



Performance Determinants in Undergraduate Economics Classes: The Effect of Cognitive Reflection

Alexei Orlov and John Roufagalas

Abstract

Economics classes claim to develop students' "analytical" abilities and "problem-solving" skills. Assessment of such claims is rather difficult, since it requires measurement of advanced thinking processes. Aptitude tests, such as the SAT, ACT (standardised tests for college admissions in the U.S.), and GRE (a standardised test for post-graduate admissions), purport to provide measures of such processes, but their questions refer to relatively simple thinking processes. Frederick (2005) has developed the "Cognitive Reflection Test" (CRT) to measure human ability to think deeper. The test contains questions whose apparent answers are incorrect and therefore further reflection is needed to arrive at the correct answers. Many economic problems fall into such a category. Our results suggest that CRT scores are higher and have a significant positive effect on exam performance in upper-level economics classes. In these classes, students with the highest CRT score outperform, on average, students with the lowest CRT score by more than half a letter grade, everything else the same.

JEL classification: A22, C13

1. Introduction

The terms "analytical abilities" and "problem-solving skills" appear, almost universally, among the goals of undergraduate economics classes. The nature of economic problems and the methodology of economics are well suited for the development and honing of such skills and abilities. Heterogeneity, along several dimensions, among students implies that they will face varying degrees of difficulty in developing these skills and abilities. Such difficulty may explain why some students fall in love with economics, while others develop a strong distaste for the subject.

A few questions naturally arise. First, is there a relationship between students' analytical/problem-solving abilities and their performance in economics classes? If so, is this relationship economically and statistically significant even after controlling for other factors that may affect student performance? Second, do students exposed to Principles of Economics exhibit improved analytical/problem-solving skills? Third, if they do, what is the most plausible explanation for this improvement and are these improved skills associated with better performance in subsequent economics courses? To answer such questions, measures of the analytical/problem-solving predisposition or exposure and student performance, as well as performance determinants, have to be discussed and decided upon.

Frederick (2005) analyses decision-making processes and identifies a category of problems, whose statement suggests that they are simpler than they actually are. As such, the first answer that comes to mind is incorrect. Further thinking, or "cognitive reflection" as Frederick calls it, is needed to arrive at

the correct answer. Frederick has developed the “Cognitive Reflection Test” (CRT) as a way to measure a person’s ability to arrive at the correct answer by cognitive reflection. It is interesting to note that many economic problems have the characteristic that the intuitive answer that first comes to mind is usually wrong and further reflection is needed to arrive at the correct one.¹ Hence, it is our claim that the CRT tests students’ disposition/aptitude towards the kinds of analytical abilities and problem-solving skills that are associated with economics. Our working hypothesis is that students performing better in the CRT will, on average, perform better in economics classes. Just as most of the literature on this topic, our paper does not take a stand on the direction of causation between cognitive reflection and performance, as our goal is to investigate the association between the two while controlling for as many other variables for which we have collected data, in addition to the CRT, that may affect student performance.

Siegfried and Fels (1979) review the literature and enumerate the factors that may affect student performance in economics classes, especially principles. These factors are classified in four categories: Student Human Capital, Faculty Human Capital, College Environment, and Student Effort. In the subsequent three decades more research on some of these factors has been carried out. Table 1 provides a summary of Siegfried and Fels’ (1979) findings with respect to the factors affecting students’ performance. The table also summarises findings reported in the more recent extant literature regarding these as well as several additional factors. Papers cited in the fourth column contain additional references.

Table 1: Literature Review

Category	Factor	Effect on Student Performance in Siegfried and Fels (1979)	Recent Literature	
			Author(s)	Effect on Student Performance
Student Human Capital	Entrance exam scores (Verbal and Math SAT, ACT)	Strong positive effect; Verbal SAT has a stronger effect than Math SAT	Ballard and Johnson (2004)	Strong positive effect for Math SAT
	High school rank	Positive impact		
	Student maturity	No effect	Anderson, Benjamin and Fuss (1994)	Positive effect
	Socioeconomic background	No effect		
	High school economics	Inconclusive	Anderson, Benjamin and Fuss (1994)	Positive effects when students score high
	Pre-test score	Positive and significant effect on post-test scores		

¹ Here are two examples: (A). The Bond Price – Yield relation in the financial markets. The immediate answer to the question: “What happens to yields when bond prices increase?” is often: “Yields increase!”

(B) Comparative Advantage. “If a country has an absolute advantage in both tradable products over its partner, is trade still mutually advantageous?” For many students not trained in economics (and some trained!) the intuitive answer is “No”. Many other examples can be given.

	Gap closing measures	Knowing more at the beginning implies less added knowledge		
Faculty Human Capital	Years of teaching	Positive	Hoffman and Oreopoulos (2006)	No effect
	TUCE (Test of Understanding College Economics) scores by faculty	Positive		
	Graduate school grades	Positive		
College Environment	Class size	No effect	Raimondo, Esposito and Gershenson (1990), Monks and Schmidt (2010)	Negative effects on (a) essay exams, (b) future intermediate macroeconomics performance, and (c) self-reported learning outcomes
	Larger colleges	Positive		
	High school-wide SAT scores	Positive		
	Two semester series	More understanding than one course		
	Choice of textbook	No effect		
Student Effort	Study time	No effect		
	Attendance	No effect	Marburger (2001), Stanca (2006), Arulampalam, Naylor and Smith (2007)	Positive effects
	Class Loads	No effect		
	Average GPA	Positive effect	Grove, Wasserman and Grovner (2006)	Positive effect
Additional Factors	Student gender		Anderson, Benjamin and Fuss (1994), Ballard and Johnson (2004), Borg and Stranahan (2002), Robb and Robb (1999)	Females score lower
	Instructor gender		Hoffman and Oreopoulos (2007)	Small positive effect on same gender students
	Student and instructor personality type		Borg and Stranahan (2002), Borg and Shapiro (1996)	Introvert, sensing/judging students perform better

Some of the above effects are statistically significant but quantitatively small². When both statistical and quantitative significance are taken into consideration, it seems that average (college) grade point average (GPA), high school GPA or rank, verbal and mathematical aptitude, student gender and student attendance are the most significant factors among those explored in the literature.

