# Students' Time-Allocation, Attitudes and Performance on Multiple-Choice Tests

Vladimir Hlasny*

**DRAFT – Please contact the author before quoting**

## Abstract

This study addresses students' executive functioning and time allocation while taking Principles of Economics multiple-choice exams, their attitudes toward these exams, and their performance on them and in the course overall. Students' decision how much time to spend on an exam and whether to review answers before submitting the exam is linked to students' attitudes toward multiple-choice exams and toward grades, students' background, their score on the exam, and their eventual performance in the course. I find that time allocation varies significantly across cultural settings, students' demographic and academic backgrounds, and students' skills and values. Most results validate a hypothesis that students' expected marginal productivity on exam questions affects their effort positively, and falls as time goes by.

Students in Korea forsake significantly less of their exam time than students in the U.S. Men, Black students, and Asian students forsake less time than other groups, possibly for cultural reasons. Quantitative-science students forsake less time than students of applied science or humanities, suggesting that one's familiarity with the material improves one's productivity and motivation to exert marginal effort on exams. Similarly, students who scored well on a previous exam tend to spend more time on the following exam – likely a self-selection by student marginal productivity on exams. Also, controlling for other student characteristics, upper-classmen are shown to forsake less time than lower-classmen.

Regarding students' personalities I find that students who reportedly typically check their answers before submitting them indeed take more time to finish exams. Students who attribute more importance to grades also spend more time on exams. However, there is no indication that students who reportedly usually finish their exams on time submit their answer sheets earlier.

Keywords: Test-taking, time management, executive functioning, multiple choice tests, Principles of Economics
JEL Codes: A22, I24

***** Associate professor, Economics, Ewha Womans University, 401 Ewha-Posco Building, Seoul 120-750, Korea. Office: +82-2-3277-4565.

## Introduction

Standardized multiple-choice tests have become an important part of the international movement to make schools accountable for their usage of resources and students' learning achievements. This came to the forefront under George W. Bush's administration following the enactment of the 2001 No Child Left Behind Act (Wagner, 2010). In Economics, particularly in the Principles courses that millions of college students globally take every year, multiple-choice exams have become a staple, popular with instructors and really deemed as necessary. The same test banks are used in the same way around the world.

Omnipresence of multiple-choice tests and their ease of content analysis have attracted some scrutiny in economics of education. Existing research has evaluated some features of exams across different settings, and across demographic groups among students (Brown 1991). Student learning has been compared across different content types of exams, and between class-room, online and hybrid courses (Brown & Liedholm 2002). These studies have found significant differences in performance across students from different backgrounds, and across exam settings. Qualitative studies have identified differences in students' concept comprehension and information retention across different exam formats. Behavioral explanations of these differences are, however, missing. Students' test-taking behavior, and time allocation in particular, have not been explored formally using quantitative methods. Features of examinations beside basic question-content and question-format have been omitted.

This study investigates the role of cultural setting, demographic and academic background, and student skills and values, in explaining students' test-taking behavior. It uses classroom data on students' exam-taking practices, survey data on students' attitudes toward multiple-choice tests administered at the beginning of the semester, and records of students' prior GPA, performance on prior exams in the course, and eventual grade on the exam and in the course overall. I study how much time students take to complete an exam, whether they review their answers, how their timing on an exam relates to their self-reported behavior on other exams, their attitude toward exams and grades, and their eventual performance in the course. This study attempts to identify which students cope well with the multiple-choice exam format, and suggest ways to bring other groups of students aboard. As a byproduct, this study allows educators to predict which students will succeed in problem-solving and learning later in their studies of economics from observing their time-management on a midterm or final examination (Bettinger *et al.* 2013).

## Methods and Data

This study uses classroom data on students' exam-taking practices, survey data on students' attitudes toward multiple-choice tests administered at the beginning of the semester, and records of students' prior GPA, performance on prior exams in the course, and eventual grade on the exam and in the course overall. 631 observations of exam-taking behavior were collected for 468 students over nine semesters at two large universities through in-classroom observation. Six semesters worth of data, summer 2003 through summer 2005, were collected in Principles of Microeconomics courses offered by two different instructors at Michigan

State University (East Lansing, MI, USA). Approximately one half of these observations were from online courses that met offline only three times, to administer two midterm exams and one final. The other half of observations are for standard classroom courses. Three semesters of data, spring 2008 through spring 2009, were collected in Principles of Economics courses offered by the author in standard classroom setting at Ewha Womans University (Seoul, Korea). The fact that the sample comprises heterogeneous observations from two countries, two instructors, and online versus offline settings represents an empirical challenge but also an opportunity to evaluate the various conditions under which students take exams.

In addition to amassing classroom observations, students' preferences, attitudes, typical exam-taking behavior, and a priori expectations were surveyed during the fall 2008 and spring 2009 semesters at Ewha Womans University (110 observations for 59 students). Table 1 reports the information collected, variable description and summary statistics. In addition to the variables reported in Table 1, information about honor-student status, exchange student status, native language, and current grade point average was collected for a subset of observations. These additional variables were evaluated using simple correlation analyses, but were not included in final regression models due to poor availability and limited theoretical or empirical significance.

