a course (3,255 or 59.4 percent of micro students; 2,789 or 50.1 percent of macro students). The advantage of this matched sample for the analysis of the TUCE data is that it controlled for differences in student characteristics because the same students took the pretest and posttest. The matched sample for TUCE microeconomics and the matched sample for TUCE macroeconomics will be used for this analysis of student learning of principles of economics.

Table 1 provides a description of all the variables to be used for the regression analysis. Among the test variables were the posttest score, pretest score, difference score and the learning variables (PL, RL, NL, and ZL). The TUCE data set also includes survey data collected from students to obtain demographic and other information for the test analysis. Eleven of these background variables were re-coded as dichotomous (1, 0) variables for the regression analysis to control for the demographic characteristics of students in the micro and macro data sets: (1) gender (male=1); (2) race or ethnicity (1=white); (3) year in college or university (1=freshman); (4) academic major (1=business); (5) enrollment status (1=full-time); (6) communicate better in English than another language (1=English speaker); (7) grade point average (1=3.0 to 4.0 GPA);

(8) took a course in college calculus (1=yes); (9) took a college course in economics (1=yes); and, (10) took a high school course in economics (yes=1).

[Insert table 1 about here]

The 11th variable captured the type of institution of higher education students attended as defined by Carnegie Foundation classifications (http://carnegieclassifications.iu.edu/). For the micro TUCE, the sample of 41 institutions included 7 associate's colleges (two-year or community colleges) offering degrees (8 percent of students), 4 colleges offering only baccalaureate degrees (8 percent), 25 universities offering up to a master's degree (61 percent), and 5 doctoral-granting or research universities (23 percent). The macro sample of 44 institutions included 4 associate's colleges offering two-year degrees (7 percent of students), 7 colleges offering only a baccalaureate degree (14 percent), 27 universities offering up to a master's degree (53 percent), and 6 doctoral-granting or research universities (27 percent). To simplify the institutional analysis, this variable was coded as a 1 if the institution was an associate's college (i.e., two-year or community college).

Table 2 presents the descriptive statistics for all the variables used from the micro sample. The results from two types of samples are reported. The first sample shows the means and standard deviations for the unrestricted or full TUCE sample, which for most variables has 3,255 observations. The second sample is a restricted sample with 3,052 observations that contains the means and standard deviations for the same set of variables used in the regression analysis.

[Insert table 2 about here]

The mean scores for the micro TUCE reported in table 2 show that it is a challenging test. On the pretest, students could answer correctly 31 percent of the 30 test items and by the posttest it was 43 percent. The average increase in scores was 3.39 points. The post–pre change, however, may understate what is learned as measured by the PL score (7.74) because the difference score of 3.30 is calculated as PL − NL (7.74 − 4.35). What also is evident is that the mean posttest score of 12.77 can be split into RL (5.03) and PL (7.74) components, indicating that 39 percent of the posttest score comprises retained learning and 61 percent is positive learning. Similarly, the mean pretest score of 9.39 can be divided into RL (5.03) and NL (4.35) components, showing that 54 percent of the pretest score is what students already know and will retain by the posttest and 46 percent of the pretest score is what students know at the beginning of the course but do not retain by the end of the course.

Table 3 reports the descriptive statistics for all the variables for the macro TUCE sample (n = 2,789 unrestricted; n = 2,687 restricted for regression analysis). The macro TUCE also was difficult for most students because on the pretest, students could correctly answer about 33 percent of the 30 test items and on the posttest it was 47 percent. As with the micro test, this average value-added score (4.40) may underestimate the gain in economic understanding shown by positive learning (8.52) because in the value-added score is calculated as PL − NL (8.52 − 4.13). The posttest score of 14.20 can be split into RL (5.67) and PL (8.52) components which means that 40 percent of the score is RL and 60 percent is PL. As for the pretest score of 9.80, about 58 percent of that score represents RL (5.67) and 42 percent is NL (4.13).

