Presented at the DLSU Research Congress 2015
De La Salle University, Manila, Philippines
March 2-4, 2015

"Math", "intuition", and "authentic learning" in Economics: Evidence from a randomised controlled trial

Gerardo L. Largoza[1,*], John Adrian F. Sarao[1], Stephen S. Laohoo[1] and Jose Margelle T. Cruz[1]
[1]*School of Economics, DLSU*
[*]*Corresponding Author: john_adrian_sarao@dlsu.edu.ph*

Abstract: Though academic research and postgraduate training in Economics have without doubt become more "formal" or mathematical in the last fifty years, questions remain as to appropriate mix of math (equations) and "intuition" (words and graphs) in undergraduate teaching. There are three inter-related issues: first, the nature of the discipline, which comprises scientific theory, measurement and testing, but also history, ideas, philosophy, and policy – all within a curriculum that cannot be expanded indefinitely. Second, student demographics that indicate widespread interest in economic issues but less interest in the increasing formalism that has seeped into even the earliest levels of the curriculum. Finally, and our main interest in this paper, the attainment of "authentic learning": if we at DLSU, like a growing number of universities overseas, encourage learning that allows students to perform "real-world tasks that demonstrate meaningful application of essential knowledge and skills" (Mueller, 2014), then it's worth investigating the appropriate mix of formalism and intuition that we use to achieve this type of learning.

In this paper, we provide preliminary yet clean evidence based on a randomized controlled trial conducted among 70 Economics majors enrolled in a research methods class. To uphold the crucial SUTVA (single unit treatment value assumption), we create treatments based formalist, intuitive, and mixed approaches used by the same faculty on a single concept (selection bias), using a workflow that minimizes treatment "contamination". We then examine the impact of these treatments on student performance in traditional test scores and modes of authentic assessment. Finally, we look at the impact of the treatments on behavioral indicators such as effort and subjective well-being.

Key Words: Authentic assessment; formalism in Economics; randomized controlled trial (RCT)

1. INTRODUCTION

This paper is in the scholarship of teaching and learning in Economics. We address the pedagogical concern of whether or not an intuitive or formalist approach makes a difference in facilitating traditional and authentic learning outcomes.

In economics, the amount of math in its courses has once become the center of discussion (Krugman, 1998; Chang & Aldred, 2014). It is argued that the approach can affect learning outcomes (Mearman, Wakely, Shoeb, & Webber, 2006), and we suspect that it also affects student well-being. While there is currently some student unrest against the approach (Chang & Aldred, 2014; Inman, 2014), mathematics nonetheless plays an important role in economics (Mankiw, 2006) and will continue to do so in the foreseeable future.

There is an apparent lack of empirical evidence on the subject. Perhaps it is due to the difficult of providing clean evidence for how teaching style in economics affects learning outcomes. We did not come across any studies on how teaching approaches can affect well-being.

To find out whether teaching approach can affect traditional and authentic assessment outcomes, we conduct a week-long classroom experiment wherein three groups of students are taught the theory of counterfactuals through an intuitive approach, a mathematical approach, and a

mixed approach. Their learning outcomes and well-being are then measured.

1.1 A Brief History of the Use of Mathematics in Economics

Formalism is the method of using math to explain and logically verify theories. It's a central part of economics methodology and is why economics has such a substantial amount of math. But it hasn't always played such a large role in the discipline, and this approach became a result of economics' attempt to model itself after physics (Callahan & Leeson, 2006).

Now, to be able to understand economic theory and read economics journals, a student in economics needs math Greg Mankiw (2006). The greater use of mathematics has given programs more impetus to increase the amount of math undergraduate economics students receive. But any curriculum can have only so many courses, and additional math courses mean less room for other subjects.

1.2 General Complaints and Arguments

Recently, students across 19 countries have banded together, clamoring for changes in the way economics is taught. They describe the math-centric approach as appearing to be divorced from the real world, unable to address the problems of the 21st century (Inman, 2014).

In the words of two Cambridge academics:

> "…what makes economics so unique is the fact that it is the only academic discipline in which a significant and increasing number of students are in an open revolt against the content of their degree courses (Chang & Aldred, 2014)."

It is clear that students interested in economics topics have qualms with the increasing formalism.