None of the above factors is directly associated with a student's ability to analyse and solve problems. It can be argued that factors such as mathematical ability and high school GPA or rank may be proxies for these types of ability. It is our argument that the higher-level analytical skills used in economics are different from the lower-level learning skills used in high school.³ Hence, we claim that using the CRT results provides information about a student's analytical aptitude and can predict the student's performance in economics classes.

Our goal is to examine whether cognitive reflection, as measured by the CRT, can predict the student's performance in an economics class. Our results indicate that the CRT measure is not statistically significant in predicting exam performance in principles classes. Students entering upper-level classes (one of which is required for all business majors, and the other is an elective) are able to score better in the CRT. An upper-level student who answers all three CRT questions correctly is expected to perform up to approximately 8-10 points (i.e. almost a letter grade) better in the exams than a student who answers all three questions incorrectly.

The rest of the paper is organised as follows: first, we introduce the CRT and provide an explanation of the decision model that it may reflect. Second, we briefly present our sample, followed by a discussion of the model and methodology. Finally, we discuss the results and present suggestions for further research.

2. Decision-making and the Cognitive Reflection Test (CRT)

Brain scanning techniques are greatly advancing our knowledge of the areas of the brain responsible for the various types of decision-making. Cohen (2005) reviews this knowledge and convincingly claims that decisions are the result of different, possibly competing, decision systems in the human brain. Greatly simplifying Cohen's arguments, we construct a model of decision-making consistent with his claims. According to this model, there are two decision-making systems associated with two different areas of the human brain: the pre-frontal lobe (the evolutionary newer part of the brain), and the subcortical structures (the evolutionary older part that includes areas such as the striatum and the brainstem). The subcortical structures are responsible for emotional, reflexive, routine types of decisions. These decisions are made fast and sequences of such decisions can be made in a parallel fashion, which implies that their cost is very low.

The prefrontal lobe is a more complex area. It contains what we can call a decision "controller" and a decision "processor." The processor is capable of higher-power thinking, analysis and problem-solving. It can face, and potentially solve correctly, previously un-encountered problems and come up with innovative solutions and ideas. But this processor requires concentration and can only deal with one decision at a time, i.e. it operates serially. As such it implies high costs of processing.

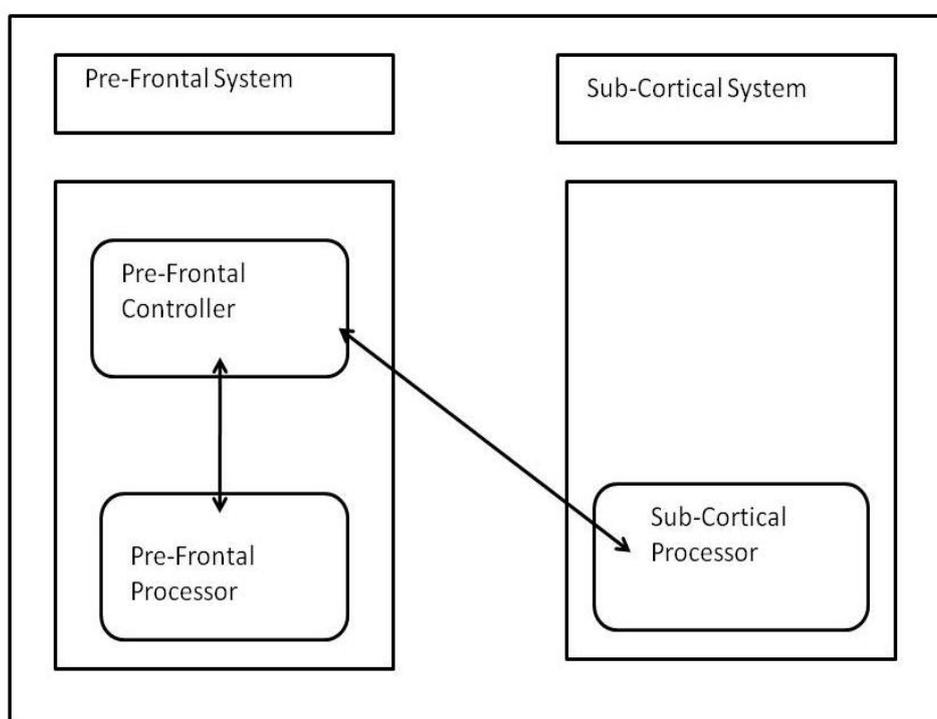
² For example, Stanca (2006) argues that a student with perfect attendance (100%) is expected to score 1.2% higher than a student with average attendance (70.8%). In the US semester system, with a class meeting three times per week for 14 weeks, each class meeting is approximately 2.38% of the overall attendance. Even if we round this number to 3% (to account for exam time, cancelled classes etc), skipping a 50-minute class is expected to cost the student 0.12% of the final grade, a small effect.

³ We have in mind something like the "proficiencies" discussed in Hansen (2001), which in a sense are parallel to the taxonomy in Bloom (1956). Low level skills involve "Accessing existing knowledge" and "Displaying command of existing knowledge," while higher level skills involve "Interpreting existing knowledge," "Interpreting and manipulating economic data," "Applying existing knowledge," and, finally, "Creating new knowledge."

The “controller” function of the pre-frontal cortex is responsible for decision allocation and evaluation: as the need for a decision arises, the controller makes an initial evaluation (based on some criteria) and allocates the actual decision either to the pre-frontal processor or to the subcortical structures. Presumably, the processor makes a rough evaluation of the expected benefits or the size of the stakes involved in the decision and decides where to allocate it. When the actual answer is received, the controller evaluates it and either announces it or remits it to the high-order processor for further processing. Figure 1 presents a graphical representation of the model.