[Insert table 3 about here]

The TUCE data also can be analyzed item by item to assess the degree of positive learning and retained learning in each item (see Appendix 1). For the micro TUCE items, the average amount of positive learning for all 30 items is 60.6 percent, which means that the average for retained learning is 39.4 percent. About the same division (60 versus 40 percent) is found for the macro TUCE items. Within the set of micro and macro items, however, there is substantial variation in the amount of positive learning reflected in items. For micro TUCE, the range runs from a high of 85.3 percent (#4) to a low of 40 percent (#22). Similarly with the macro TUCE, the range goes from a high of 86.7 percent (#5) to a low of 40.5 percent (#13). This tabulation indicates which items show the greatest amount of positive learning for students. Certain items are clearly contributing more to positive learning than other items, which may be of interest for developing tests that better measure positive learning.

REGRESSION MODEL AND SOME EXPECTATIONS

The value of analyzing the learning components derived from the disaggregation of pretest and posttest scores is best shown through regression analysis. The following basic regression model was specified for the analysis:

$$Test\ score\ (Y) = \beta_0 + \beta_i X + \beta_k Z$$

The test score, or dependent variable Y, could be any one of the relevant outcome variables: posttest, pretest, difference score, or one of the four learning components: PL, RL, NL, and ZL.

The variable X is a vector of student characteristics, which are all dummy variables (1, 0) as described in table 1. These student characteristics are all thought to have a potential influence on a test outcome and have been included in regression studies in economic education with undergraduates (Siegfried 1979; Becker 1997; Allgood, Walstad, and Siegfried 2015). The inclusion of this set of variables, however, is for comparison purposes to investigate how the coefficient values change as the dependent variable (test score) changes. The variable Z is a vector of school dummy variables to control for differences across colleges and universities. The regressions are estimated with school fixed effects to control for differences across institutions.

If the posttest alone is the dependent variable, the model estimates the effects of the control variables on student achievement at the time at the end of the course. If the pretest is used as the dependent variable, the results will show the effects of the regressors on student achievement at the time of the pretest or the start of a course. Regardless of the dependent variable used for the model, what is being estimated is the effect of the variable on the *stock* of economics understanding at a point in time (at posttest or pretest).

The *flow* of economic understanding during a course also is of value for studying what students gain, lose, retain, or simply do not know. Although the aggregate difference score is supposed to capture those changes for estimation purpose, the analysis of learning components for the difference score shows that it reduces to PL − NL. If NL is relatively small, then regression results for the difference score as the dependent variable or PL as the dependent variable will be similar because the difference score is only a slightly smaller PL. As NL increases in size, however, the regression results are likely to differ substantially depending on whether the difference score is the dependent variable or PL is the dependent variable. Such an outcome will be the case with the TUCE data because as shown by the restricted descriptive statistics in tables 2 and 3, NL is 56 percent of the size of PL in the micro sample (4.36 versus 7.77) and 48 percent of the size of PL in the macro sample (4.12 versus 8.53). The general conclusion is that the regression results with PL as the dependent variable should be more informative about factors that influence what students learn in a course because they reflect only the PL response pattern to test items whereas the difference score is influenced by both PL and NL response patterns.

The RL regression results also should be of interest in identifying factors that are most likely associated with what students already know at the start of the course and are able to retain during the course. The effects of the control variables in the RL should be more comparable with the effects of those variables on pretest scores because the pretest is (RL + NL). If NL is smaller than RL, then RL will be the dominant learning component in the pretest. The restricted TUCE data in tables 2 and 3 indicate that RL is dominant but only slightly larger than NL (5.04 versus 4.36 in micro; 5.71 versus 4.12 in macro) suggesting that there will be differences in coefficient effects depending on whether the pretest or RL is the dependent variable. By contrast, if NL, which measures what students are not able to retain, is used as the dependent variable, the regression effects of variables should be opposite of when RL is the dependent variable.