1.3 Authentic vs. traditional assessment

One of the problems of traditional schooling is that the work given to students often has no value outside of the school setting (Newmann & Wehlage, 1993). In undergraduate economics, we can see this when a student passes a course like statistics despite not being able to apply the skills outside of the school setting and into the real world.

One of the reasons this happens is because curricula can often be fixated on traditional forms of assessment, which are characterized as 'forced-choice' questions like multiple choice and true-or-false tests. These tests rely heavily on one's ability to recall.

On the other hand, there are also authentic forms of assessment, and the philosophy behind this is to see people as needing to "be capable of performing meaningful tasks in the real world." These are usually task-oriented and assess whether students can perform meaningful tasks that resemble challenges in the real world (Mueller, 2014).

De La Salle University's Expected Lasallian Graduate Attributes (ELGA), a set of qualities that, according to the university, should be embodied by every Lasallian graduate as an indicator of higher order thinking skills, are more in line with the goals of authentic. Also. as students struggle with math (Mearman, et al. 2006; Chang & Aldred, 2014), it might prevent them from achieving authentic learning.

2. METHODOLOGY

We randomly divide all DLSU economics students currently taking the Methods of Research (METRESE) course into three different groups. They are taught the concept of counterfactuals and selection bias over the course of a week, before being tested for their learning outcomes and well-being. We use the lesson of counterfactuals and selection bias because it's a lesson that is amenable to different teaching styles and is important to students of economics. We further break down each aspect of our design in the subsequent sections.

2.1 Treatment Groups

The first group we label as the intuitive group and receive a graph oriented and thus more intuitive approach to the lesson. This is our control group since the professor teaching each class, Dr. Gerardo Largoza, conventionally teaches using this approach. The second group we label as the formalist group and are taught through expected values, a more mathematical approach. The last group we label as the mixed group and are taught using a mix of both approaches. There are 25 students assigned to the the intuitive and formalist groups, and 24 to the mixed group.

Each class follows uniform schedule. The first 30 minutes are allotted for lecture, the next 30 minutes for self-review, and the last 30 minutes for administering the exam and surveys.

This setup is unusual since students are not usually tested the same day a lesson is taught to them, but we proceed this way to satisfy the SUTVA assumption, that is making sure that

treatment effects are contained within treatment groups.

During the second day, some time is allotted for administering Kahneman's instrument at the end of the period to measure well-being.

The lesson is divided into two parts. For the first class session, counterfactuals and the problem of selection bias are explained. For the second class session, the conditions for solving selection bias are explained.

Students are told that the test will serve as their midterm exam so students have an incentive to diligently answer the test. They are, however, aware that they are subjects of an experiment.

2.2 Tests Administered

The tests administered on the first day is focused on traditional assessment. There are 10 questions in total, and as a bonus, students are given the option to explain their answers for each question.

Each question has a formalist and intuitive version. Those under the intuitive and formalist groups have their questions split 70-30, the majority being in the style of their treatment. We use this split because we attempt to mimic the set-up of a departmental exam wherein a student is only familiar with about 70% of the questions, with the remaining 30% left for students to figure out on their own. The formalist test has a 50-50 split.

The test administered on the second day includes both traditional and authentic means of assessment and is somewhat consistent among the three groups. This test has eight questions in total.

In two multiple-choice questions, we illustrate real-world situations and ask students to identify sources of selection bias. Each situational question is followed up with an open-ended question that asks them to explain the situation using the approach taught to them, i.e. students under the intuitive approach are asked to explain using graphs while formalist students are expected to use expected values. Those under the mixed approach can choose between the two. We consider the situational questions and the follow-up questions as both traditional and authentic.

The last question, considered authentic, asks students to consider their own thesis proposals and identify possible sources of selection bias.

Moreover, the questions used are amenable to Bloom's revised taxonomy, and we categorize each question into three of the lower half levels of Bloom's taxonomy: remembering, understanding, and evaluating. A student in any of the groups answers eight questions under remembering, five questions under understanding, and four questions under applying.

To measure well-being, we use a variation of Kahneman's Day Reconstruction Method. We administer a survey where students list down the activities they engaged in throughout the day for both the Tuesday and Thursday of the experiment week and assign the average utility gained or lost per average minute (U.P.A.M.) for each activity, between a range of -100 to +100.

The merit of Kahneman's instrument lies in the fact that respondents are unaware of which activity we're actually interested in, negating the effects of any self-awareness. We convert each student's (U.P.A.M.) to standard normal values.