Learning is presumably a process through which routine decisions are relegated from the high-order processor to the subcortical structures. For example, multiplication tables for a 7-year-old are most probably processed in the pre-frontal processor. For most educated adults the routine calculations involved with multiplication tables have been relegated to the subcortical structures.

Figure 1: Cohen’s Brain Decision-Making Model



Source: Authors’ adaptation of arguments in Cohen (2005).

Heterogeneity with respect to costs of processing and benefits of specific decisions imply that different individuals will allocate their decisions differently and they may come up with different answers when faced with the same problem. An example from finance may illustrate the point: consider two investors with the same portfolio decision - to buy 1000 shares of company X. The decision may be made in the subcortical structures using a simple rule like: “If the share price has increased in the last two weeks, buy; otherwise do not buy.” Or it can be made in the pre-frontal processor by collecting information about the prospects of company X, calculating financial ratios, estimating the effect of the overall economy, and so on. The investor with high cost of using the pre-frontal processor and with a controller that estimates that stakes involved (such as the probability of losing a large chunk of the investment, etc.) are relatively small will most likely allocate the decision to the cheaper subcortical structures. On the other hand, an investor who considers the cost of using the pre-frontal cortex relatively low and the stakes involved relatively high will most likely use the pre-frontal processor.

Figure 2: The Cognitive Reflection Test (CRT)

<p>1. A bat and a ball cost \$1.10 in total. The bat costs \$1.00 more than the ball. How much does the ball cost?</p> <p style="text-align: center;">_____ cents</p> <p>2. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?</p> <p style="text-align: center;">_____ minutes</p> <p>3. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?</p> <p style="text-align: center;">_____ days</p>

Source: Frederick (2005)

Frederick (2005) has identified a set of three simple questions, shown in Figure 2, that apparently the controller initially estimates can be solved by the subcortical structures.⁴ Based on learned routines, such as averages, analogies, proportions, etc., the subcortical processor provides fast, reasonable-looking answers, which are actually incorrect. In particular, for Question 1, the automatic answer is 10 cents, the result of a simple subtraction of \$1.00, the price of the bat, from \$1.10, the price of the bundle. The actual answer is \$1.05 for the bat and \$0.05 for the ball, which can be found by solving a simple system of two equations with two unknowns. For question 2, the “apparent” (and incorrect) answer comes from analogies that imply 5/5/5 should be analogous to 100/100/100, hence 100 minutes. The correct answer is 5 minutes, which is found after noting that it takes each machine 5 minutes to make 1 widget and that machines can work simultaneously. Finally, in Question 3, the automatic answer comes from proportions: if it takes 48 days to cover the entire lake, it would take 24 days (or half of 48) to cover half of the lake. The correct answer is, of course, 47 days: the patch would double in size on day 48 and thus cover the entire lake.

Presumably, the pre-frontal controller evaluates the answers. Some of the test-takers provide these incorrect automatic answers either because the controller considers them correct, or because the controller lacks the skills to evaluate them, or the controller thinks that it is too costly to evaluate them in detail. For some other respondents, the controller evaluation demonstrates that the answers are incorrect and the problems are relegated to the pre-frontal processor. These respondents will most likely come up with correct answers. This is obvious from the reaction of the respondents when the correct answers are explained to them.

As was pointed out above, many economics problems have the characteristic that routine reflexive answers are incorrect and more cognitive reflection is needed to come up with the correct answer. It is

⁴ Frederick (2005) has developed his own terminology about the brain systems that answer questions. He uses the term “System 1” for what we call subcortical structures, and the term “System 2” for pre-frontal processing.

our hypothesis that students who are more “reflective” in the manner measured by the CRT will perform better in economics classes. It should be clear from the preceding description of the CRT that to score high on the test, the respondents need to resist the urge to provide apparent and incorrect answers. Respondents can arrive at correct answers only upon further reflection, which requires a higher level of thinking. Thus the CRT intends to measure – and does measure – “the ability or disposition to resist reporting the response that first comes to mind” (Frederick, 2005, p. 35).

The CRT has been shown to be both reliable and consistent. Reliability of the CRT can be seen in Frederick’s (2005) comparison of CRT with four other measures of cognitive ability. All alternative measures correlate positively and significantly with the CRT. Frederick also documents a strong correspondence between performance on the CRT and time preference (or patience), thereby confirming the notion that people with higher cognitive abilities are more patient, i.e., have lower discount rates. Conversely, the “cognitively impulsive” group (those scoring low on the CRT) is found to be less patient. Frederick’s (2005) study of the links between the CRT and time preferences, and the CRT and risk preferences attests to the consistency of the CRT measure.

3. Data

Various measures of student performance have been proposed: Letter Grades (A-F), Percent Total Score (0-100), Percent of Correct Exam Questions (based on exam scores only), etc. The choice of measure determines the econometric method of analysis used (e.g. Ordered Probit models are used with Letter grades), but no measure has been proven superior. Hence, the choice of measure hinges on data availability, though percent total score or percent of correct exam questions contain more detailed information. As argued below, our data allow us to use a more detailed measure.

We administered the 3-question CRT in all our Fall 2006 and Spring 2007 classes during the first week. The classes included: ECON 105: Principles of Macroeconomics (4 sections), ECON 330: Money and Banking (2 sections), and ECON 340: Global Economy (2 sections). ECON 105 and 340 are required for all business students and are General Education options for the rest of the university. They have no prerequisites, but students typically take ECON 340 after the Principles sequence, and in the sequence they take Principles of Macroeconomics before Principles of Microeconomics. ECON 330, an intermediate-level course, has ECON 105 as its prerequisite, is required for all economics majors, and attracts very few non-economics majors.

Our principles course is taught as an average, run-of-the-mill course with the use of McConnell and Brue’s textbook. Money and Banking is also a standard course that uses Mishkin’s textbook. All sections of the Global Economy and Business course are taught using Hill’s “Global Business Today”. The exam questions in all of our classes are a mixture of analytical and non-analytical (definitional/term recognition) questions, many of which come directly from the textbooks’ test banks.