Theoretically, students' decision how much time to spend on an exam depends on the marginal cost of students' presence in the classroom, and marginal productivity of students' effort on the exam in terms of grade reported on academic transcript and a measure of warm glow from following classroom norms. Students' cost is in terms of forgone earnings or disutility from spending time in the classroom, and may depend on students' demographics. Marginal productivity depends on student's academic preparation, other skills (attention span and ability to review and improve answers), and possibility to retake the course later. Finally, students' value from following classroom norms may depend on the cultural setting and student demographics. Behaviorally, student values, expectations and attitudes toward multiple-choice exam format may affect their behavior. Because many of these variables are unobservable, students' self-reported up-to-date performance (grade point average, and behavior on other exams) and attitudes toward grades and exams are surveyed.

Empirically, students' time to completion of an exam is thought to depend on students' semester-invariant characteristics (gender, age, class level, major, GPA, experience with multiple-choice exams in previous classes, attitudes toward grades and exams); students' experience in the class (expected grade, grades on previous assignments); test characteristics

(count and type of questions, time allowed, average grade, variation in grades); and general characteristics of the class (spring/summer/fall semester, class size, country).

Table 1. Definition of Variables Used in Statistical Analysis

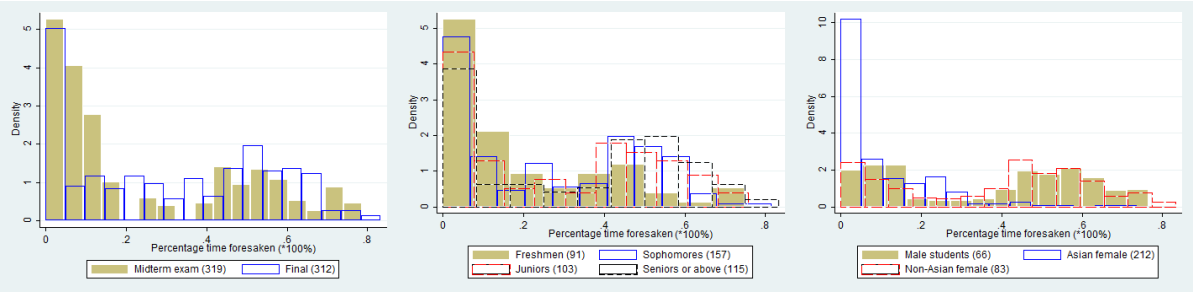
| Name | Definition [Units] | Observations | Avg. (St.Dev.)[i] | Min. – Max |
|---|---|---|---|---|
| **Information observed in classroom** | | | | |
| Time | Time taken by student on an exam [minutes] | 631 | 61.73 (22.07) | 18 – 120 |
| Timewaste | % difference between one's minutes and max minutes by any student on the same exam, (maxminutes-minutes)/maxminutes [%/100] | 631 | 0.277 (0.246) | 0 – 0.833 |
| Season | Binary indicators (spring 55.5%, summer 17.7%, fall 26.8%) | 631 | -- | -- |
| Examorder | Binary indicators for the order of exam (1st midterm exam 41.2%, 2nd midterm exam 9.4%, final exam 49.4%) | 631 | -- | -- |
| Male | Male student [binary] | 624 | 0.335 (0.472) | 0 – 1 |
| Race | Binary indicators (Asian 60.5%, Black 7.1%, Hispanic 0.7%, White 32.2%) | 365 | 0.071 (0.258) | 0 – 1 |
| Class level | Student's class level [1=freshman,…,4=senior,5=above] | 466 | 2.526 (1.078) | 1 – 5 |
| Major | Binary indicators (social & political science 43.6%, quantitative science 5.4%, applied science 15.5%, natural science 6.6%, humanities, language & education 14.3%, finance & management 10.8%) | 427 | -- | -- |
| Grade | Grade on the exam, out of all possible points [%/100] | 435 | 0.705 (0.169) | 0.20 – 1.00 |
| Course grade | Overall course grade, of 100 [%/100] | 477 | 0.729 (0.167) | 0.20 – 0.96 |
| Previous grade | Grade on previous exam, of all possible points [%/100] | 164 | 0.757 (0.165) | 0.20 – 0.99 |
| **Students' self-reported information** | | | | |
| Preveconcls | Count of previous Economics classes in college, 0.5 for high-school classes | 110 | 0.723 (0.954) | 0 – 4 |
| Repeat | Student is re-taking the class [binary] | 110 | 0.073 (0.261) | 0 – 1 |
| Grasplecture | How easy is it to follow lectures in English? [1=very uncomfortable,…,5=most comfortable] | 110 | 3.791 (0.927) | 1 – 5 |
| Grasptextbook | How easy is it to follow the English-language textbook? [1=very uncomfortable,…,5=most comfortable] | 110 | 3.909 (0.929) | 1 – 5 |
| Grasp mult. choice | How easy is it to read multiple-choice questions in English? [1=very uncomfortable,…,5=most comfortable] | 108 | 3.944 (0.846) | 2 – 5 |
| Comfortable w/ mult.choice | How comfortable are you solving multiple-choice tests? [1=very uncomfortable,…,5=most comfortable] | 106 | 3.967 (0.892) | 1 – 5 |
| Essaypref | Rank essay-type, concept-definition, graphing, mathematical & multiple-choice questions in terms of preference [1=essay-type least preferred,…,5=most] | 109 | 2.385 (1.199) | 1 – 4.5 |
| Essayperform | Rank 5 question types in how you perform on them relative to other students [1=essay-type worst,…,5=best] | 102 | 2.647 (1.232) | 1 – 5 |
| Conceptpref | [1=concept definition questions least preferred…5=most] | 109 | 3.046 (1.265) | 1 – 5 |
| Conceptperform | [1=concept questions perform worst,…,5=perform best] | 102 | 3.147 (1.155) | 1 – 5 |
| Math preference | [1=math questions least preferred,…,5=most preferred] | 109 | 2.927 (1.310) | 1 – 5 |
| Math perform. | [1=math questions perform worst,…,5=perform best] | 103 | 3.039 (1.196) | 1 – 5 |
| Graph prefer. | [1=graph questions least preferred,…,5=most preferred] | 109 | 3.376 (1.251) | 1 – 5 |
| Graph perform. | [1=graph questions perform worst,…,5=perform best] | 103 | 3.223 (1.120) | 1 – 5 |
| Mult. ch. pref. | [1=mult.choice least preferred,…,5=most preferred] | 109 | 3.899 (1.071) | 1 – 5 |
| Mult.ch.perform | [1=mult.choice perform worst,…,5=perform best] | 103 | 3.515 (1.136) | 1 – 5 |
| Comparative advantage | Student has comparative advantage in multiple-choice exams relative to other students [1=great disadvantage,…,5=great advantage] | 110 | 3.255 (0.806) | 2 – 5 |
| On-time usually | Do you usually have enough time to complete multiple-choice exam? [1=never,…5=always] | 110 | 3.536 (1.155) | 1 – 5 |
| Check answers usually | After completing multiple-choice exam, do you check your answers? [1=never,…5=always] | 108 | 3.875 (1.162) | 1 – 5 |
| Expected grade | Expected course grade, out of 100 [%/100] | 103 | 0.892 (0.053) | 0.70 – 0.95 |
| Grade important | Importance of 'good' course grade [1=not | 106 | 4.660 (0.660) | 2 – 5 |