The fourth learning component to be studied with the regression analysis is ZL, which shows what students do not know at the pretest or the posttest. The number wrong on the posttest is ZL + NL. If ZL dominates NL, then ZL accounts for most of the wrong answers on the posttest, which is the case with the restricted TUCE data (12.83 versus 4.36 in micro; 11.64 versus 4.12 in macro). As a result, a regression with the ZL score as the dependent variable or a regression with the posttest score as the dependent variable will generally produce coefficients of a similar size, but the effects will be opposite of each other because as the posttest score rises, the ZL score falls, and vice versa.

One final point from the regression analysis is the expectation for the coefficient estimates for the test scores. The disaggregation of aggregate test scores in learning components splits the posttest score into RL + PL, divides the pretest score into RL + NL, and shows that the difference score to be PL − NL. The regression results will show that the coefficient for a variable in the posttest equation is the sum of the coefficients for the same variable in the RL and PL equations. The coefficient for a variable in the pretest equation is the total of RL and NL coefficients. The coefficient for a variable in the difference score equation is the PL minus the NL coefficient.

REGRESSION RESULTS

Table 4 reports the regression results for the micro TUCE sample. To see the value of the learning analysis, consider the equation results for the posttest (column 1) and its learning components, RL (column 4) and PL (column 5). In the case of the gender variable, as expected, the male coefficient (1.0772) equals the sum of the coefficients for RL (0.6660) and PL (0.4112). This analysis shows that the majority of the male effect on the posttest is from RL (62 percent) rather than PL (38 percent). This gender effect is attributable largely to prior economic understanding as has been found in previous studies (Siegfried 1979; Becker and Powers 2001).

[Insert table 4 about here]

A similar conclusion about the larger contribution of RL to the posttest score can be drawn by reviewing the relative size of RL and PL coefficients for other variables. White is positive and statistically significant in the RL equation, but not in the PL equation, which suggests that this white effect is 83 percent RL. For other variables, a coefficient is negative and statistically significant in the RL equation and there is no significant effect in the PL equation: business major (78 percent RL); full-time enrolled (70 percent RL), and attending a two-year college (93 percent RL). Still other variables have a positive and statistically significant coefficient in both the RL and PL equations, but the contribution from RL is larger: freshman (62 percent RL); 3.0 to 4.0 GPA (72 percent RL); and, took a college calculus course (63 percent). With a few other variables, coefficient signs and statistical significant are mixed: the effect of taking a high school economics course is positive and statistically significant in the RL equation, but negative and insignificant in the PL equation (−9 percent); the effect of being an English language speaker is insignificant and negative in the RL equation, but significant and negative in the PL equation.

As expected the regression analysis shows that the variable coefficients in the posttest equation are most comparable with their counterparts in the ZL equation (column 7), but with the reverse sign for each coefficient. The posttest score measures the number of right answers. The number wrong is the combination of ZL + NL. Since ZL is the major part of the number wrong score, it will vary in the opposite direction from the posttest score, so the coefficients across variables in the two equations largely mirror each other.

The coefficients in the pretest equation (column 2) are the sum of the coefficients for RL (column 4) and NL (column 6). The male coefficient of 0.6271 [(0.6660) + (−0.0389)], which indicates that RL explains about all the gender effect for the pretest. The NL results also suggest that females are no more likely than males to give a correct answer on the pretest, but then give an incorrect answer on the posttest, or show negative learning during a course.

Some variables that are statistically significant in the pretest equation also are statistically significant in the RL equation but not in the NL equation. These results indicate that RL explains most of the coefficient effect on the pretest for these variables: white (0.6945); business major (−0.5003); took a college calculus course (0.4134); and, took high school economics (0.3655). A statistically significant RL effect, however, can be offset by a statistically significant NL effect, as shown in the variable results for freshman and full-time enrolled (which are both insignificant in the pretest), GPA, and attending a two-year college. By contrast, the pretest effect for took a college economics course (0.8628) largely comes from a RL contribution (75 percent) that is augmented by a NL contribution (25 percent). This split indicates that a quarter of the pretest effect is from students who had taken a prior economics course and gave correct pretest answers to some items, but could not retain that economic understanding on the posttest.