In ECON 105 and 330, attendance is required and monitored via the use of a “clicker” system. Each student is required to purchase a “clicker”, which, via a remote sensor communicates with the class computer. At the beginning of each lecture, a 3-7 question quiz on past material was administered. Performance on the quizzes and attendance accounted for 20% of the final grade in the ECON 105 class, and 10% in ECON 330. The variable ATTEND describes attendance frequency.

At the beginning of each semester the registrar routinely provides faculty with class rosters that contain information about the students’ majors. Based on these rosters, we assigned students into colleges. These Colleges are: Business and Economics (COBE), Information Systems and Technology (CIST), Visual

and Performing Arts (CVPA), Arts and Sciences (CAS), Education (EDUC), and Human and Health Sciences (CHHS). Students who did not declare a major were classified as Pre-major.

We were also provided data from the registrar's office that included (a) the gender of the student; (b) SAT⁵ scores, verbal (SAT_V), maths (SAT_M) and total (SAT); (c) high school or transfer grade point average (HS_GPA); (d) whether the student transferred into the university; (e) the number of credits completed when the student enrolled in the class (CREDITS) and (f) the student's total GPA at the end of the semester (GPA). Note that some of the data were missing, which affects the size of the sample in some of the regressions below.

Our choice of control variables is motivated in no small part by the extant literature. Differences in student performance between male and female students have been documented by Anderson *et al.* (1994), Robb and Robb (1999), and Borg and Stranahan (2002), among others. Siegfried and Fels (1979) emphasise the importance of the initial level of human capital – as measured, for example, by verbal, maths, and total SAT scores – when trying to account for student performance. In addition to the SAT scores, we also use high school or transfer GPA as a proxy for the initial level of the students' analytical abilities. Previous studies suggested that attendance may affect student performance (e.g. Marburger, 2001; Stanca, 2006; Arulampalam *et al.*, 2007). Each student's total GPA at the end of the semester is a measure of student effort; the use of this variable has been advocated by Siegfried and Fels (1979) and Grove *et al.* (2006). Finally, we believe that students' college experience, which we measure by the number of credits completed, should also belong in the set of control variables when trying to account for student performance.

Student gender was described by a dummy variable (GENDER), which took the value of 1 if the student was female. Similarly, dummy variable TRANSFER takes the value of 1 if the student transferred into the university. Dummy variables ECON 330 and ECON 340 were introduced to capture the difference of these courses from the excluded category: ECON 105: Principles of Macro. With these variables we intend to capture the effects of (a) different instructors; (b) different attendance policies; and (c) different level of instruction. One co-author has taught all sections of the Principles and Money and Banking classes, while the other co-author taught all Global Economy classes. Thus, the dummy variables for ECON 330 and ECON 340 represent different instructors as well. In all sections of each course we used the same textbook and the same exams.

Our main focus is the explanatory variable called "CRT Score" which takes values (0, 1, 2, 3) to signify the number of correct answers. We use exam averages (i.e. percent correct answers averaged over all exams, per student), as the dependent variable. Our choice of dependent variable is motivated by the fact that final grades across different classes are determined differently. For example, attendance and open-notes (but limited-time) quizzes affect principles final grade, while presentations and in-class exercises affect ECON 340 grades. We want to preclude items unrelated to analytical ability, such as attendance or public speaking skills, from affecting our results.

Table 2 presents information about our sample and shows that about 40% of the sample are female and the majority of our students, especially in ECON 340 and ECON 330, are College of Business and Economics students.

⁵ The SAT is a standardised test used in the United States for college admissions.

Table 2: Sample Information

	ECON 105	ECON 340	ECON 330	TOTAL
Sample	199	67	46	312
Gender				
Females	83	30	14	127
Males	116	37	32	185
Colleges				
COBE	92	59	41	192
CIST	5	3	0	8
CVPA	19	1	0	20
CAS	44	4	4	52
CHHS	5	0	0	5
EDUC	11	0	1	12
Pre-major	23	0	0	23

Source: Authors' calculations

Note: Table 2 presents information about the size of the sample and the distribution of students with respect to classes, gender, and college.

Table 3 shows enrollment data for the various classes included in the sample. It can be seen that initial enrollment ranged from a low of 22 students to a high of 95 students, i.e. we have a mixture of small and larger classes. The second column shows the distribution of the 312 observations used in the sample. The sample excludes the students who did not have a CRT score (i.e. they were absent when the test was administered, column 3) and/or did not complete the course (i.e. did not take all the exams and did not have a final grade, column 4).

It also shows how many of the students who did not complete the course had taken the CRT test (column 5), as well as the number of students who failed the course (column 5). The failing students are part of the sample. Note that addition of the columns "Used CRT Responses", "Dropped; CRT Taken" and "No CRT Taken" yields the initial enrollment.

As can be seen, only 15 students, of a total of 379, failed to complete the courses, i.e. dropped out, and another 14 received a failing grade. Attrition does not seem to be a significant problem in our sample.

Table 3: Enrollment Data

Term/Class	Initial Enrollment	Used CRT Responses	No CRT Taken	Dropped Total	Dropped; CRT Taken	Failed
Fall 2006						
ECON 105–09	40	40	0	0	0	0
ECON 105–10	95	76	16	4	3	6
ECON 330–01	22	21	1	0	0	1
ECON 340–01	58	39	17	6	2	2
ECON 340–05	37	28	9	2	0	2
Spring 2007						
ECON 105–08	39	31	8	0	0	0
ECON 105–11	57	52	3	3	2	3
ECON 330–01	31	25	6	0	0	0
TOTAL	379	312	60	15	7	14

We split the overall sample into two sub-samples: one includes only the principles classes (Principles Level), while the other includes the students in the ECON 330 and 340 classes (Upper Level).