important,…,5=very important]

[i] Evaluated in an unbalanced panel of 468 students for which dependent variable is non-missing (631 available observations). Sources: In-classroom observation during summer '03 – spring '09; Student questionnaire administered after assigning some practice multiple-choice problems just before midterm exam in fall '08 & spring '09 semesters.

As a dependent variable in the following statistical analysis, the percentage difference between student *i*'s time on an exam and the maximum time taken by any student on that exam is used: $timewaste_i = \left[\max_j(time_j) - time_i\right] \Big/ \max_j(time_j) \forall j$. The class-maximum time – rather than, say, the maximum instructor-allowed time – is used to account for unmeasured exam-specific characteristics that affect students' expected time to completion. The percentage difference – rather than simple time to completion or absolute difference – is used to account for systematic differences in the lengths of midterm and final exams, and across exams from different semesters.[1] The dependent variable has a nice interpretation as the share of effective time allotment that a student forsook or wasted by submitting his answer sheet early. Figure 1 shows the distribution of the variable by exam type and by selected student characteristics. The distribution appears censored at zero, with a significant fraction of observations submitting their answer sheets at the last minute. This suggests that some students may have wanted to submit their answers even later (i.e., *timewaste<0*) if not prevented by the deadline.

Figure 1. Distribution of the Time Forsaken by Students by Type of Exam, Students' Class Level, Gender and Race




The last histogram in Figure 1 suggests that the censoring at zero is most severe among Asian female students (in Korea), most likely because of the cultural setting placing high value on students' demonstration of effort to the last minute. The censoring at zero is also particularly marked in the case of final exams (when stake is higher), and among freshmen students (whose motivation or concentration has not eroded yet). However, these density

[1] Systematic differences across semesters may come from adjustments unwittingly done by the instructor in response to students' performance in prior semesters, from the changing population of exam questions available for assignment on new exams, and from the evolution of students' preparation for exams (e.g., observing exams from previous semesters, repeating the same class, etc.).

functions are not monotonically declining, and sometimes have multiple significant peaks. For instance, the distribution of time-wastage on final exams has peaks at 0% and at 50%. One interpretation is that on final exams, students are fairly certain about the effect of the exam score on their overall course performance. Students who are between two grades have a high motivation to exert all effort, while students who are certain of their grade only need to assure themselves of a minimum score.