3. RESULTS AND DISCUSSION

Among the 79 students expected to participate, only 69 do. The number of students who turn up on both days for each group is 27, 23, and 20 for the formalist, intuitive, and mixed groups, respectively. The difference is due to some students being absent, while some students attend a class different from the one assigned.

There are slightly more females than males in the population, and the average cumulative grade point average for the population is 2.82 (4.0 being the highest mark). There is no statistically significant difference in grades or gender between groups.

In the following figures illustrating our regression results, the control group, the intuitive group, is used as the baseline, and the treatment groups are expressed in standard deviations away from the baseline. Also, *** indicates statistical significance at the 99% level, ** at the 95% level, and * at the 90% level. Conventionally in educational studies, a .1 difference in standard deviation is considered as slight, .2-.3 as moderate, and .5 as large (Greene, 1997).

5.1 Does an Intuitive, Formalist, or Mixed Approach to Teaching Affect Scores in Traditional Assessment?

In traditional assessment tests, both with or without the bonus, students under the mixed treatment group score the highest, followed by the formalist treatment group, and then the intuitive treatment group. Using the intuitive group as the baseline, the mixed treatment amounts to a gain of 0.76σ in correct answers, while the formalist treatment amounts to a gain of 0.42σ, both without the bonus. The mixed treatment effect is significant

at the 95% level while the formalist treatment effect is not significant even at the 90% level. When we control for gender and CGPA, the mixed treatment effect is still statistically significant at the 90% level but no longer at the 95% level.

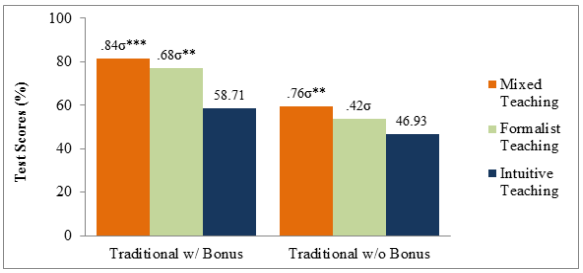

Fig. 1. Traditional test scores with bonus

When we take into account the bonus, the treatment ranking remains the same; the treatment effects, however, become slightly more pronounced.

5.2 Does an Intuitive, Formalist, or Mixed Approach to Teaching Affect Scores in Authentic Assessment?

There are two ways to check the performance in the authentic portion. The first is to focus on the lone 'purely' authentic question; the second is to add to the first the set of questions that are considered as both authentic and traditional; we'll call this second 'semi-authentic.'

For the purely authentic question, the mixed treatment scores the highest followed by the formalist treatment, but the effects are not statistically significant.

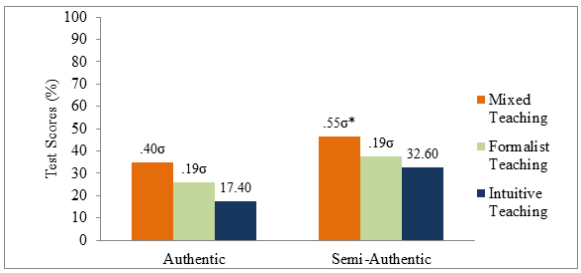

Fig. 2. Authentic test scores

For the semi-authentic questions, the mixed treatment performs best, followed by the formalist. The mixed group has a score .55σ higher than the baseline, the only statistically significant treatment effect at the 90% level. The level of significance is the same when we control for the average grade for the given economics subjects and gender, but it drops when CGPA is used instead of the average grade.

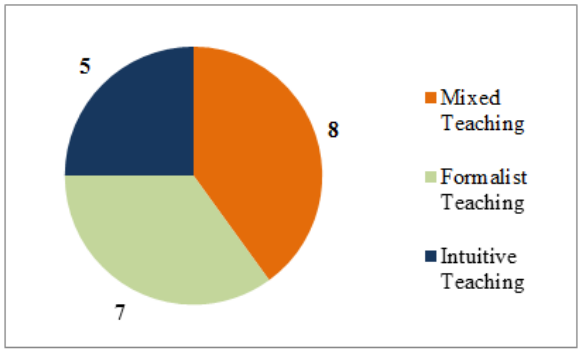

Fig. 3. Treatment group of semi-authentic passers

It is worth noting that the average of each class is below a passing rate of 60%. Individually, only 20 out of the 70 students passed. Of the ones who passed, 8 are from the mixed class, 7 from the formalist class, and 5 from the intuitive class. The CGPA of those who passed are slightly higher than those who failed. Finally, there are 11 females and 9 males.