Table 4: Descriptive Statistics

	Total Sample			Principles Level			Upper Level		
	Obs	Average	Std Deviation	Obs	Average	Std Deviation	Obs	Average	Std Deviation
Exam Avg.	312	77.03	10.13	199	76.96	9.91	113	77.13	10.57
CRT Score	312	0.67	0.91	199	0.50	0.79	113	0.97	1.04
SAT_V	238	495.37	70.88	161	497.57	69.92	77	490.77	73.08
SAT_M	238	510.33	67.26	161	508.88	66.14	77	513.37	69.87
SAT	238	1005.71	122.95	161	1006.46	119.35	77	1004.16	130.94
HS_GPA	304	3.00	0.42	197	3.02	0.39	107	2.96	0.47
CREDITS	311	45.93	31.36	198	29.58	25.17	113	74.59	17.52
ATTEND	245	0.82	0.17	199	0.83	0.16	46	0.77	0.19
TRANSFER	312	0.12	0.33	199	0.10	0.30	113	0.15	0.36
EXECPGA	312	2.89	0.61	198	2.86	0.66	113	2.96	0.50

Source: Authors' calculations.

Note: The table gives the number of observations, the mean and standard deviation of the various variables in the 3 samples.

Table 4 presents descriptive statistics for the variables in the three samples.⁶ In all samples the average exam grade was about 77 points. SAT scores are very similar, with verbal in the 490s and maths at around 510. High school or transfer GPAs are also similar, averaging around 3.0. As expected, students in the upper-level classes have more credits, about 75; while the average principles student has only 30

⁶ It should be noted that our paper does not include all of the variables that the past research on student performance has suggested over the past three decades (summarised in Table 1), and thus there is still a possibility of an omitted variable bias. We do not include some of the variables in our study for two (often overlapping) reasons: (i) some of the variables were deemed unimportant by the extant literature in accounting for student performance, and (ii) we do not have data for some of the variables. All in all, however, we believe that our study contains most of the important control variables.

credits (and is barely a sophomore⁷) before (s)he takes the Principles class. The attendance frequency for principles classes was about 83%, which reflects the fact that attendance was mandatory. Finally, about 10% to 15% of the students are transfers. Variable EXECGPA represents the EXcluding EConomics GPA, i.e. the student's GPA at the end of the semester in which (s)he enrolled in the respective economics class, excluding the grade (s)he received in the economics class. The value of EXECGPA remains roughly constant among the three samples at about 2.9.

Of particular interest is the average CRT score. Overall, the average student answered about 2/3 of questions out of possible 3. Sample decomposition shows that principles students answer about half a question correctly (0.50), while upper-level students are able to improve their score by almost 100% and answer about 1 question correctly (0.97).

4. Model and methodology

The goal of the analysis is to determine whether student performance in the 3-question CRT score has any power in predicting student exam performance, above and beyond the usual variables used in the literature. The model to be used is:

$$\text{Exam Average} = f(\text{CRT score, control variables})$$

Control variables include SAT scores, high school GPA, whether the student transferred, the gender of the student, and the number of college credits the student has accumulated.

Table 5 presents the simple correlation coefficients among the various variables in our model.

Table 5: Pearson Correlation Coefficients

	Exam Average	CRT Score	Gender	HS GPA	EXECGPA	SAT_M	SAT_V	Credits	EC 330	EC 340
Exam Average	1.0000	0.2519	0.0658	0.2508	0.4288	0.3130	0.4020	0.1928	0.0023	0.0434
CRT Score		1.0000	-0.1877	0.1020	0.1423	0.4399	0.2992	0.2141	0.1115	0.1387
Gender			1.0000	0.0072	0.1270	0.2547	0.1154	0.0199	-0.0039	-0.0794
HS GPA				1.0000	0.3301	0.1935	0.2670	-0.0159	0.0262	0.0283
EXECGPA					1.0000	0.2187	0.1862	0.0735	0.1059	-0.0106
SAT_M						1.0000	0.5936	0.0353	0.0968	-0.0557
SAT_V							1.0000	0.0258	0.0187	-0.0623
Credits								1.0000	0.6205	0.3013
EC 330									1.0000	-0.1897
EC 340										1.0000

Top number: Correlation Coefficient

Bottom Number: p -value for $H_0: \beta=0$

⁷ In the second year of university.

As expected, the exam average is significantly correlated with (in diminishing order) the EXECGPA ($r=42.88\%$), the Verbal SAT ($r=40.2\%$), the Maths SAT ($r=31.3\%$), the CRT Score ($r=25.2\%$), high school GPA ($r=25.1\%$), and the number of credits ($r=19.3\%$). Gender and upper-level class dummies are not significantly correlated with the Exam average. It is also interesting to note the CRT scores are significantly correlated with both SAT scores, the EXECGPA and the number of credits (i.e. the college experience) of the student. Note that the correlation coefficient between CRT scores and Maths SAT scores, while statistically significant, is less than 50%, suggesting that the two measures do not reflect exactly the same abilities.

5. Results

Tables 6, 7 and 8 present our regression results. Table 4 above shows that, as expected, students in upper-level classes have accumulated more credits than students in the principles classes; hence collinearity is expected between the variable CREDITS and the ECON 330 and ECON 340 dummies. Indeed, Table 5 shows a simple correlation coefficient of about 62% between CREDITS and ECON 330. Similarly, the correlation between the verbal and maths SAT scores is about 59%. In order to avoid collinearity problems, we use regressions with either total SAT scores or maths SAT scores, and either number of credits or dummies for the upper-level classes.