Turning to the results for the difference score (column 3), there are six statistically significant variables: male; freshman; English speaker; high GPA; took a college calculus course; and, took a college economics course. The relevant learning components for the comparative analysis for the difference score are PL (column 5) and NL (column 6). For the male coefficient, most of the effect of 0.4501 comes from PL (0.4112). PL also accounts for most of the contribution to the difference score coefficients for English speaker and took a college calculus course. For the other three variables (freshman, high GPA, and took a college economics course), the coefficient effects for the difference score may be misleading because of a large and significant NL effect. A more reasonable approach for assessing learning from pretest to posttest would be to focus on the PL results because they are based on only one type of learning and are not contaminated by a NL factor.

To add robustness to the regression analysis, the macro TUCE data are analyzed in the same way as the micro TUCE data (table 5). In the posttest analysis for macro (column 1), the variables for male, white, full-time enrolled, English speaker, high GPA, took a college calculus course, and attended a two-year college are all statistically significant. The results for the posttest coefficient for male indicate that about 78 percent of this gender effect is from RL and 22 percent is from PL. RL also had the greater effect than PL on the posttest score for GPA (61 percent versus 38 percent) and took a college calculus course (69 percent versus 21 percent). The influence of being white on the posttest score is about equally divided between RL and PL. The negative effects on the macro posttest of full-time enrolled, English speaker, and attending a two-year college are mostly affected by the influence of RL.

[Insert table 5 about here]

The results for the macro pretest (column 2) show that the statistically significant variables of male, white, business major, full-time enrollment, GPA, took college calculus, took college economics, and attended a two-year college are essentially explained by the effects of those variables in RL (column 4) and not NL (column six). When the NL coefficients are statistically significant (full-time enrolled, high GPA, and took a college calculus course), they have an opposite effect than do these same significant variables in the RL equation, thus serving to lower the size of the pretest coefficient.

The third comparison to make with the macro data is with the difference score and PL estimates. For gender, those coefficient estimates are about the same, but male is statistically significant only in PL. Also, when coefficients are positive and statistically significant, as they are for white, business major, GPA, and took a college calculus course, the difference score estimates are larger than PL estimates because of the NL factor affecting the difference score. Conversely, when coefficients are negative and statistically significant, as they are for full-time enrolled, English speaker, and took a college economics course, difference score estimates are smaller than the PL coefficients, again because of how NL affects the calculation. These results show that if the purpose of a regression analysis is to recognize how variables affect student learning during an economics course, the PL estimation appears to offer less biased estimates because the analysis is based on only one type of student learning.

CONCLUSION

In this study, we offer a comprehensive method for using pretest and posttest data multiple choice format to analyze student understanding and learning in economics. The four possible response patterns for student answers on a pretest and a posttest define scores for four types of learning: Positive (PL) (incorrect pre and correct post); Retained (RL) (correct pre and correct post); Negative (NL) (correct pre and incorrect post); and Zero (ZL) (incorrect pre and incorrect post). The posttest score can be defined as the combination of PL + RL. The pretest score is the total of RL + NL. The difference score or value-added score is PL − NL. The number wrong score on the posttest is ZL + NL and the number wrong on the pretest is ZL + PL.

It is common practice in economic education research to use a difference or value-added score as a measure of learning in a regression that is estimating effects of control variables on student learning. This use of the difference score as the dependent variable, however, has potential problems because it is a combination of a desirable type of learning (PL) and an undesirable type of learning (NL). This positive and negative combination makes it difficult to interpret this difference score because there will always be an adjustment to PL that could cause an estimate for a variable to be overstated or understated relative to a PL estimate. A better alternative for measuring the desirable part of learning is to use PL estimates as indicated by the results from the regression analysis with the micro and macro TUCE data.

The PL analysis also can contribute new insights to regression studies that use the posttest or the pretest as a dependent variable when matched pretest and posttest data are available for the analysis. This contribution arises because the posttest is composed of RL and PL and the pretest is a mixture of RL and NL, so through the analysis of learning components it becomes possible to identify the type of learning affecting variables that influence the posttest or pretest outcome.