5.3 How do Teaching Styles Affect the Performance of Students in Specific Levels of Bloom's Taxonomy?

The data shows that the mixed treatment group scores the highest across the three levels of thinking, though only the treatment effect in remembering questions is substantial with a gain of .87σ, significant at the 99% level.

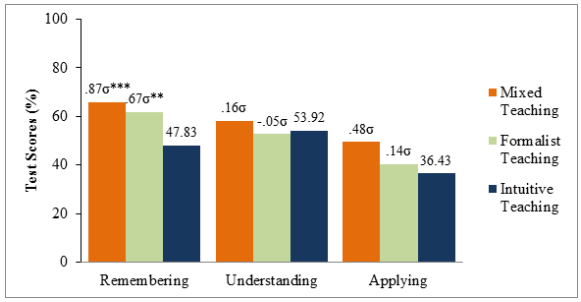

Fig. 4. Average test scores per Bloom's Levels

The formalist treatment exhibit mixed results. The only significant formalist treatment effect is in remembering, with a gain of .67σ, significant at the 95% level. Also, the formalist group actually scores lower in the understanding category while scoring higher in the applying

category; both results, however, are less than .2σ in magnitude.

5.4 Does Teaching Style Affect Whether or Not Students Explain Their Answers for Bonus Points?

We use the number of optional explanations attempted by a student during the first exam, regardless of correctness, as an indicator of effort.

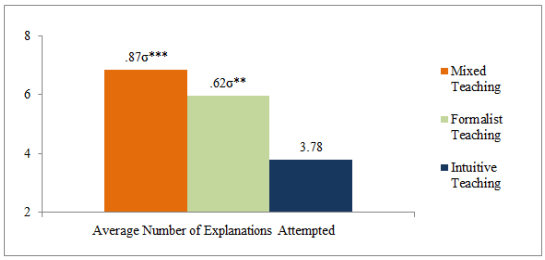

Fig. 5. Avereage number of explanations per style

On average, students attempt to explain their answers on 5.5 out of the 10 possible questions. The mixed treatment group has the highest average at 6.85, a 0.87σ difference from the baseline, significant at the 99% level. Meanwhile, those under the formalist teaching have a gain of 0.62σ from the baseline, significant at the 95% level. When controlling for gender and CGPA, the effect of both treatments remain above half a standard deviation, both significant at the 95% level.

5.5 Do Those Under the Mixed Treatment Group Prefer Using Intuitive or Formalist Modes of Expression in Explaining?

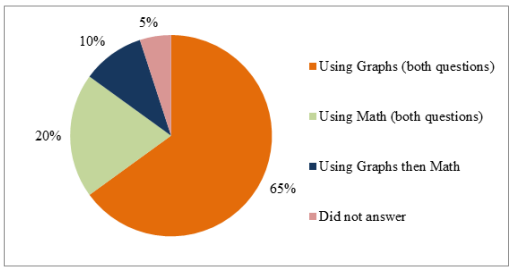

Fig. 6. Choice of method in explaining

In the two questions where the 20 students under the mixed treatment group are given the choice to explain situations either in terms of expected values or graphs, 13 consistently prefer to explain using graphs in both questions, 4 consistently prefer to explain using expected values, 2 use graphs in the first question before switching over to expected values in the second question, and 1 (5%) does not answer both questions.

5.7 Which Teaching Style Gives Students the Most Utility Per Average Minute?

The mixed treatment has the highest z-score, followed by the formalist, and then the intuitive group.

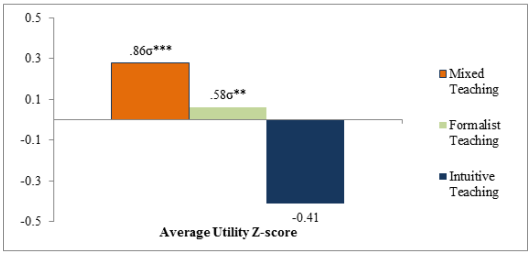

Fig. 7. Utility z-score based on teaching style

The average utility z-score of students under the mixed treatment is 0.28, which is 0.86 standard deviations higher than the baseline intuitive group and is significant at the 99% level. The formalist treatment, on the other hand, averages 0.06, which is 0.58 standard deviations higher than the baseline and is significant at the 95% level. Only the intuitive group is below zero.