Notice that college GPA is not part of our control variables set, as inclusion of this variable presents considerable problems. First, there is a conceptual problem with the interpretation of its coefficient. If it is positive and significant it leads to the conclusion that “good students do well in economics and below average students do badly in economics.” It is highly likely that this statement holds for every other class as well and it does not explain why a student does well. Second, GPA is a composite of individual class grades that are in part determined by a similar set of variables, such as SAT scores, high school GPA, gender etc. Hence, the GPA reflects how these variables affect performance in the average class included. As a result, the effects of the other control variables are obscured because they affect the dependent variable both indirectly, through the GPA, and directly. Putting it in another way, inclusion of the EXECGPA introduces the so-called “dominant variable” effect (Rao and Miller, 1971, pp. 41-43). Rao and Miller argue in favour of excluding the dominant variable to better understand the underlying relationship.⁸

Given our control variables, Table 6 shows that the CRT score has a marginally significant effect in the overall sample (significant at the 90% level in two regressions and insignificant in the other two). It is worth reminding the reader that it is very difficult to capture the effects of critical thinking on the basis of a rough measure like a standardised test, especially such a simple, three-item test as the CRT. Consequently, this type of measured significance that we find may be all that one can expect, and it is therefore meaningful that any statistically significant correlation was obtained.

Inspection of Tables 7 and 8 reveals that the CRT score has no significant effect on student performance in the Principles level classes, but a strong positive effect on the upper-level classes. The effect is also large quantitatively: for each CRT question an upper-level student answers correctly, he or she is expected to obtain, on average, a 2.5 to 3 percentage points higher exam grade overall. In other words, a student who answers all 3 CRT questions correctly is expected to achieve a grade higher by almost one whole letter grade. This effect is so strong that it carries over to the whole sample, although with lower significance and about two-thirds of the size.

The difference between the principles and upper-level cohorts might arise for a number of reasons. Upper-level courses are designed in ways that require problem-solving ability, especially ECON 330,

⁸ Inclusion of college GPA measures in ECON performance equations nearly doubles their R -square levels, strongly suggesting a dominant variable effect.

while lower-level courses often do not require this ability. Such a distinction could manifest in the differences in the assessment format: exams in upper-level courses are more likely to emphasise and reward problem-solving abilities. (In our case this explanation is less plausible than the others in light of the fact that exam questions in all classes are a mixture of analytical and definitional questions). It is also likely that students who self-select into the economics major and upper-level courses have better problem solving skills (although it must be pointed out that not every student in upper-level economics courses majors in economics). Finally, students' problem-solving abilities may improve as a result of their learning experience during a principles course.

The results reported in Tables 7 and 8 are based on splitting the sample into the lower-level (or principles) sub-sample and the upper-level sub-sample. Since ECON 340 has no prerequisites (just as principles has none), one may wonder to what extent this course is truly "upper-level". Consequently, as a robustness check, we rerun the four regressions reported in Table 8 excluding the ECON 340 students from the upper-level sub-sample. Without the ECON 340 observations in the sub-sample the CRT effect becomes even stronger.⁹ Including only ECON 330 responses almost doubles the magnitude of the CRT effect – from about 3 to more than 5 points. *F*-values are now lower but still significant. Excluding the ECON 340 observations reduces the SAT coefficients by one order of magnitude (i.e. about 10 times) and eliminates their statistical significance. Thus, excluding ECON 340 from the upper-level group makes our results even stronger, and therefore the results reported in Table 8 constitute conservative estimates of the CRT effect.

The tables show a number of other interesting results, some in agreement with the literature and some in contrast. In Table 6 we can see that gender, SAT scores, high school or transfer GPA, and number of credits have positive and significant effects on student performance. Comparison with tables 7 and 8 reveals that most of the statistical significance carries over from the Principles sub-sample. With the exception of the SAT scores, the other three variables are not significant in the upper-level sub-sample.

In contrast with the previous research (e.g. Anderson *et al.*, 1994; Borg and Stranahan, 2002; Ballard and Johnson, 2004), we obtain that female students are expected to do about 3% better than their male colleagues. This might be due to the fact that presently there are more women in academe than men and, as a consequence, more high-quality female students who are likely to self-select into higher-paying majors such as economics. The effects of gender are significant in the overall sample, but mostly statistically insignificant in the two sub-samples, though their size is of the same order.

SAT scores have strong effects on performance, which is in line with past research (e.g. Ballard and Johnson, 2004). In principles classes, a student entering with a Maths SAT score of 600, is expected to achieve a grade about 6 points higher than a student entering with a Maths SAT of 400 (i.e. the difference of 200 points times the coefficient 0.03). The effect of Maths SATs increases quantitatively in upper-level classes as Table 8 indicates.

The effects of the high school or transfer GPA are strong, statistically and quantitatively, in the principles sub-sample and carry over in the total sample, but are insignificant and smaller in the upper-level sub-sample. This is to be expected, as students with stronger backgrounds are likely to perform better in college, and this starting-point effect should become less pronounced as students advance in their college careers. In the principles sample, each additional HSGPA (high school GPA) point is expected to add about 5 points (i.e. half a letter grade) on the student performance measure. It should be noted that this effect is reduced to 3 points when attendance is taken into consideration.

Unlike the existing literature, we obtain a strong, statistically significant effect of attendance. With the caveats that attendance was required and we have complete data only for the principles classes, and

⁹ The results of this robustness check are not reported in the paper to conserve space. We thank one of the referees for suggesting this sensitivity analysis.

hence only the principles sub-sample estimates are reliable, we obtain that complete attendance (i.e. 100% or 1) improves performance by 0.9 points over 90% attendance or 9 points over 0% attendance!

Student college experience, as can be surmised by the number of credits the student had completed before entering the examined courses, also has strong positive effects that are mostly concentrated in the principles sub-sample. This is not surprising, as the marginal impact on student performance of one additional credit-hour earned by a freshman¹⁰ is likely to be higher, on average, than that earned by a junior or a senior¹¹.

Finally, whether a student has transferred into the university seems to have a negative effect on performance, which is rather large quantitatively but statistically insignificant. The regressions explain 18% to 28% of the performance measure variation (as determined by the Adjusted *R*-squared values) and the *F*-tests of all regressions are highly statistically significant. Regression diagnostics revealed no problems with heteroskedasticity or multicollinearity.