REFERENCES

Allgood, S., W. B. Walstad, and J. J. Siegfried. 2015. Research on teaching economics to undergraduates. *Journal of Economic Literature* 53 (2): 285–325.

Becker, W. E. 1997. Teaching economics to undergraduates. *Journal of Economic Literature* 35 (3): 1347–73.

Becker, W. E., and J. R. Powers. 2001. Student performance, attrition, and class size given missing data. *Economics of Education Review* 20 (4): 377–88.

Dickie, M. 2006. Do classroom experiments increase learning in introductory microeconomics? *Journal of Economic Education* 37 (3): 267–88.

Koedel, C., K. Mihaly, and J. Rockoff. 2015. Value-added modeling: A review. *Economics of Education Review* 47 (August): 180–95.

Maxwell, N. L., J. R. Mergendoller, and Y. Bellisimo. 2005. Problem-based learning and high school macroeconomics: A comparative study of instructional methods. *Journal of Economic Education* 36 (4): 315–31.

Petscher, Y., and C. Schatschneider. 2011. A simulation study on the performance of simple difference and covariance-adjusted scores in randomized experimental designs. *Journal of Educational Measurement* 48 (1): 31–43.

Saunders, P., and J. R. Powers. 1995. Reallocating content coverage in principles of microeconomics to increase student learning. *American Economic Review* 85 (2): 339–42.

Siegfried, J. J. 1979. Male-female differences in economics education: A survey. *Journal of Economic Education* 10 (2): 1–11.

Siegfried, J. J., and R. Fels. 1979. Research on teaching college economics: A survey. *Journal of Economic Literature* 17 (3): 923–69.

Walstad, W. B., M. Watts, and K. Rebeck. 2007. *Test of understanding of college economics: Examiner's manual.* 4th ed. New York: Council for Economic Education.

APPENDIX 1: TUCE Item Data for Positive Learning

	Microeconomics				Macroeconomics			
Item	Posttest Proportion Correct	Retained Learning	Positive Learning	PL as % of Post	Posttest Proportion Correct	Retained Learning	Positive Learning	PL as % of Post
1	0.501	0.244	0.257	**51.3%**	0.530	0.168	0.362	**68.3%**
2	0.397	0.171	0.226	**57.0**	0.607	0.348	0.259	**42.7**
3	0.502	0.210	0.292	**58.2**	0.688	0.338	0.350	**50.9**
4	0.569	0.084	0.486	**85.3**	0.458	0.219	0.239	**52.2**
5	0.457	0.198	0.258	**56.5**	0.592	0.079	0.513	**86.7**
6	0.456	0.149	0.307	**67.3**	0.469	0.158	0.311	**66.3**
7	0.490	0.283	0.207	**42.3**	0.600	0.333	0.267	**44.5**
8	0.368	0.142	0.226	**61.4**	0.497	0.234	0.262	**52.8**
9	0.309	0.087	0.222	**71.9**	0.326	0.088	0.238	**73.0**
10	0.439	0.206	0.233	**53.0**	0.407	0.169	0.238	**58.6**
11	0.315	0.053	0.263	**83.3**	0.593	0.261	0.332	**55.9**
12	0.451	0.118	0.333	**73.8**	0.551	0.239	0.313	**56.7**
13	0.504	0.213	0.292	**57.8**	0.635	0.378	0.257	**40.5**
14	0.453	0.143	0.310	**68.4**	0.475	0.139	0.336	**70.6**
15	0.338	0.114	0.225	**66.4**	0.612	0.331	0.281	**46.0**
16	0.497	0.245	0.251	**50.6**	0.380	0.112	0.269	**70.7**
17	0.433	0.167	0.266	**61.4**	0.368	0.128	0.239	**65.1**
18	0.411	0.150	0.261	**63.4**	0.449	0.087	0.361	**80.6**
19	0.435	0.227	0.208	**47.7**	0.396	0.147	0.249	**62.9**
20	0.307	0.060	0.247	**80.6**	0.605	0.356	0.248	**41.1**
21	0.449	0.222	0.226	**50.4**	0.418	0.119	0.299	**71.5**
22	0.592	0.355	0.237	**40.0**	0.328	0.092	0.236	**71.9**
23	0.312	0.104	0.208	**66.6**	0.364	0.107	0.256	**70.5**
24	0.488	0.221	0.268	**54.8**	0.328	0.103	0.225	**68.6**
25	0.335	0.116	0.219	**65.3**	0.600	0.286	0.314	**52.3**
26	0.343	0.113	0.229	**66.9**	0.308	0.078	0.230	**74.7**
27	0.409	0.151	0.258	**63.2**	0.325	0.076	0.249	**76.6**
28	0.349	0.107	0.242	**69.3**	0.509	0.209	0.300	**59.0**
29	0.371	0.145	0.226	**61.0**	0.341	0.115	0.226	**66.3**
30	0.493	0.234	0.259	**52.6**	0.439	0.177	0.262	**59.7**
Average	**0.426**	**0.168**	**0.258**	**60.6**	**0.473**	**0.189**	**0.284**	**60.0**

TABLE 1: TUCE Variables

Variables	Variable Description
Learning Variables	
Posttest score	Number of correct answers out of 30 on the posttest
Pretest score	Number of correct answers out of 30 on the pretest
Post−Pre	Difference between the posttest and pretest scores
PL	Positive Learning—Incorrect pretest, correct posttest
RL	Retained Learning—Correct pretest and posttest
NL	Negative Learning—Correct pretest, incorrect posttest
ZL	Zero Learning—Incorrect pretest and posttest
Student Characteristics	
Male	1=male; 0=female
White	1=white; 0=else
Freshman	1= freshman; 0=else
Business Major	1=business major; 0=else
Full-Time enrolled	1=full-time; 0=part-time
English Speaker	1= communicates better in English; 0=does not
3.00–4.00 GPA	1=3.00–4.00 GPA; 0=less than 3.0 GPA
Took college calculus	1=1 or more college calculus courses taken; 0=no courses
Took college economics	1=1 or more college economics courses taken; 0=no courses
Took high school economics	1=high school economics course taken; 0= no course
Attended a 2-year college	1=2-year or associate's college; 0=else

TABLE 2: Micro TUCE Descriptive Statistics

	Unrestricted			Restricted		
	Count	Mean	Std. Dev.	Count	Mean	Std. Dev.
Micro pretest score	3255	9.3849	3.3162	3052	9.3994	3.2919
Micro posttest score	3255	12.7736	4.6816	3052	12.8070	4.6871
Micro post−pre	3255	3.3886	4.4206	3052	3.4076	4.4232
Retained learning (RL)	3255	5.0320	3.1944	3052	5.0364	3.1904
Positive learning (PL)	3255	7.7416	3.0839	3052	7.7706	3.0856
Negative learning (NL)	3255	4.3530	2.1675	3052	4.3630	2.1723
Zero learning (ZL)	3255	12.8734	4.1686	3052	12.8300	4.1595
Male	3232	0.5718	0.4949	3052	0.5685	0.4954
White	3219	0.6847	0.4647	3052	0.6923	0.4616
Freshman	3241	0.2379	0.4259	3052	0.2382	0.4261
Business major	3186	0.5138	0.4999	3052	0.5164	0.4998
Full-time enrolled	3212	0.9278	0.2589	3052	0.9276	0.2592
English speaker	3219	0.8524	0.3547	3052	0.8552	0.3520
3.00-4.00 GPA	3213	0.5338	0.4989	3052	0.5337	0.4989
Took college calculus	3232	0.6463	0.4782	3052	0.3555	0.4787
Took college economics	3234	0.6054	0.4888	3052	0.3938	0.4887
Took high school econ	3204	0.4416	0.4967	3052	0.4413	0.4966
Attended 2-year college	3255	0.0774	0.2673	3052	0.0803	0.2718