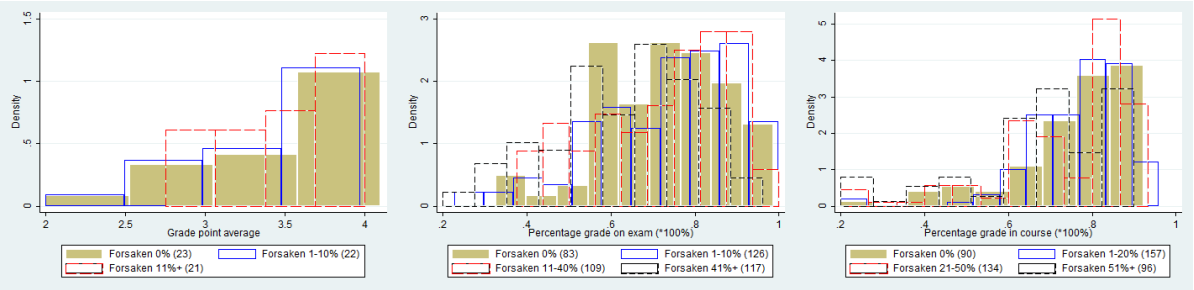
The distribution of the dependent variable suggests that censored-variable regression models or models allowing non-normal distribution of errors may perform better than ordinary, normal-distribution based models. The following empirical analysis thus uses tobit censored-variable models. Since there are two or three observations per student (in semesters when two midterms were administered), we can take advantage of the panel structure of the dataset to avoid estimator inconsistency due to unobserved time-constant heterogeneity across students. Tobit models are thus augmented by random or fixed effects. Random effects are expected to be more sensible in the available dataset due to the unbalanced nature of the sample, short time dimension, and limited variation in most regressors over time. Hausman specification is performed on selected model specifications to evaluate the existence of inconsistency in the ordinary tobit and random effects specifications.

Ordinary linear least squares models, and linear models with random or fixed effects are used for reference and as a robustness check. Finally, because of the non-monotonic distribution of the dependent variable, regression residuals are expected to be heteroskedastic. Serial correlation in residuals is also suspected across observations for the same students. Hence, standard errors of model estimates are corrected for heteroskedasticity and autocorrelation of model residuals of an unknown type.

To evaluate the bearing of *timewaste* on students' academic performance, students' grades are regressed on their forsaken time on exams. First, as a preliminary check of students' self-selection, *timewaste* is related to students' preexisting grade point average. The first histogram in Figure 2 shows that the relationship between students' academic status and their decision to waste time on exams is very weak. If anything, it appears that better students have less motivation or benefit from spending additional time on exams. In terms of grades received on the given exam or in the overall course, there appears to be a non-monotonic relationship: students forsaking 0% or more than 40% of the allowed time perform clearly worse on average than students who forsake just a few minutes. This suggests that students who (expect to) fare poorly on the exam and in the overall course are more likely to either give up on the exam significantly early or to scramble to the last minute. Students who

(expect to) perform well typically have a small time-reserve that they may choose to forsake. Alternatively, we may interpret these facts as showing that the observed marginal productivity of effort on exams is diminishing – in part due to students' self-selection. Initially, spending more time on the exam may improve performance, but eventually any observable marginal benefits disappear. This could be due to an interplay of three effects – productivity gains, endogenous self-selection, and attention erosion: 1) students spending incremental time on the exam may improve their answers; 2) only students who expect to benefit from the incremental time – or who are desperate to salvage their grade – decide to stay longer; 3) students spending too much time on the exam may change their correct answers to incorrect answers (particularly since these are special students have who have self-selected themselves into staying to the last minute).

Figure 2. Distribution of the Grade on the Exam, Course Grade, and Grade Point Average by Time Forsaken by Students on Exams




## Findings

Tables 3 and 4 report the results of tobit and random-effects tobit regressions of students' timewaste on the cultural setting of the class, students' demographic and academic background, and students' self-reported attitudes, habits and expectations. Students in Korea forsake significantly less of their exam time than students in the U.S. Men, Black students, and Asian students forsake less time than other groups, possibly for cultural reasons. Quantitative-science students forsake less time than students of applied science or humanities, suggesting that one's familiarity with the material improves one's productivity and motivation to exert marginal effort on exams. Similarly, students who scored well on a previous exam tend to spend more time on the following exam – likely a self-selection by student marginal productivity on exams. Also, controlling for other student characteristics, upper-classmen are shown to forsake less time than lower-classmen.

Table 2. Results of Tobit Models

| | Basic model | Gender & race | Student major | Full model | Student customs | Student attitude |
|---|---|---|---|---|---|---|
| Spring semester | 0.067** | -0.056* | -0.058* | -0.033 | | |
| | (.022)[.027] | (.032)[.029] | (.034)[.030] | (.032)[.030] | | |
| Summer semester | 0.068* | -0.036 | -0.166*** | -0.089 | | |
| | (.029)[.040] | (.064)[.070] | (.050)[.057] | (.064)[.077] | | |
| Korea | -0.342*** | -0.400*** | -0.422*** | -0.316*** | | |
| | (.021)[.023] | (.059)[.068] | (.032)[.035] | (.063)[.071] | | |
| Final exam | 0.038* | -0.008 | 0.014 | 0.015 | -0.046** | -0.037* |
| | (.021)[.020] | (.025)[.022] | (.026)[.021] | (.024)[.020] | (.025)[.022] | (.023)[.021] |
| Class level | | -0.025** | -0.008 | -0.020 | -0.041*** | -0.042*** |
| | | (.011)[.013] | (.010)[.012] | (.012)[.015] | (.015)[.015] | (.013)[.012] |
| Male | | -0.056 | | -0.027 | | |
| | | (.035)[.051] | | (.042)[.062] | | |
| Black | | -0.022 | | -0.090 | | |
| | | (.048)[.054] | | (.057)[.069] | | |
| Asian | | 0.010 | | -0.089 | | |
| | | (.059)[.075] | | (.065)[.078] | | |
| Social & political science | | | -0.004 | -0.023 | 0.049 | -0.042 |
| | | | (.031)[.038] | (.046)[.069] | (.087)[.095] | (.083)[.099] |
| Applied science | | | 0.011 | -0.035 | 0.438*** | 0.404*** |
| | | | (.032)[.040] | (.046)[.071] | (.115)[.092] | (.099)[.091] |
| Humanities, lang. & educ. | | | 0.038 | 0.052 | 0.151 | 0.065 |
| | | | (.037)[.044] | (.049)[.070] | (.091)[.097] | (.086)[.101] |
| Quantitative sciences | | | -0.054 | -0.119 | | |
| | | | (.049)[.072] | (.062)[.100] | | |
| On-time usually | | | | | 0.019 | -0.002 |
| | | | | | (.013)[.013] | (.014)[.013] |
| Check answers usually | | | | | -0.016 | -0.005 |
| | | | | | (.012)[.014] | (.012)[.015] |
| Grade important | | | | | | -0.072*** |
| | | | | | | (.019)[.019] |
| Grasp multiple choice | | | | | | 0.007 |
| | | | | | | (.016)[.022] |
| Comfortable w/ multiple choice | | | | | | -0.011 |
| | | | | | | (.015)[.022] |
| Constant | 0.303*** | 0.517*** | 0.494*** | 0.506*** | 0.068 | 0.536*** |
| | (.021)[.029] | (.067)[.077] | (.052)[.049] | (.072)[.094] | (.110)[.109] | (.167)[.166] |
| Observations | 631 | 325 | 427 | 287 | 97 | 91 |
| Students | 468 | 203 | 307 | 167 | 51 | 48 |
| Pseudo R-square | 0.753 | 1.048 | 1.030 | 1.140 | -5.699 | -13.006 |
| F-statistic | 83.43*** | 26.04*** | 44.85*** | 10.75*** | 89.38*** | 93.77*** |

Standard errors robust to arbitrary heteroskedasticity and autocorrelation at the student level are in brackets. Non-robust standard errors are in parentheses. * significant at 10%, ** 5%, *** 1% two-sided test using robust standard errors. Math & science major indicators are omitted in columns 5-6 due to collinearity.

Table 3. Results of Panel Tobit Models with Random Effects

| | Basic model | Gender & race | Student major | Full model | Student customs | Student attitude |
|---|---|---|---|---|---|---|
| Spring semester | 0.050** | -0.056 | -0.065 | -0.036 | | |
| | (0.025) | (0.039) | (0.042) | (0.038) | | |
| Summer semester | 0.038 | -0.041 | -0.199*** | -0.130* | | |
| | (0.032) | (0.067) | (0.055) | (0.069) | | |
| Korea | -0.356*** | -0.371*** | -0.421*** | -0.296*** | | |
| | (0.025) | (0.061) | (0.033) | (0.068) | | |
| Final exam | 0.022* | -0.006 | 0.004 | 0.006 | -0.043** | -0.038* |
| | (0.013) | (0.017) | (0.017) | (0.017) | (0.022) | (0.023) |
| Class level | | -0.018 | -0.003 | -0.016 | -0.039** | -0.042*** |
| | | (0.012) | (0.011) | (0.014) | (0.016) | (0.013) |
| Male | | -0.026 | | 0.004 | | |
| | | (0.038) | | (0.047) | | |
| Black | | -0.011 | | -0.087 | | |
| | | (0.053) | | (0.066) | | |
| Asian | | -0.008 | | -0.110 | | |
| | | (0.063) | | (0.071) | | |
| Social & political science | | | 0.002 | -0.017 | 0.047 | -0.042 |
| | | | (0.033) | (0.052) | (0.095) | (0.083) |
| Applied science | | | 0.019 | -0.032 | 0.435*** | 0.404*** |
| | | | (0.034) | (0.052) | (0.129) | (0.099) |
| Humanities, lang. & educ. | | | 0.026 | 0.053 | 0.148 | 0.065 |
| | | | (0.040) | (0.056) | (0.100) | (0.086) |
| Quantitative sciences | | | 0.000 | -0.073 | | |
| | | | (0.056) | (0.073) | | |
| On-time usually | | | | | 0.019 | -0.002 |
| | | | | | (0.015) | (0.014) |
| Check answers usually | | | | | -0.018 | -0.005 |
| | | | | | (0.014) | (0.012) |
| Grade important | | | | | | -0.072*** |
| | | | | | | (0.019) |
| Grasp multiple choice | | | | | | 0.007 |
| | | | | | | (0.016) |
| Comfortable w/ multiple choice | | | | | | -0.011 |
| | | | | | | (0.015) |
| Constant | 0.345*** | 0.496*** | 0.493*** | 0.503*** | 0.072 | 0.536*** |
| | (0.023) | (0.069) | (0.054) | (0.076) | (0.122) | (0.167) |
| Observations | 631 | 325 | 427 | 287 | 97 | 91 |
| Students | 468 | 203 | 307 | 167 | 51 | 48 |
| F-statistic | 60.74*** | 25.12*** | 32.35*** | 11.49*** | 4.98*** | 6.74*** |

Standard errors robust to arbitrary heteroskedasticity and autocorrelation at the student level are in brackets. Non-robust standard errors are in parentheses. * significant at 10%, ** 5%, *** 1% two-sided test using robust standard errors. Math & science major indicators are omitted in columns 5-6 due to collinearity.

Regarding students' personalities I find that students who reportedly typically check their answers before submitting them indeed take more time to finish exams. Students who attribute more importance to grades also spend more time on exams. However, there is no indication that students who reportedly usually finish their exams on time submit their answer sheets earlier.

Overall, students' allocation of time on exams is found to vary significantly by cultural setting of the class, students' demographic and academic background, and students' latent

skills and values. Most results validate a hypothesis that students' expected marginal productivity on exam questions affects their effort positively.

## Conclusions and Test-Format Recommendations

This study has evaluated students' time allocation on multiple-choice exams in Principles of (micro)economics courses. Broadly, this study contributes to the literature explaining the process of production of education among students (Brown & Saks 1980, 1984; Hanushek 1986). The study can be viewed as indirectly evaluating the role of the Big Five personality traits or soft skills – openness, conscientiousness, extraversion, agreeableness and neuroticism (OCEAN) – on young adults' academic performance. Among these five, this study suggests that meticulousness and distractibility are important to students' performance and their decision to exert effort on assignments.

The most interesting finding for educators is that students do not appear to utilize the available time efficiently. Their chosen time allocation cannot be attributed to a particular leisure-grade tradeoff, or a tradeoff of productivities at different tasks. A significant portion of students submit the exam a long time before the official time expires. Students who submit significantly early tend to score less than students who stay longer. However, students who leave a few minutes early tend to outperform those who stay the entire allowed time. Even among students who submit early, some students leave unanswered questions (even though the penalty for wrong answers is the same as for unanswered questions).

Students' motivation may explain this behavior in part. Students with self-reported poor reading comprehension or comparative disadvantage at multiple-choice tests, those expecting lower grades, and those attributing lower importance to grades exert less effort on exams, and exhibit myopia. They view marginal effort on exams as bringing them insufficient instantaneous returns in terms of grade or overall welfare.

Another explanation has to do with poor allocation of time across activities. Students who tend to submit early are those who self-reportedly do not return to individual questions multiple times, and do not check their answers, instead running through the exam a single time before handing in their solutions. These students may falsely believe that they have answered the exam correctly and in full. In agreement with this idea, among students who leave early, unanswered questions tend to occur in the first half of the exam, whereas students who stay the entire allowed time may leave unanswered questions near the end of the exam.

There is also some evidence that students do not allocate time across exam-taking tasks efficiently, when their time is constrained. Students who self-reportedly check their answers

after running through the exam perform worse than students who read questions multiple times before choosing their answers. On average, students appear to allocate too much time to checking their answers, and too little time attempting to comprehend questions (or searching for hints in the rest of the exam) prior to answering them.

These findings yield several prescriptions for educators: Analogously to the Stroop test, I find that the format of the examination affects ease of students' comprehension and speed of their completion of the examination. Length of the exam and time allotment are important, because they force students to trade off multiple tasks, in face of attention attrition and myopia. The order of questions on multiple-choice exams is important, as it affects students' need to check answers or pre-read questions before answering them, and students' willingness to answer each question carefully. Ultimately, there are tradeoffs involved in designing an exam for a heterogeneous group of students, but the best design clearly depends on the background of the median test-taker. Finally, especially in heterogeneous groups, counseling of students about their strategies, and about the format and content of upcoming exams can be very valuable. Working through a set of sample questions in classroom, including pre-reading of questions and checking of answers, is recommended. Trial exams under real test-day conditions (self-administered by students) are also helpful. Students with alternative learning needs should be given special instructions on time management, as well as real-time guidance during exams.

To the extent that some students may not have experience with the multiple-choice format, they should be instructed to prepare for multiple-choice tests effectively, and to learn strategies for focusing attention, reading comprehension and checking of responses. For students with alternative learning needs, changes related to content order, instructions and guidance through the exam, time allotment and test settings may be warranted.

## References


Bettinger, Eric P., Brent J. Evans, and Devin G. Pope (2013), Improving College Performance and Retention the Easy Way: Unpacking the ACT Exam, NBER Working Paper No. 17119.

Brown, Byron W. (1991), How gender and socioeconomic status affect reading and mathematics achievement, Economics of Education Review 10(4):343-357.

Brown, Byron W., and Carl E. Liedholm (2002), Can Web Courses Replace the Classroom in Principles of Microeconomics? American Economic Review 92(2):444-448.

-- (2004), Student preferences in using online learning resources, Social Science Computer Review 22(4):479-492.

Brown, Byron W., and Daniel H. Saks (1980), Production technologies and resource allocations within classrooms and schools: Theory and measurement, The analysis of educational productivity 1:53-117.
-- (1984), An economic view of the acquisition of reasoning skills: Agenda for research in the information age, Review of Educational Research 54(4):560-576.
-- (1986), Measuring the effects of instructional time on student learning: Evidence from the beginning teacher evaluation study, American Journal of Education 94(4):480.
-- (1987), The microeconomics of the allocation of teachers' time and student learning, Economics of Education Review 6(4):319-332.
Hanushek, Eric A. (1986), The economics of schooling: Production and efficiency in public schools, Journal of economic literature 24(3):1141-1177.
Kaplan, D. A. (2009), The Standardized-Test Smackdown, Fortune, 30 November 2009.
Wagner, Tony (2010), The global achievement gap, New York: Basic Books.

## Appendix 1. Additional Results

Table A1. OLS Regression Results

| | Basic model | Gender & race | Student major | Full model | Student customs | Student attitude |
|---|---|---|---|---|---|---|
| Spring semester | 0.078*** | -0.022 | -0.026 | -0.009 | | |
| | (.019)[.024] | (.026)[.020] | (.028)[.020] | (.025)[.021] | | |
| Summer semester | 0.079** | 0.008 | -0.117** | -0.061 | | |
| | (.026)[.037] | (.052)[.062] | (.041)[.048] | (.051)[.068] | | |
| Korea | -0.292*** | -0.319*** | -0.349*** | -0.243*** | | |
| | (.018)[.018] | (.048)[.061] | (.026)[.027] | (.050)[.062] | | |
| Final exam | 0.048*** | 0.018 | 0.039** | 0.037** | -0.012 | -0.008 |
| | (.018)[.017] | (.020)[.016] | (.021)[.015] | (.019)[.015] | (.017)[.016] | (.016)[.015] |
| Class level | | -0.014 | 0.000 | -0.009 | -0.019** | -0.020*** |
| | | (.009)[.010] | (.008)[.010] | (.009)[.011] | (.009)[.008] | (.008)[.008] |
| Male | | -0.050 | | -0.022 | | |
| | | (.029)[.049] | | (.034)[.060] | | |
| Black | | -0.020 | | -0.090 | | |
| | | (.040)[.051] | | (.046)[.067] | | |
| Asian | | -0.002 | | -0.100 | | |
| | | (.048)[.068] | | (.053)[.069] | | |
| Social & political science | | | -0.002 | -0.021 | 0.036 | -0.017 |
| | | | (.026)[.034] | (.037)[.063] | (.054)[.034] | (.052)[.040] |
| Applied science | | | 0.009 | -0.032 | 0.361*** | 0.350*** |
| | | | (.027)[.038] | (.038)[.068] | (.078)[.030] | (.069)[.035] |
| Humanities, lang. & educ. | | | 0.021 | 0.032 | 0.102** | 0.057 |
| | | | (.031)[.040] | (.040)[.065] | (.058)[.040] | (.056)[.046] |
| Quantitative sciences | | | -0.053 | -0.112 | | |
| | | | (.041)[.058] | (.049)[.082] | | |
| On-time usually | | | | | 0.014 | -0.001 |
| | | | | | (.009)[.009] | (.009)[.007] |
| Check answers usually | | | | | -0.012 | -0.005 |
| | | | | | (.008)[.010] | (.008)[.011] |
| Grade important | | | | | | -0.057*** |
| | | | | | | (.014)[.015] |
| Grasp multiple choice | | | | | | 0.003 |
| | | | | | | (.011)[.015] |
| Comfortable w/ multiple choice | | | | | | -0.003 |
| | | | | | | (.011)[.016] |
| Constant | 0.295*** | 0.440*** | 0.424*** | 0.441*** | 0.064 | 0.402*** |
| | (.018)[.026] | (.054)[.067] | (.042)[.038] | (.057)[.085] | (.073)[.053] | (.116)[.107] |
| Observations | 631 | 325 | 427 | 287 | 97 | 91 |
| Students | 468 | 203 | 307 | 167 | 51 | 48 |
| R-squared | 0.36 | 0.48 | 0.52 | 0.45 | 0.35 | 0.50 |
| F-statistic | 107.18*** | 23.41*** | 50.67*** | 9.32*** | 85.66*** | 89.49*** |

Standard errors robust to arbitrary heteroskedasticity and autocorrelation at the student level are in brackets. Non-robust standard errors are in parentheses. * significant at 10%, ** 5%, *** 1% two-sided test using robust standard errors. Math & science major indicators are omitted in columns 5-6 due to collinearity.

Table A2. Results of Panel OLS Models with Random Effects

| | Basic model | Gender & race | Student major | Full model | Student customs | Student attitude |
|---|---|---|---|---|---|---|
| Spring semester | 0.059** | -0.021 | -0.028 | -0.010 | | |
| | (.022)[.023] | (.033)[.018] | (.035)[.018] | (.031)[.019] | | |
| Summer semester | 0.048 | 0.003 | -0.154*** | -0.105* | | |
| | (.030)[.033] | (.058)[.060] | (.047)[.043] | (.058)[.065] | | |
| Korea | -0.311*** | -0.301*** | -0.355*** | -0.230*** | | |
| | (.022)[.017] | (.053)[.063] | (.028)[.026] | (.058)[.063] | | |
| Final exam | 0.032*** | 0.015 | 0.022* | 0.022* | -0.011 | -0.008 |
| | (.010)[.011] | (.012)[.012] | (.011)[.012] | (.011)[.012] | (.015)[.016] | (.014)[.015] |
| Class level | | -0.008 | 0.004 | -0.006 | -0.018** | -0.020*** |
| | | (.011)[.011] | (.010)[.010] | (.012)[.011] | (.010)[.008] | (.009)[.008] |
| Male | | -0.021 | | 0.009 | | |
| | | (.034)[.044] | | (.040)[.054] | | |
| Black | | -0.010 | | -0.093 | | |
| | | (.048)[.049] | | (.058)[.060] | | |
| Asian | | -0.015 | | -0.119* | | |
| | | (.056)[.075] | | (.062)[.071] | | |
| Social & political science | | | 0.004 | -0.016 | 0.035 | -0.021 |
| | | | (.029)[.033] | (.045)[.062] | (.059)[.036] | (.058)[.042] |
| Applied science | | | 0.016 | -0.031 | 0.359*** | 0.347*** |
| | | | (.030)[.035] | (.045)[.063] | (.086)[.031] | (.077)[.037] |
| Humanities, lang. & educ. | | | 0.013 | 0.035 | 0.101** | 0.052 |
| | | | (.036)[.040] | (.048)[.065] | (.063)[.042] | (.063)[.048] |
| Quantitative sciences | | | -0.003 | -0.077 | | |
| | | | (.050)[.045] | (.063)[.075] | | |
| On-time usually | | | | | 0.014 | -0.001 |
| | | | | | (.010)[.009] | (.011)[.007] |
| Check answers usually | | | | | -0.013 | -0.006 |
| | | | | | (.009)[.010] | (.009)[.011] |
| Grade important | | | | | | -0.056*** |
| | | | | | | (.015)[.015] |
| Grasp multiple choice | | | | | | 0.005 |
| | | | | | | (.012)[.015] |
| Comfortable w/ multiple choice | | | | | | -0.004 |
| | | | | | | (.012)[.017] |
| Constant | 0.339*** | 0.423*** | 0.428*** | 0.444*** | 0.065 | 0.403*** |
| | (.021)[.023] | (.059)[.064] | (.046)[.037] | (.064)[.086] | (.080)[.053] | (.130)[.109] |
| Observations | 631 | 325 | 427 | 287 | 97 | 91 |
| Students | 468 | 203 | 307 | 167 | 51 | 48 |
| Overall R-square | 0.35 | 0.48 | 0.51 | 0.44 | 0.35 | 0.50 |
| F-statistic | 119.40*** | 24.59*** | 50.34*** | 9.22*** | 253.98*** | 290.69*** |

Standard errors robust to arbitrary heteroskedasticity and autocorrelation at the student level are in brackets. Non-robust standard errors are in parentheses. * significant at 10%, ** 5%, *** 1% two-sided test using robust standard errors. Math & science major indicators are omitted in columns 5-6 due to collinearity.

Table A3. Results of Panel Tobit and OLS Models with Fixed Effects

| | Tobit | | OLS | |
|---|---|---|---|---|
| Spring semester | -0.034 | -0.123 | -0.006*** | -0.348** |
| | (.080) | (.113) | (.090)[.000] | (.163)[.141] |
| Final exam | 0.050*** | 0.077*** | 0.030*** | 0.064*** |
| | (.019) | (.029) | (.011)[.011] | (.019)[.018] |
| Previous grade | | -0.902* | | -0.580** |
| | | (.552) | | (.245)[.236] |
| Constant | | | 0.265*** | 0.689*** |
| | | | (.050)[.005] | (.256)[.246] |
| Observations | 631 | 164 | 631 | 164 |
| Students | 468 | 149 | 468 | 149 |
| Within R-square | -- | -- | 0.050 | 0.620 |
| F-statistic | 7.16*** | 43.45*** | 1.95** | 2.15*** |

Regressors are limited to those varying for individuals over time. Standard errors robust to arbitrary heteroskedasticity and autocorrelation at the student level are in brackets. Non-robust standard errors are in parentheses. * significant at 10%, ** 5%, *** 1% two-sided test using non-robust errors in tobit models, and using robust standard errors in OLS models.

## Appendix 2. Student Questionnaire about Examinations in the Course

Name: …….

1) On a scale from 1 to 5, how comfortable are you understanding lectures in English at Ewha (not just this class)?
(1 very uncomfortable … 5 most comfortable)

2) On a scale from 1 to 5, how comfortable are you understanding the English in our textbook?
(1 very uncomfortable … 5 most comfortable)

3) On a scale from 1 to 5, how comfortable are you understanding the English in multiple-choice questions in this class?
(1 very uncomfortable … 5 most comfortable)

4) On a scale from 1 to 5, how comfortable are you taking multiple choice tests?
(1 very uncomfortable … 5 most comfortable)

5) Please rank the following types of tests in how you prefer them (1 least preferred … 5 most preferred) & in how you usually perform on them compared to other students (1 perform worst … 5 perform best):

Essay questions __ & __
Questions about definitions of concepts __ & __
Mathematical questions __ & __
Graphing questions __ & __
Multiple-choice questions __ & __

6) On a scale from 1 to 5, would you say you have comparative advantage to other students in taking multiple-choice tests, compared to other tests?
(1 great disadvantage … 5 great advantage)

7) On a scale from 1 to 5, can you usually complete a multiple choice test on time?
(1 never enough time … 5 always enough time)

8) On a scale from 1 to 5, after completing a multiple-choice test, do you go back and check your answers?
(1 never check answers … 5 always check answers)

9) How many Economics classes have you taken before? Were they high school or college classes?

10) What grade do you expect to receive in this class?

11) On a scale from 1 to 5, how important is it to you to get a good grade in this class?
(1 not important … 5 very important)

12) What is your current grade point average?

13) Do you have any other requests or feedback about assignments or exams in this class?