Table 6: Regression results (Overall Sample)

Dependent Variable	Exam Average	Exam Average	Exam Average	Exam Average
Constant	34.79***[5.56]	36.84***[5.84]	41.21***[6.45]	33.39***[4.50]
CRT Score	1.05 [1.39]	1.40* [1.83]	1.36* [1.75]	1.31 [1.49]
GENDER	3.04***[2.83]	3.33***[2.83]	3.01** [2.5]	2.50* [1.87]
SAT	0.026***[4.91]	0.026***[4.84]		
SAT_M			0.03***[3.38]	0.03***[2.95]
HS_GPA	3.73** [2.53]	3.52** [2.35]	4.56***[3.05]	4.43***[2.63]
TRANSFER	-2.39 [-1.21]	-2.51 [-1.24]	-3.34* [-1.65]	-3.30 [-1.40]
CREDITS	0.05***[2.74]		0.05***[2.63]	0.06***[2.74]
ECON_330		-0.58 [-.36]		
ECON_340		1.66 [1.04]		
ATTEND				10.36***[2.58]
Obs	236	236	236	197
Adj. R-squared	0.23	0.20	0.19	0.21
F-Value	12.66***	9.69***	10.11***	8.52***

Source: Authors' calculations.

***, **, * signify 1%, 5% and 10% levels, respectively.

Numbers in brackets are *t*-test values.

¹⁰ Student in the first year of university.

¹¹ Students in third and fourth years of university, respectively.

Table 7: Regression Results (Principles Level Sub-sample)

Dependent Variable	Exam Average	Exam Average	Exam Average
Constant	27.65***[3.81]	35.15***[4.61]	28.91***[3.58]
CRT Score	0.51 [0.58]	0.78 [0.85]	0.54 [0.59]
GENDER	2.37* [1.56]	1.97 [1.40]	2.05 [1.47]
SAT	0.027***[4.58]		
SAT_M		0.032***[2.88]	0.036***[3.22]
HS_GPA	5.56*** [3.22]	6.68*** [3.79]	5.54***[3.04]
TRANSFER	-3.40 [-1.44]	-4.12* [-1.69]	-3.90 [-1.61]
CREDITS	0.11***[3.85]	0.11***[3.92]	0.12***[4.23]
ATTEND			8.92** [2.13]
Obs	161	161	161
Adj. R-squared	0.28	0.22	0.24
F-Value	11.33***	8.64***	8.22***

Source: Authors' calculations.

***, **, * signify 1%, 5% and 10% levels, respectively.

Numbers in brackets are *t*-test values.

Table 8: Regression Results (Upper Level Sub-sample)

Dependent Variable	Exam Average	Exam Average	Exam Average	Exam Average
Constant	46.89***[3.68]	49.92***[4.15]	47.39***[3.85]	49.40***[4.24]
CRT Score	2.91** [1.94]	2.80* [1.85]	2.82* [1.87]	2.60* [1.72]
GENDER	3.31 [1.43]	3.64 [1.68]	3.50 [1.52]	3.84* [1.68]
SAT	0.020 [1.69]	0.022* [1.92]		
SAT_M			0.039* [1.78]	0.046** [2.10]
HS_GPA	0.85 [0.30]	0.99 [0.17]	0.81 [0.29]	0.41 [0.15]
TRANSFER	0.27 [0.07]	-0.26 [-0.07]	0.05 [0.01]	-0.41 [-0.11]
CREDITS	0.05 [0.79]		0.04 [0.69]	
ECON_330		-1.92 [-0.86]		-2.22 [-0.99]
Obs	75	75	75	75
Adj. R-squared	0.18	0.18	0.18	0.19
F-Value	3.71***	3.73***	3.78***	3.88***

Source: Authors' calculations.

***, **, * signify 1%, 5% and 10% levels, respectively.

Numbers in brackets are *t*-test values.

6. Conclusion

This paper proposes to use Frederick's (2005) CRT in accounting for student performance. Our results indicate that the test does indeed have predictive power in the context of student performance, while carrying very low costs of implementation. Frederick (2005) points out that "the CRT is an attractive test: it involves only three items and can be administered in a minute or two, yet its predictive validity equals or exceeds other cognitive tests that involve up to 215 items and take up to 3½ hours to complete (or which involve self-reports that cannot be readily verified)" (p.37).

The results reported in this paper are quite interesting. Higher levels of cognitive reflection help students perform considerably better in upper-level economics classes, though not that much better in the principles courses. It could be that upper-level economics classes require more cognitive reflection than principles classes where a student can do well, sometimes, by using lower-level skills such as memorisation. Alternatively, it may take advanced training in economics for the critical thinking ability to manifest itself. This makes sense especially to the extent that lower-level courses emphasise learning of institutional facts and features of the economy, while upper-level courses emphasise application of economic theory, which arguably is where the correlation with critical thinking occurs. These interesting questions are left for future research; what we hope to have accomplished is to lay a foundation for others to build upon.

Admittedly, most of the literature on student performance in economics classes focuses on principles students. Our study, in its principles sub-sample, verifies most (but not all) of the literature results. SAT scores, high school (or transfer) GPA, attendance, and student college experience have significant positive effects on principles class performance. Unlike most of the literature, which finds that male students achieve up to 5% higher grades in economics as compared to females, we find that females gain 3% more than males in our complete sample. This may reflect changes in student body composition in the last twenty years.

The coefficient on the CRT scores in the regressions for the principles courses is insignificant because, presumably, cognitive skills are not essential to succeeding in those courses. However, those cognitive skills play an increasingly important role as the student advances in his or her college career, and so the CRT scores become significant in the upper-level regressions. This is a novel result relative to the extant literature.

Our review of the extant literature reveals that over the years researchers suggested numerous explanatory variables that may help account for student performance. Needless to say, it is very difficult to uncover new variables that would have a significant effect on student performance after controlling for the "usual suspects". It is therefore quite remarkable that the majority of the regressions reported in this paper produce significant results for the CRT variable at the 90% level or better. Our results seem particularly impressive in the face of our expanded set of control variables and in light of the strong correlations between our dependent variable and the controls.