TABLE 3: Macro TUCE Descriptive Statistics

	Unrestricted			Restricted		
	Count	Mean	Std. Dev.	Count	Mean	Std. Dev.
Macro pretest score	2789	9.8017	3.4820	2687	9.8292	3.4883
Macro posttest score	2789	14.1979	5.2878	2687	14.2374	5.2821
Macro post−pre	2789	4.3962	4.8049	2687	4.4083	4.7907
Retained learning (RL)	2789	5.6744	3.6261	2687	5.7071	3.6297
Positive learning (PL)	2789	8.5235	3.3700	2687	8.5303	3.3566
Negative learning (NL)	2789	4.1273	2.1928	2687	4.1221	2.1918
Zero learning (ZL)	2789	11.6748	4.4738	2687	11.6405	4.4738
Male	2776	0.5947	0.4910	2687	0.5932	0.4913
White	2777	0.7105	0.4536	2687	0.7142	0.4519
Freshman	2784	0.1724	0.3778	2687	0.1727	0.3780
Business major	2762	0.5261	0.4994	2687	0.5277	0.4993
Full-time enrolled	2768	0.9292	0.2566	2687	0.9293	0.2564
English speaker	2774	0.8792	0.3259	2687	0.8817	0.3231
3.00–4.00 GPA	2773	0.5424	0.4983	2687	0.5415	0.4984
Took college calculus	2780	0.3755	0.4843	2687	0.3792	0.4853
Took college Economics	2782	0.3404	0.4739	2687	0.3420	0.4745
Took high school econ	2764	0.4005	0.4901	2687	0.4012	0.4902
Attended 2-year college	2789	0.0667	0.2495	2687	0.0674	0.2507

Table 4: Micro TUCE Regressions

	(1) Posttest	(2) Pretest	(3) Post−Pre	(4) RL	(5) PL	(6) NL	(7) ZL
Male	1.0772***	0.6271***	0.4501***	0.6660***	0.4112***	−0.0389	−1.0383***
	(0.154)	(0.117)	(0.153)	(0.110)	(0.107)	(0.079)	(0.140)
White	0.8760***	0.6945***	0.1816	0.7238***	0.1522	−0.0294	−0.8467***
	(0.190)	(0.144)	(0.188)	(0.136)	(0.131)	(0.098)	(0.173)
Freshman	0.7918***	0.1856	0.6061***	0.4917***	0.3000**	−0.3061***	−0.4857**
	(0.214)	(0.162)	(0.212)	(0.153)	(0.148)	(0.110)	(0.194)
Business major	−0.6393***	−0.5003***	−0.1390	−0.5013***	−0.1380	0.0010	0.6383***
	(0.164)	(0.124)	(0.162)	(0.117)	(0.113)	(0.084)	(0.149)
Full-time enrolled	−0.6283**	−0.3140	−0.3142	−0.4384**	−0.1899	0.1244	0.5039*
	(0.310)	(0.235)	(0.307)	(0.221)	(0.214)	(0.159)	(0.282)
English speaker	−0.6301***	−0.1727	−0.4575**	−0.1941	−0.4361***	0.0214	0.6088***
	(0.219)	(0.166)	(0.217)	(0.156)	(0.151)	(0.113)	(0.199)
3.00–4.00 GPA	1.3875***	0.7818***	0.6057***	0.9921***	0.3954***	-0.2103***	−1.1772***
	(0.154)	(0.117)	(0.153)	(0.110)	(0.107)	(0.079)	(0.140)
Took college calculus	0.8494***	0.4134***	0.4359***	0.5373***	0.3120***	−0.1239	−0.7255***
	(0.163)	(0.123)	(0.161)	(0.116)	(0.112)	(0.084)	(0.148)
Took college economics	0.2733	0.8628***	−0.5895***	0.6492***	−0.3759***	0.2136**	−0.4869***
	(0.186)	(0.141)	(0.184)	(0.133)	(0.129)	(0.096)	(0.169)
Took high school econ	0.3111*	0.3655***	−0.0544	0.3398***	−0.0287	0.0257	−0.3368**
	(0.162)	(0.123)	(0.160)	(0.116)	(0.112)	(0.083)	(0.147)
Attended 2-year college	−2.2972*	−0.7893	−1.5078	−2.1318**	−0.1654	1.3424*	0.9548
	(1.365)	(1.033)	(1.351)	(0.975)	(0.944)	(0.702)	(1.240)
School fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	10.0940	9.7374	0.3566	4.6686	5.4254	5.0688	14.8372
	(1.095)	(0.828)	(1.084)	(0.782)	(0.757)	(0.563)	(0.994)
Pseudo R^2	.2533	.1343	.1789	.1781	.1764	.0818	.2180
Observations	3052	3052	3052	3052	3052	3052	3052

Standard errors in parentheses
*$p < .10$; **$p < .05$; ***$p < .01$

TABLE 5: Macro TUCE Regressions

	(1) Posttest	(2) Pretest	(3) Post−Pre	(4) RL	(5) PL	(6) NL	(7) ZL
Male	1.1712***	0.9049***	0.2664	0.9140***	0.2573**	−0.0091	−1.1622***
	(0.175)	(0.134)	(0.175)	(0.131)	(0.123)	(0.086)	(0.153)
White	1.1365***	0.4553***	0.6813***	0.5639***	0.5727***	−0.1086	−1.0280***
	(0.209)	(0.160)	(0.209)	(0.156)	(0.146)	(0.102)	(0.182)
Freshman	0.0986	0.2851	−0.1866	0.3012	−0.2026	−0.0161	−0.0825
	(0.269)	(0.206)	(0.269)	(0.201)	(0.189)	(0.132)	(0.235)
Business major	−0.0717	−0.4074***	0.3357*	−0.3364**	0.2647**	−0.0710	0.1427
	(0.177)	(0.136)	(0.177)	(0.132)	(0.124)	(0.087)	(0.155)
Full-time enrolled	−1.1467***	−0.5235**	−0.6232*	−0.9083***	−0.2383	0.3849**	0.7618**
	(0.347)	(0.265)	(0.347)	(0.258)	(0.243)	(0.170)	(0.302)
English speaker	−0.6446**	−0.1517	−0.4929*	−0.4435**	−0.2011	0.2918**	0.3528
	(0.264)	(0.202)	(0.264)	(0.196)	(0.185)	(0.129)	(0.230)
3.00-4.00 GPA	2.1291***	1.0390***	1.0902***	1.3906***	0.7385***	−0.3517***	−1.7775***
	(0.174)	(0.133)	(0.174)	(0.130)	(0.122)	(0.085)	(0.152)
Took college calculus	1.2243***	0.5731***	0.6512***	0.8390***	0.3852***	−0.2659***	−0.9583***
	(0.196)	(0.150)	(0.196)	(0.146)	(0.137)	(0.096)	(0.171)
Took college economics	0.1749	0.5930***	−0.4181*	0.4994***	−0.3245**	0.0937	−0.2686
	(0.220)	(0.168)	(0.220)	(0.164)	(0.154)	(0.108)	(0.192)
Took high school econ	0.1191	0.1409	−0.0219	0.0560	0.0631	0.0850	−0.2040
	(0.186)	(0.142)	(0.186)	(0.138)	(0.130)	(0.091)	(0.162)
Attended 2-year college	−5.2090***	−3.6870***	−1.5220	−4.8059***	−0.4030	1.1190	4.0900***
	(1.628)	(1.244)	(1.628)	(1.212)	(1.140)	(0.796)	(1.419)
School fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	18.9829	14.9612	4.0216	11.6455	7.3374	3.3157	7.7014
	(1.498)	(1.145)	(1.498)	(1.116)	(1.049)	(0.733)	(1.306)
Pseudo R^2	.3464	.1245	.2057	.2328	.2063	.0926	.3079
Observations	2687	2687	2687	2687	2687	2687	2687

Standard errors in parentheses
*$p < .10$; **$p < .05$; ***$p < .01$