When the control variables of gender and cumulative grade point average are added, the level of significance for mixed and formalist treatment effects drop to the 90% confidence level, but the magnitude of both treatment effects remain above half a standard deviation.

4. CONCLUSIONS

In a randomized classroom experiment, we find that students who are taught through a mixed mathematical and intuitive approach perform better in traditional multiple-choice tests and authentic assessment tests, enjoy their classes more, and put in more effort as opposed to students who are taught through an exclusively mathematical or exclusively intuitive approach. Additionally, those under the formalist approach perform better than those under the intuitive approach. However, students given the choice prefer to explain scenarios intuitively as opposed to mathematically.

The evidence provided by our study has several practical implications in pedagogy. For economics professors at the undergraduate level, a

more balanced teaching style might lead to better student performance, in addition to better enjoyment. In DLSU, it might help to supplement the teaching of math-heavy economics subjects like econometrics or statistics with graphs and other intuitive teaching styles.

Another point of interest is that only 20 out of the 70 students pass the semi-authentic assessment portion of the test, while only 18 get the purely authentic item correct. This would indicate that even students who are capable of passing traditional tests have trouble applying this knowledge to real world scenarios, regardless of teaching style.

Considering repetition and the use of more real-world examples may help improve the achievement of authentic learning outcomes. Given DLSUs trimestral system, the terms could be too short to realistically expect students to be able to both learn and apply lessons and to expect teachers to be able to repeat each lesson sufficiently.

For future studies, we suggest conducting similar experiments over a lengthier time period to gauge the effects of teaching approaches when it comes to long-term retention of learning outcomes.

5. ACKNOWLEDGMENTS

We thank Marites M. Tiongco and Mitzie Irene P. Conchada for helpful comments at various technical workshops.

6. REFERENCES

Callahan, G. & Leeson, P. (2006). *Scientism in the Way of Science: Hope for Heterodoxy in Economics.*

Chang, H. & Aldred, J. (2014, May 11). After the crash, we need a revolution in the way we teach economics. *The Guardian.* Retrieved from http://www.theguardian.com/business/2014/may/11/after-crash-need-revolution-in-economics-teaching-chang-aldred

Chen, J., & Lin, T. (2008). Class attendance and exam performance: A randomized experiment. *Journal of Economic Education.* Retrieved from http://nccur.lib.nccu.edu.tw/bitstream/140.119/23213/1

De Dios, E. (2014) *Economics education for a future middle class.* Retrieved from http://www.econ.upd.edu.ph/perse/?p=3786

Genesee, F. H. (2009). Early Childhood Bilingualism: Perils and possibilities. *Journal of Applied Research on Learning, 2(Special Issue), Article 2,* pp.1-21.

Greene, J. (1997). A meta-analysis of the rossell and baker review of bilingual education research. *Bilingual Research Journal,* 103-122.

Inman, P. (2014, May 4). Economics students call for shakeup of the way their subject is taught. *The Guardian.* Retrieved from http://www.theguardian.com/education/2014/may/04/economics-students-overhaul-subject-teaching

Krugman, P. (1998). Two Cheers for Formalism. *The Economic Journal, 108(451),* 1829-1836.

Mueller, J. (2014). *What is Authentic Assessment?* Retrieved from http://jfmueller.faculty.noctrl.edu/toolbox/whatisit.htm

Mankiw, G. (2006, September 15). *Why Aspiring Economists Need Math.* Retrieved from http://gregmankiw.blogspot.com/2006/09/why-aspiring-economists-need-math.html

McCloskey, D. (2000). How to be Human: Though an Economist. *Ann Arbor, MI: The University of Michigan Press.*

Mearman, A., Wakely, T., Shoib, G., & Webber, D. (2011). Does Pluralism in Economics Education Make Better Educated, Happier Students? A Qualitative Analysis. *International Review of Economic Education, 10(2),* 50-62.

Newmann, F., & Wehlage, G. (1993). Five Standards of Authentic Instruction. Educational Leadership, 50(7), 8-12.

Expected Lasallian Graduate Attributes for DLSU GE Curriculum (ELGAs) (n.d.). Retrieved from http://www.dlsu.edu.ph/offices/iaa/downloads/iaa1213t1-elga.pdf