Two caveats about our results should be mentioned. First, there may be an issue with self-selection in our data. It can be argued that mostly students who do well in principles classes continue on to upper-level classes. Our counterargument is that one of our upper-level classes is a required course for all business students (ECON 340), which means that business students cannot self-select. Also, the regression results show that there is no difference between the ECON 340 and the ECON 330 students (the dummy variable that differentiates them has no statistical significance). We admit that non-business students do not usually enrol in upper-level economics classes and that may still be a source of

self-selection, but we have no way of testing this and we suspect that the effect will probably be very small.

Second, another significant issue is that of causality that was raised against much of the literature by Siegfried and Fels (1979). The issue is this: does cognitive reflection cause course performance, or course performance cause cognitive reflection, or are both performance and cognitive reflection caused by a third variable, such as student effort, which is largely unobserved? While we are aware of this problem, we do not have the data detail needed to deal with it. Hence our results, as almost all other results in this strand of literature, should be interpreted with this caveat in mind.

Future research can build on the analysis presented in this paper in at least two ways. Economics is usually believed to help students master cognitive reflection skills. If this is so, will students exposed to Principles of Economics perform better on the CRT relative to those who didn't have introductory economics? To test this hypothesis one can compare random samples of students who took introductory economics with those who did not. Related to this is the question of whether the students who were successful in acquiring cognitive reflection skills in the principles classes will do well in upper-level economics courses. Students' success at acquiring cognitive reflection skills can be measured by the CRT, and testing the following hypothesis can provide the answer: Students exposed to Principles of Economics who score better in the CRT will perform better in subsequent economics classes.

Taking a somewhat broader perspective, the CRT analysis can help assess if students who are good at cognitive reflection tend to gravitate toward economics and similar "deep-thinking" disciplines. Cross-section variation in CRT scores can be used to explain the students' initial major choice, as well as changes in major during their college career. The CRT may be a predictor of students' success in other social sciences, engineering, humanities, business, etc. If the CRT is found to have such predictive powers, colleges and universities could use the CRT (along with other assessment tools) to help students identify fields of study especially suited to their abilities and interests.

References

- Anderson, G., Benjamin, D. and Fuss, M. A. (1994). 'The Determinants of Success in University Introductory Economics Courses', *Journal of Economic Education*, Vol. 25(2), pp. 99–119.
- Arulampalam, W., Naylor, R. A. and Smith, J. (2007). 'Am I Missing Something? The Effects of Absence from Class on Student Performance', Working Paper, University of Warwick, Department of Economics, The Warwick Economics Research Paper Series.
- Ballard, C. L., and Johnson, M. F. (2004). 'Basic Math Skills and Performance in an Introductory Economics Class', *Journal of Economic Education*, Vol. 35(1), pp. 3–23.
- Bloom, B. S. (1956). *Taxonomy of Educational Objectives, Handbook I: Cognitive Domain*, New York, NY: David McKay.
- Borg, M. O., and Shapiro, S. L., (1996). 'Personality Type and Student Performance in Principles of Economics', *Journal of Economic Education*, Vol. 27(1), pp. 3–25.
- Borg, M. O. and Stranahan, H. A. (2002). 'Personality Type and Student Performance in Upper-Level Economics Courses: The Importance of Race and Gender', *Journal of Economic Education*, Vol. 33(1), pp. 3–14.

- Cohen, J. D. (2005). 'The Vulcanization of the Human Brain: A Neural Perspective on Interactions Between Cognition and Emotion', *Journal of Economic Perspectives*, Vol. 19(4), pp. 3–24.
- Frederick, S. (2005). 'Cognitive Reflection and Decision Making', *Journal of Economic Perspectives*, Vol. 19(4), pp. 25–42.
- Grove, W. A., Wasserman, T. and Grodner, A. (2006). 'Choosing a Proxy for Academic Aptitude', *Journal of Economic Education*, Vol. 37(2), pp. 131–47.
- Hansen, W. L. (2001). 'Expected Proficiencies for Undergraduate Economics Majors', *Journal of Economic Education*, Vol. 32(3), pp. 231–42.
- Hoffmann, F. and Oreopoulos, P. (2006). 'Professor Qualities and Student Achievement', NBER Working Paper No. 12596.
- Hoffmann, F. and Oreopoulos, P. (2007). 'A Professor Like Me: The Influence of Instructor Gender on College Achievement', NBER Working Paper No. 13182.
- Marburger, D. R. (2001). 'Absenteeism and Undergraduate Exam Performance', *Journal of Economic Education*, Vol. 32(2), pp. 99–109.
- Monks, J. and Schmidt, R. (2010). 'The Impact of Class Size and Number of Students on Outcomes in Higher Education', Cornell University, School of Industrial and Labor Relations Working Paper No. 114.
- Raimondo, H. J., Esposito, L. and Gershensberg, I. (1990). 'Introductory Class Size and Student Performance in Intermediate Theory Courses', *Journal of Economic Education*, Vol. 21(4), pp. 369–81.
- Rao, P. and Miller, R. L. (1971). *Applied Econometrics*, Belmont: Wadsworth.
- Robb, R. E. and Robb, A. L. (1999). 'Gender and the Study of Economics: The Role of Gender of the Instructor', *Journal of Economic Education*, Vol. 30(1), pp. 3–19.
- Siegfried, J. J. and Fels, R. (1979). 'Research on Teaching College Economics: A Survey', *Journal of Economic Literature*, Vol. 17(3), pp. 923–69.
- Stanca, L. (2006). 'The Effects of Attendance on Academic Performance: Panel Data Evidence for Introductory Microeconomics', *Journal of Economic Education*, Vol. 37(3), pp. 251–66.

Contact details

Alexei G. Orlov
Department of Economics,
Radford University,
Radford, VA 24142,
United States
Email: aorlov@radford.edu

John Roufagalas
Department of Economics,
Radford University,
Radford, VA 24142,
United States
Email: