MATERIALS APPENDIX FOR: CORRELATION NEGLECT IN STUDENT-TO-SCHOOL MATCHING

ALEX REES-JONES, RAN SHORRER, AND CHLOE TERGIMAN

In this appendix we present the instructions that the subjects received in print and that were read out loud.

The instructions are for the treatment in which the subjects saw the uncorrelated arm first. Note that the instructions for the treatment in which the subjects saw the correlated arm first are identical except for the order of those two modules.

After the instructions, we also present the consent form and display the sequence of screens that subjects who saw the uncorrelated arm first faced.

At the end of this document, we present our preregistration for reference. It is also available at https://aspredicted.org/8st27.pdf.

Date: **January 2020**.

1. Instructions and Screen Shots

INSTRUCTIONS - Preamble

This LEMA session consists of two Studies, Study 1 and Study 2. Study 1 is in 3 Parts. Study 2 is in 2 Parts. You will be paid for both Studies. Furthermore, payment in all Parts are completely independent: nothing you do in one Part has any impact whatsoever in any other Part and no decision you make at any point can increase, decrease, or change in any other way your opportunities in any other Part.

Study 1 - Preamble

Study 1 consists of 3 Parts. Each Part will consist of 9 Rounds for a total of 27 Rounds. In Study 1, your earnings will be determined by your choices in only one, randomly chosen, Round. Thus it is in your best interest to treat each Round in isolation, as if it were the one that mattered for payment.

PART I

Part I is in 9 Rounds. In this part of the study, you will face a simulation of applying to college. In each Round, there are three colleges accepting applications, College A, College B and College C, and you will choose where to apply.

Your earnings in a Round:

Your earnings will depend on the college where you ultimately enroll.

- If you enroll in College A, you will receive a $10 bonus.
- If you enroll in College B, you will receive a $5 bonus.
- If you enroll in College C, you will receive a $2.5 bonus.

However, you may only apply to two colleges, and you have to tell us your first choice and your second choice.

(1) If your first choice admits you, you will enroll in your first choice college and earn the bonus associated with that college.
(2) If your first choice rejects you and your second choice admits you, you will enroll in your second choice college and earn the bonus associated with that college.
(3) If both your first choice and second choice reject you, you will receive no bonus.

College Admissions:

College admissions depend on test scores. If you apply to a college, they will admit you only if your test score is greater than or equal to the minimum score that they accept.

(1) Each college has its own admissions test.
(2) Your tests scores are randomly and independently generated between 0 and 99. That means that:
    - Your test score for College A and has an equal probability of being any whole number from 0 to 99.
    - Your test score for College B also has an equal probability of being any whole number from 0 to 99.
    - Your test score for College C also has an equal probability of being any whole number from 0 to 99.

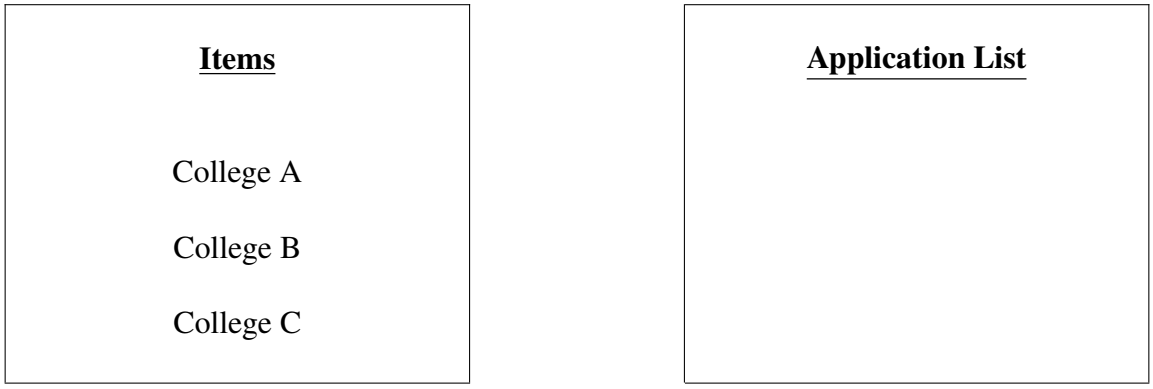
    Your test scores for Colleges A, B and C have no bearing on one another: each is a whole number that is randomly chosen from the 0 to 99 interval.
(3) Your test scores are randomly generated at the start of Part I and will remain the same for all the Rounds in Part I.
(4) Those test scores will not be revealed to you. Thus you will move from one Round to the next without knowing which college you enrolled in (if any) in the previous Round.

All the Rounds will have the same format. The minimum score a college accepts will be presented to you in a table. For example, that table will look like:

	Bonus if you enroll	**Minimum test score**
College A	$10	65
College B	$5	45
College C	$2.5	15

Here, to be able to be admitted to College A, your test score for College A has to be at least 65. In order to be admitted to College B, your test score for College B has to be at least 45. In order to be admitted to College C, your test score for College C has to be at least 15.

Below that table, you will see the three colleges in a list and will have to rank two of them by dragging and dropping those two, in order, into the table called "Application List":



If this Part is chosen to count, only one Round will matter for payment. So it is in your best interest to treat each Round in isolation, as if it were the one that mattered for payment.

Do you have any questions?

Before we start Part I, you will face an example scenario (practice Round) where you can familiarize yourself with the screen and procedures. Nothing you do in this practice Round impacts your earnings in any way.

PART II

Part II is in 9 Rounds. This Part of the study is identical to Part I, except that now, **the same test score will be used for ALL colleges.**

Your earnings in a Round:

Just like in Part I, your earnings will depend on the college where you ultimately enroll.

- If you enroll in College A, you will receive a $10 bonus.
- If you enroll in College B, you will receive a $5 bonus.
- If you enroll in College C, you will receive a $2.5 bonus.

Similarly to Part I, you may only apply to two colleges, and you have to tell us your first choice and your second choice.

(1) If your first choice admits you, you will enroll in your first choice college and earn the bonus associated with that college.
(2) If your first choice rejects you and your second choice admits you, you will enroll in your second choice college and earn the bonus associated with that college.
(3) If both your first choice and second choice reject you, you will receive no bonus.

College Admissions:

College admissions depend on a test score. If you apply to a college, they will admit you only if your test score is greater than or equal to the minimum score that they accept.

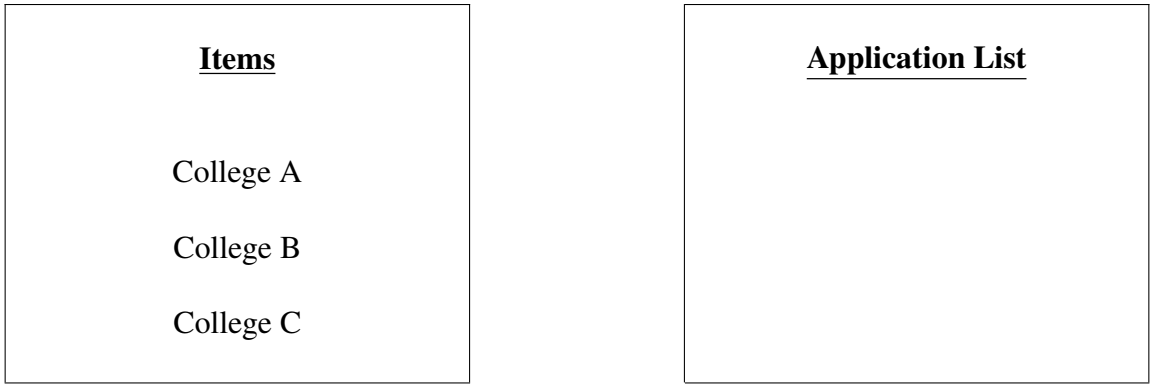
(1) The same admissions test is accepted by all colleges.
(2) Your test score is randomly generated, and has an equal probability of being any whole number from 0 to 99.
(3) Your test score is randomly generated at the start of Part II and will remain the same for all the Rounds in Part II.
(4) Your test score will not be revealed to you. Thus you will move from one Round to the next without knowing which college you enrolled in (if any) in the previous Round.

All the Rounds will have the same format. The minimum score a college accepts will be presented to you in a table. For example, that table will look like:

	Bonus if you enroll	**Minimum test score**
College A	$10	65
College B	$5	45
College C	$2.5	15

Here, to be able to be admitted to College A, your test score has to be at least 65. In order to be admitted to College B, your test score has to be at least 45. In order to be admitted to College C, your test score has to be at least 15.

Below that table, you will see the three colleges in a list and will have to rank two of them by dragging and dropping those two, in order, into the table called "Application List":



If this Part is chosen to count, only one Round will matter for payment. So it is in your best interest to treat each Round in isolation as if it were the one that mattered for payment.

Do you have any questions?

Before we start Part II, you will face an example scenario (practice Round) where you can familiarize yourself with the screen and procedures. Nothing you do in this practice Round impacts your earnings in any way.

PART III

Part III is in 9 Rounds. In each Round of this Part of the study, you will be asked to choose between two options. Each option is a lottery.

Here is an example of a Round that you can encounter (the choices you face will be different and will vary from Round to Round):

In this example, you have a choice between:

- a lottery in which you have a 45% chance of earning $10 and a 55% chance of earning $0.
  OR
- a lottery in which you have a 40% chance of earning $5 and a 60% chance of earning $2.50.

Here are how your earnings would be determined if this were the Round that mattered for payment:

- If you chose the lottery on the left, if this were the Round that mattered for payment, the computer would run the lottery you selected. Then, with 45% chance you would earn $10 and with %55 chance you would earn $0.
- If instead you chose the pair on the right, if this were the Round that mattered for payment, the computer would run the lottery you selected. Then, with 40% chance you would earn $5 and with 60% chance you would earn $2.50.

If this Part is chosen to count, only one Round will matter for payment. So it is in your best interest to treat each Round as if it were the one that mattered for payment.

Do you have any questions?

STUDY 2

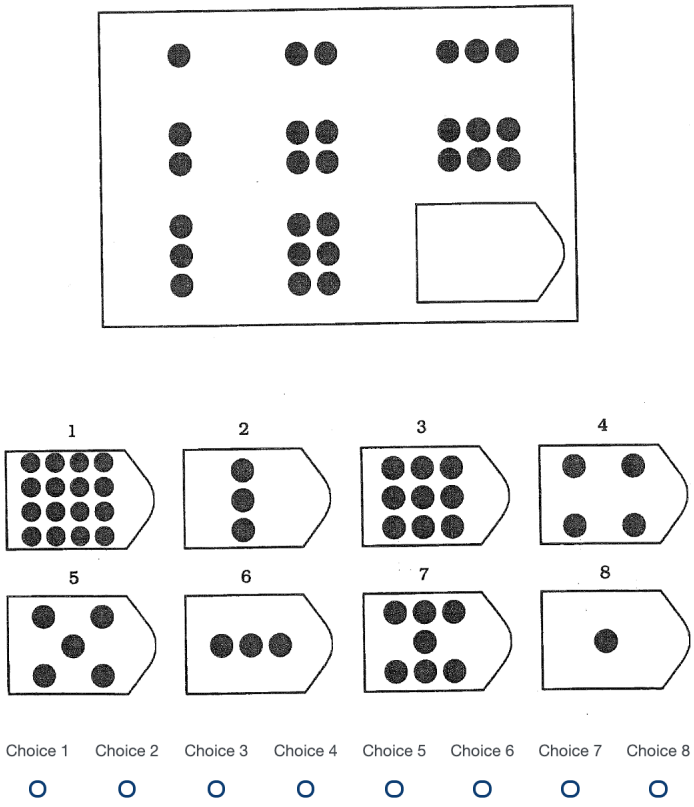
Welcome to Study 2. Study 2 consists of 2 Parts. You will be paid for each Part. We will hand out instructions for Part II when you have completed Part I.

PART I

Part I lasts for exactly 5 minutes. During those 5 minutes you will face a series of problems. These problems are taken from a test of spatial reasoning.

Each problem consists of a series of patterns. One of the patterns has been removed. Your task is to find the missing pattern to complete problem.

An example of such a problem is below:



You have 8 options to choose from to complete the pattern in each problem. Simply select your choice by clicking the number of the choice that completes the pattern.

You will receive $1 per correctly solved problem. Please solve as many of these problems as you can in 5 minutes.

There is a timer at the top of the page for your reference.

PART II

This Part consists of 3 Rounds. **In each Round, you will have to estimate a different, unknown number X.** That is, the X number for the first Round has been randomly determined, and that X number is independent from the X number that has randomly been determined for the second Round, and that X number is independent from the X number that has been randomly determined for the third Round. In each Round, the number X that has been chosen will not be revealed to you.

As will be explained in more detail below, you will receive some information about this number. Then you will be asked to provide an estimate for X.

Your earnings in a Round will depend on how precisely you estimate X, i.e., how close your estimate is to the actual number X. Only one of the three Rounds will count for payment, and you will be paid according to the precision of your estimate in that Round. This will be explained in more detail in the next section. You can earn either \$10 or \$0. Below we describe how you can earn \$10.

After you have provided your estimate, we will compare your estimate to the true value of X and see how close you are to being right. You will receive \$10 if the square of the distance between your estimate and the actual number X is less than or equal to a number k (which is unknown to you). If the square of the distance is larger than k, you will receive 0 dollars. This number k lies somewhere between 0 and 10,000, and the computer chose k randomly with each number between 0 and 10,000 being equally likely. The determination of your earnings can be expressed with the following formula:

- Payment = \$10, if $(X - \text{"your estimate"})^2 \leq k$
- Payment = \$0, if $(X - \text{"your estimate"})^2 > k$

While this formula might look complicated, the underlying principle is very simple: the better your estimate, i.e., the smaller the distance between your estimate and the true value of X, the higher the likelihood that you will receive the \$10. In other words, you should try to estimate X as best you can.

Consider the following example: suppose that in a Round the computer drew the number **X = 150**. In addition, suppose that **k = 500**. You would then earn the following amount of money depending on your estimate:

$$\text{If Your Estimate} = 50 \quad \rightarrow (150 - \text{Your Estimate})^2 = (150-50)^2 = 100^2 = 10,000 > k \quad \rightarrow earnings = \$0$$

$$\text{If Your Estimate} = 100 \quad \rightarrow (150 - \text{Your Estimate})^2 = (150-100)^2 = 50^2 = 2,500 > k \quad \rightarrow earnings = \$0$$

$$\text{If Your Estimate} = 130 \quad \rightarrow (150 - \text{Your Estimate})^2 = (150-130)^2 = 20^2 = 400 \leq k \quad \rightarrow earnings = \$10$$

$$\text{If Your Estimate} = 150 \quad \rightarrow (150 - \text{Your Estimate})^2 = (150-150)^2 = 0^2 = 0 \leq k \quad \rightarrow earnings = \$10$$

$$\text{If Your Estimate} = 170 \quad \rightarrow (150 - \text{Your Estimate})^2 = (150-170)^2 = (-20)^2 = 400 \leq k \quad \rightarrow earnings = \$10$$

$$\text{If Your Estimate} = 200 \quad \rightarrow (150 - \text{Your Estimate})^2 = (150-200)^2 = (-50)^2 = 2,500 > k \quad \rightarrow earnings = \$0$$

$$\text{If Your Estimate} = 250 \quad \rightarrow (150 - \text{Your Estimate})^2 = (150-250)^2 = (-100)^2 = 10,000 > k \quad \rightarrow earnings = \$0$$

You can see that your chances of receiving \$10 only depend on the absolute distance. That is, it does not matter whether you overestimate or underestimate the true number by the same amount, since the square of the distance will be the same.

The estimation task:

As already mentioned, in each Round you will have to provide an estimate regarding an unknown number X. Also, as already mentioned, for each Round the the number X is different. You will not know this number. A computer has drawn this number for each Round from a probability distribution, that is displayed below. The distribution you see in Figure 1 is what is called a normal distribution. The distribution has a mean of 0 and a standard deviation of 500. Although you will not know the number X, the graph tells you something about the range from which X was drawn by the computer.

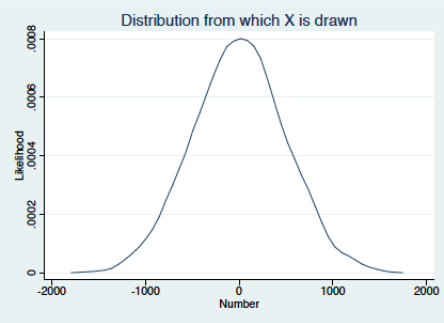


FIGURE 1. Distribution from which the computer drew X from.

Information regarding the estimation tasks:

Your task in each Round is to provide an estimate about a randomly drawn number X (unknown to you). To help you estimate X, in each Round you will receive different computer-generated pieces of information about the correct value of X. In each Round, you will see this information and then enter your own estimate.

This information comes from the following two types of devices:

- First, there are four "Estimation Devices" (Estimation Devices A, B, C and D), which themselves provide an estimate of X.
- Second, there are four "Communication Machines" (Communication Machines 1, 2, 3 and 4), which observe the estimates of the Estimation Devices and compute their own estimate from these reports.

The Estimation Devices provide an estimate about the number X, and the estimates of these devices are completely independent from each other. The Estimation Devices have determined an estimate by randomly drawing a number from a normal distribution, as seen below in Figure 2. The Estimation Devices all have the same quality, i.e., they are equally good in determining estimates.

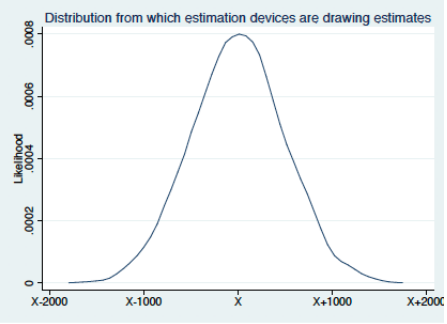


FIGURE 2. Distribution from which the Estimation Devices draw their estimates.

Importantly, this distribution takes as mean the number X, and a standard deviation of 500. You can see that the highest point of the bell curve is at the number X, i.e., the correct value. The further you move away from X, the less likely it is that the corresponding numbers are drawn from the Estimation Devices.

This means that the Estimation Devices are good at solving the estimation task. If the Estimation Devices would provide a large number of estimates, then the average of these estimates would be correct. While almost every individual estimate will be incorrect, the average taken over many estimates will be very precise. In addition, many estimates will be rather close to the correct value.

In short:

(1) The Estimation Devices give you an estimate that they have drawn from a normal distribution with mean X. This means that the Estimation Devices are good at solving the estimation task. If the Estimation Devices would provide a large number of estimates, i.e., if they would draw many times from the normal distribution, then the average of these estimates would be correct (or very precise).
(2) The Estimation Devices make mistakes, but it is much more likely that the estimate is close to the true value than that it is very far away.
(3) For every estimation task, there are a total of four Estimation Devices (A, B, C, D). These four devices, which are completely independent and separate from each other, each randomly draw an estimate from the normal distribution (with mean X and a standard deviation of 500).

Apart from the Estimation Devices, there are also four Communication Machines (1, 2, 3, 4). These Communication Machines do not determine their own estimates. Figure 3 below shows how the Estimation Devices A, B, C and D transmit their estimates to the Communication Machines 1, 2, 3 and 4, and then how those Communication Machines communicate that information to you.

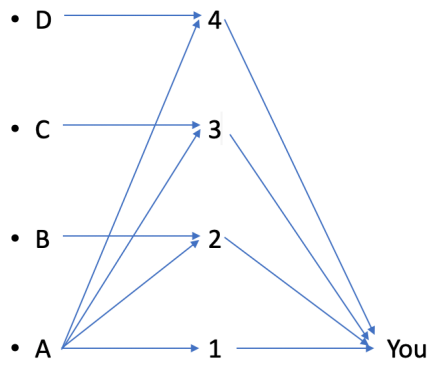


FIGURE 3. Communication Machine 1 transmits the estimate generated by Estimation Device A. Each of the three other Communication Machines receives the estimate of Estimation Device A and of one other Estimation Device, as described by the arrows. They compute their own estimate by taking the average of the two estimates.

This means that you will receive the following information:

(1) As is evident from Figure 3, Communication Machine 1 receives the estimate from Estimation Device A and reports A's estimate to you.
(2) All the other Communication Machines all see the estimate of Estimation Device A and of one other Estimation Device. As you can see in Figure 3:
    - Communication Machine 2 receives the estimates of Estimation Devices A and B.
    - Communication Machine 3 receives the estimates of Estimation Devices A and C.
    - Communication Machine 4 receives the estimates of Estimation Devices A and D.

    Communication Machines 2, 3, 4 take the average of the two estimates they see and each report this average to you.

The following simple example illustrates this. We again assume that the value of X is 150. Let's assume for this example that the estimates of the four Estimation Devices are as follows:

- Estimation Device A: 81
- Estimation Device B: 127
- Estimation Device C: 209
- Estimation Device D: 176

Communication Machine 1 would then report the estimate of Estimation Device A (i.e., 81). Communication Machines 2, 3 and 4 would take the average of the two estimates they see, as described above. Communication Machines 1, 2, 3 and 4 would thus report the following estimates:

- Communication Machine 1: 81
- Communication Machine 2: 104 (104 is the average of 81 and 127)
- Communication Machine 3: 145 (145 is the average of 81 and 209)
- Communication Machine 4: 128.5 (128.5 is the average of 81 and 176).

Thus, for this estimation task, you would see the following information on your computer screen:

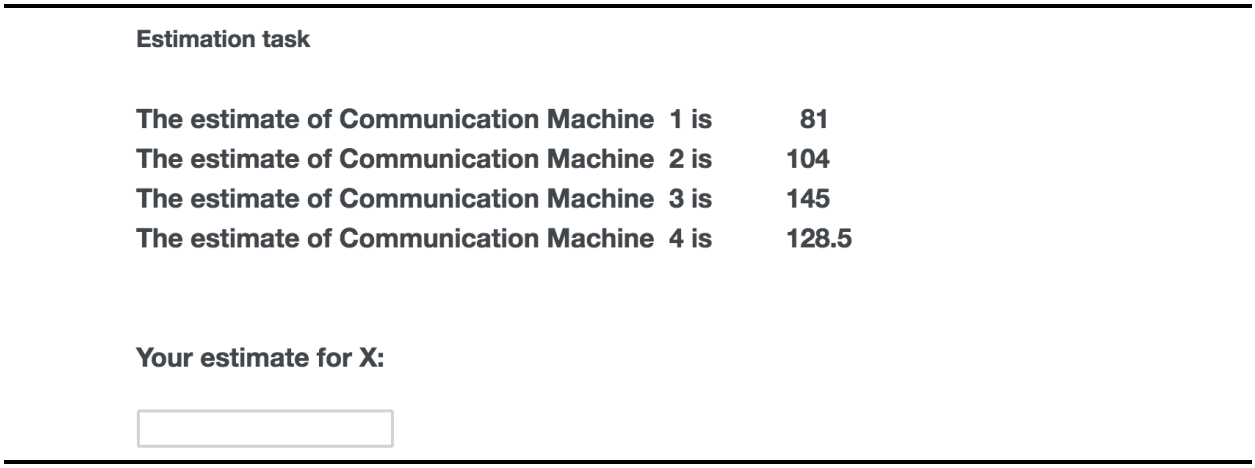


FIGURE 4. Screen shot example.

SUMMARY

- In this Part, you will face 3 different estimations tasks (1 in each Round).
- Your earnings depend on how precise your answer in the estimation task was, i.e., how close it was to the correct value. Only one of these 3 Rounds matter for your earnings. Since each of the 3 estimation tasks are potentially payoff-relevant, you should answer carefully in all 3 Rounds.
- For each estimation task, you will be provided with computer-generated information regarding the number X you have to estimate.
- The structure of the information will be as follows: In each Round, each of the four Estimation Devices provide estimates that come from a random draw from a normal distribution with mean X. These devices are always equally good at estimating X.
- In addition, there are four Communication Machines which process the estimates they observe from the Estimation Devices and then report an estimate derived from these estimates.
- Please take a look again at Figure 3, which we've reproduced below. There you can see which estimates the respective Communication Machines observe, how they are processed and how they map into the estimate which the respective Communication Machine report.
- For each estimation task, you receive the following information: You will see the estimates from Communication Machines 1, 2, 3 and 4.
- After observing this information, you will then be given a maximum of 5 minutes to think about your own estimate and then enter it into your computer. You can enter any number (positive or negative) for your estimate of X.

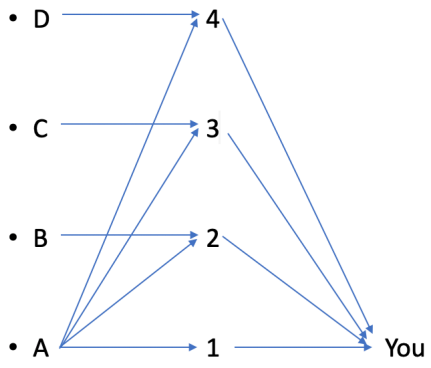


FIGURE 3

INFORMED CONSENT FORM FOR SOCIAL SCIENCE RESEARCH
The Pennsylvania State University

Title of Project: Decision-Making.

Principal Investigator: Chloe Tergiman, 332A Business Bldg
University Park, PA 16802
(814)863-4372; chloejt@gmail.com

1. Purpose of the Study: The purpose of this research study is to study decision-making.

2. Procedures to be followed: Subjects will be tasked to read instructions and make decisions at a computer terminal.

3. Discomforts and Risks: There are no risks in participating in this research beyond those experienced in everyday life.

4. Benefits: Subjects may develop economic decision-making skills and earn monetary pay-off

5. Duration: The study will last 1 hour.

6. Statement of Confidentiality: Your participation in this research is confidential. No identify information is stored in the data file. If this research is published, no information that would identify you will be written. The following may review and copy records related to this research: The Office of Human Research Protections in the U.S. Department of Health and Human Services, the Social Science Institutional Review Board and the PSU Office for Research Protections.

7. Right to Ask Questions: Please contact Chloe Tergiman at (814)863-4372 with questions, complaints or concerns about the research. You can also call this number if you feel this study has harmed you. If you have questions about your rights as a research participant, contact Penn State's Office for Research Protections at (814) 865-1775.

8. Compensation: Participants will receive $7 plus whatever you earn in the experiments you participate in. No other compensation is provided. You must stay until the end to receive your earnings beyond the $7 show up fee.

9. Voluntary Participation: You do not have to participate in this research. You can end your participation at any time by telling the person in charge. Refusal to take part in or withdrawing from this study will involve no penalty or loss of benefits you would receive otherwise.

10. The aggregate data collected in this experiment will be used in future academic publications.

Completion and submission of the survey is considered your implied consent to participate in this study. Please keep this form for your records.

FIGURE 4. Consent form

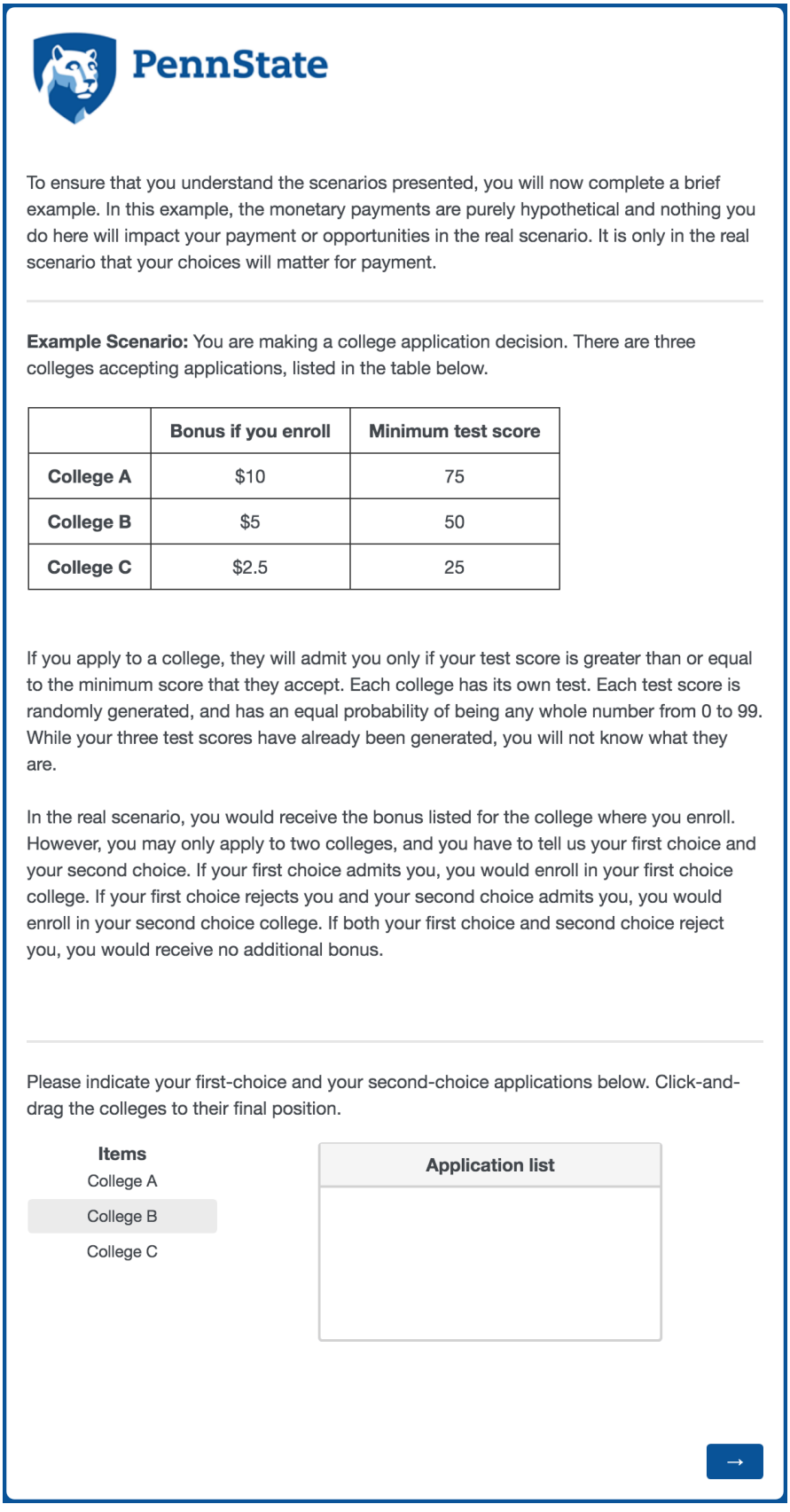


FIGURE 5. Practice scenario in uncorrelated arm

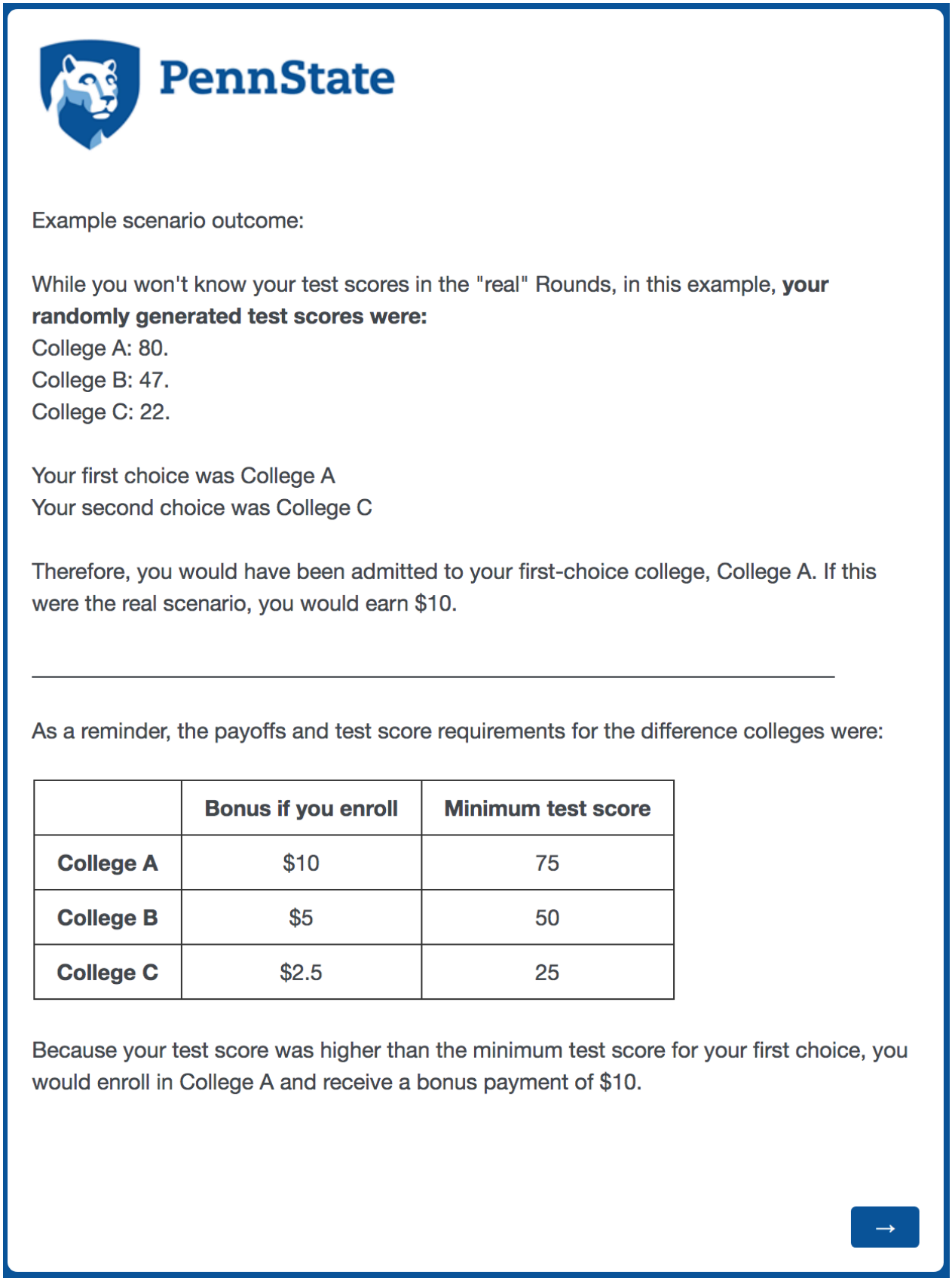


FIGURE 6. Results page for practice scenario in uncorrelated arm.

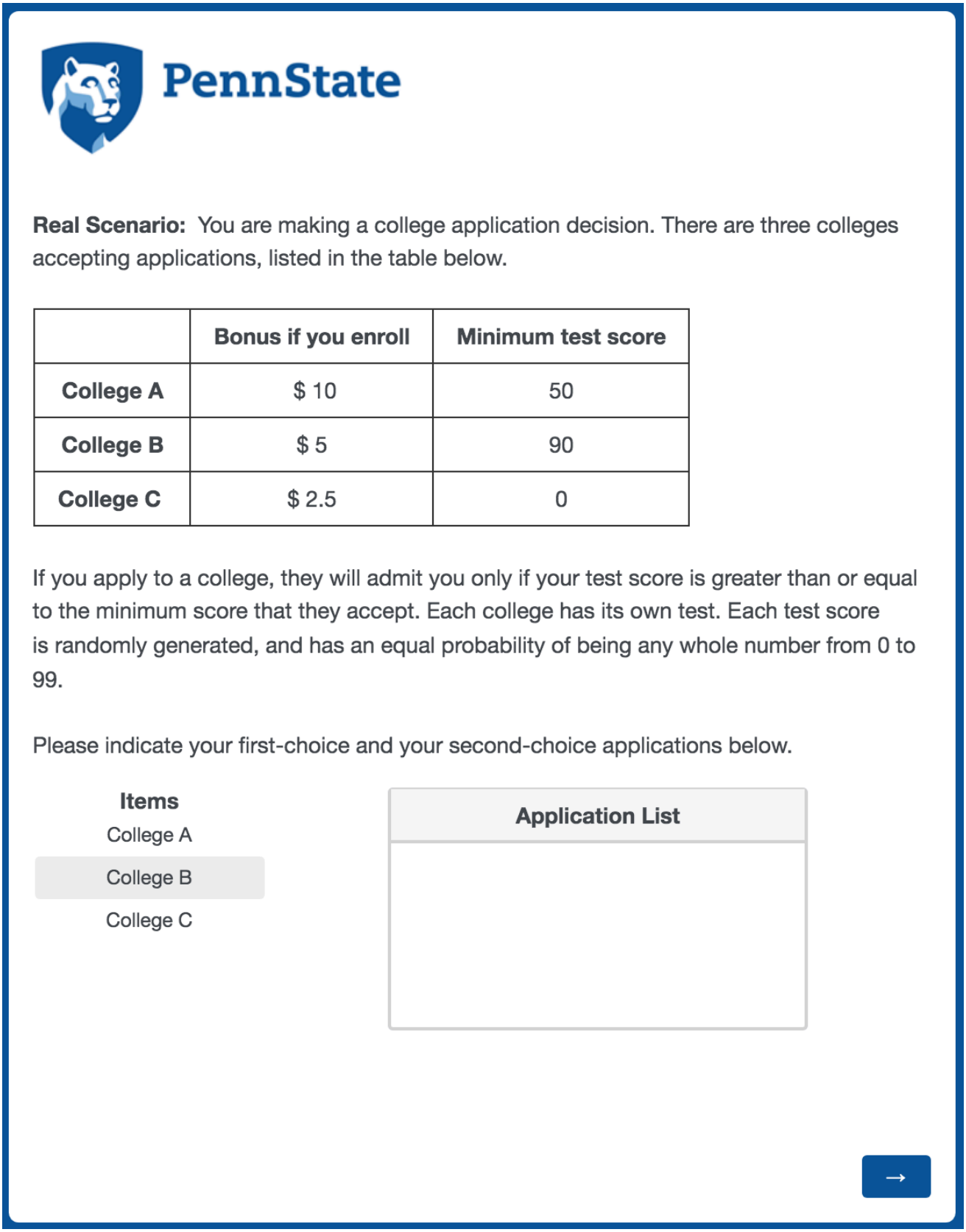


FIGURE 7. Example ("real") scenario in uncorrelated arm.

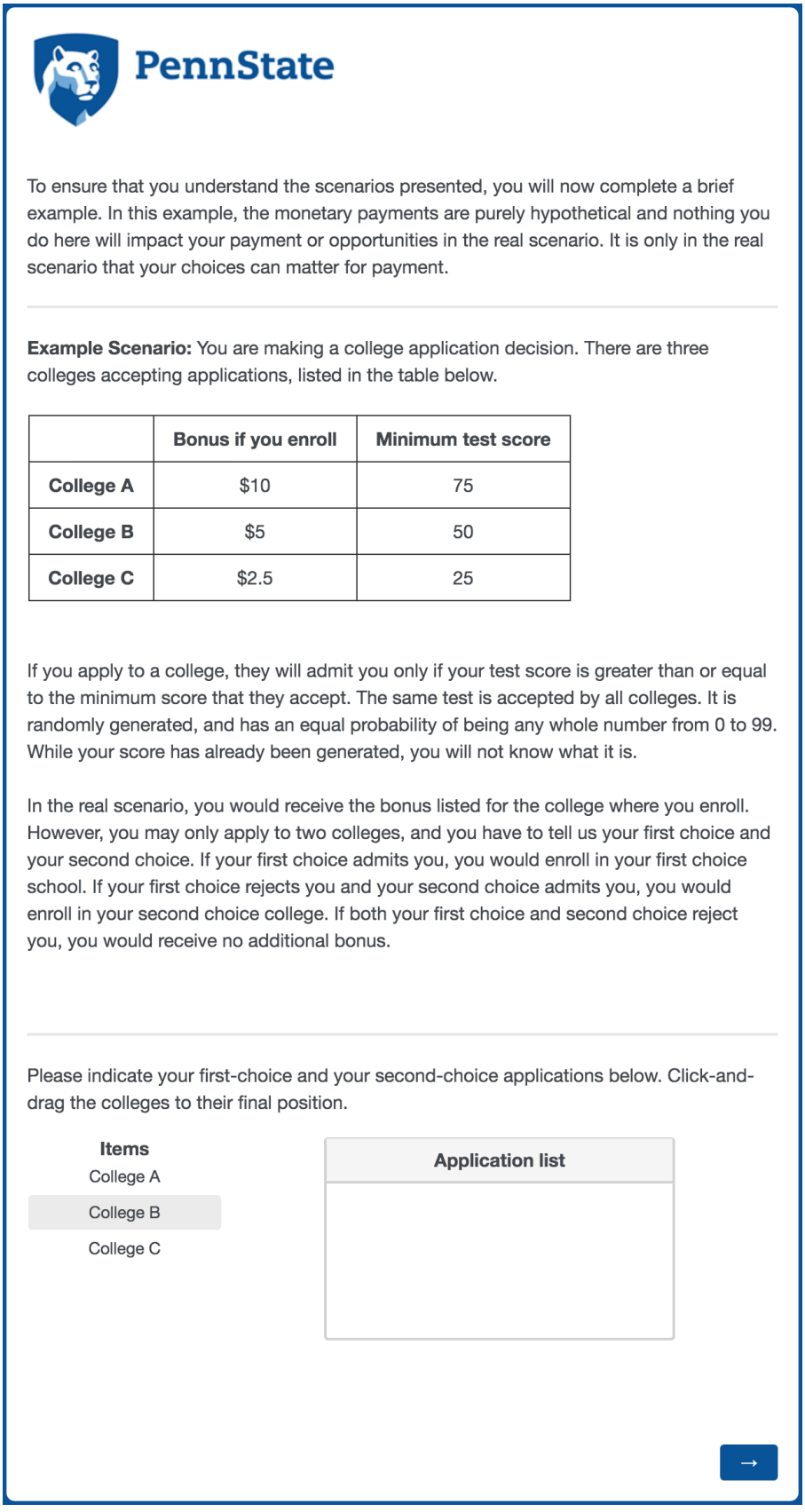


FIGURE 8. Practice scenario in correlated arm.

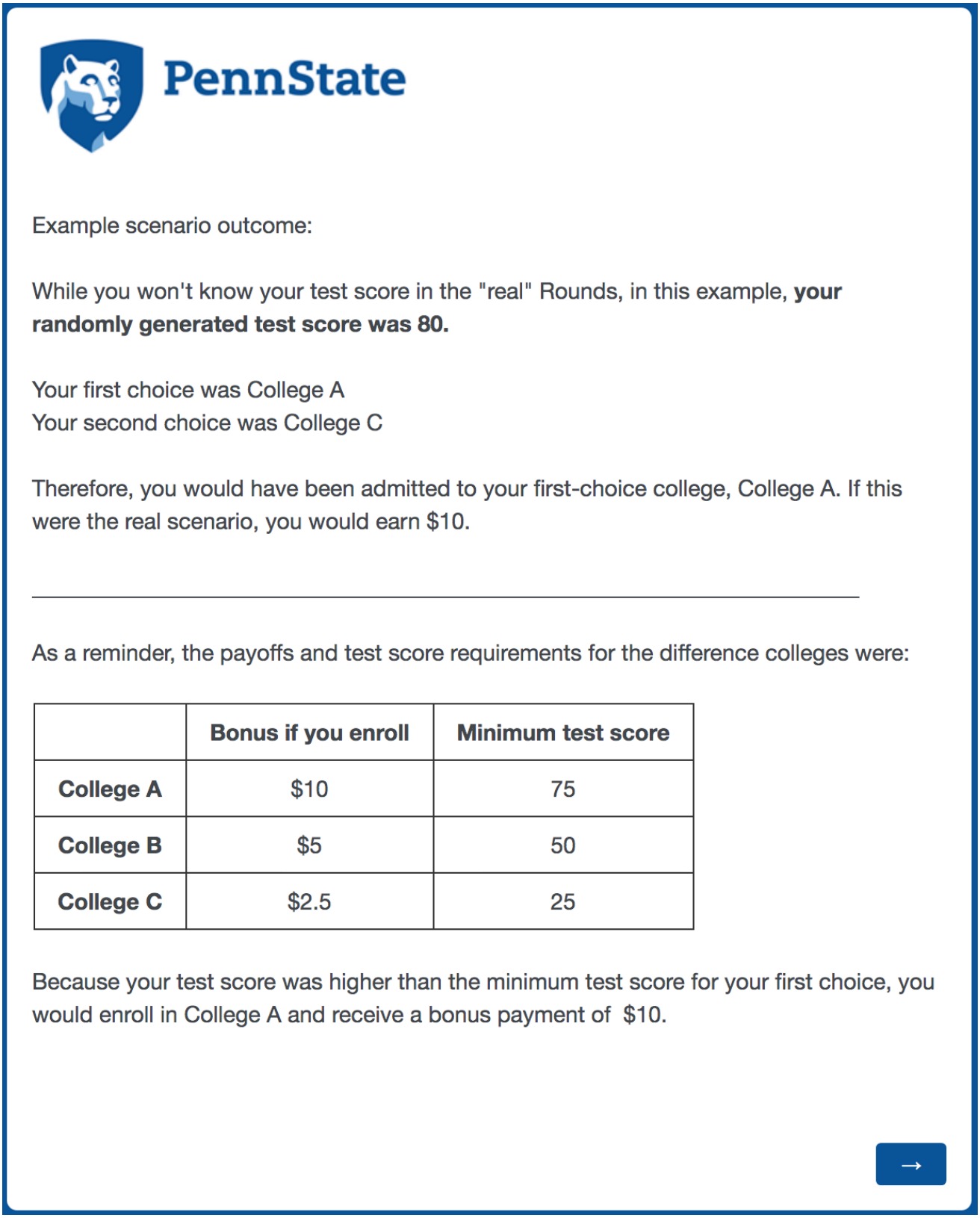


FIGURE 9. Results page for practice scenario in correlated arm.

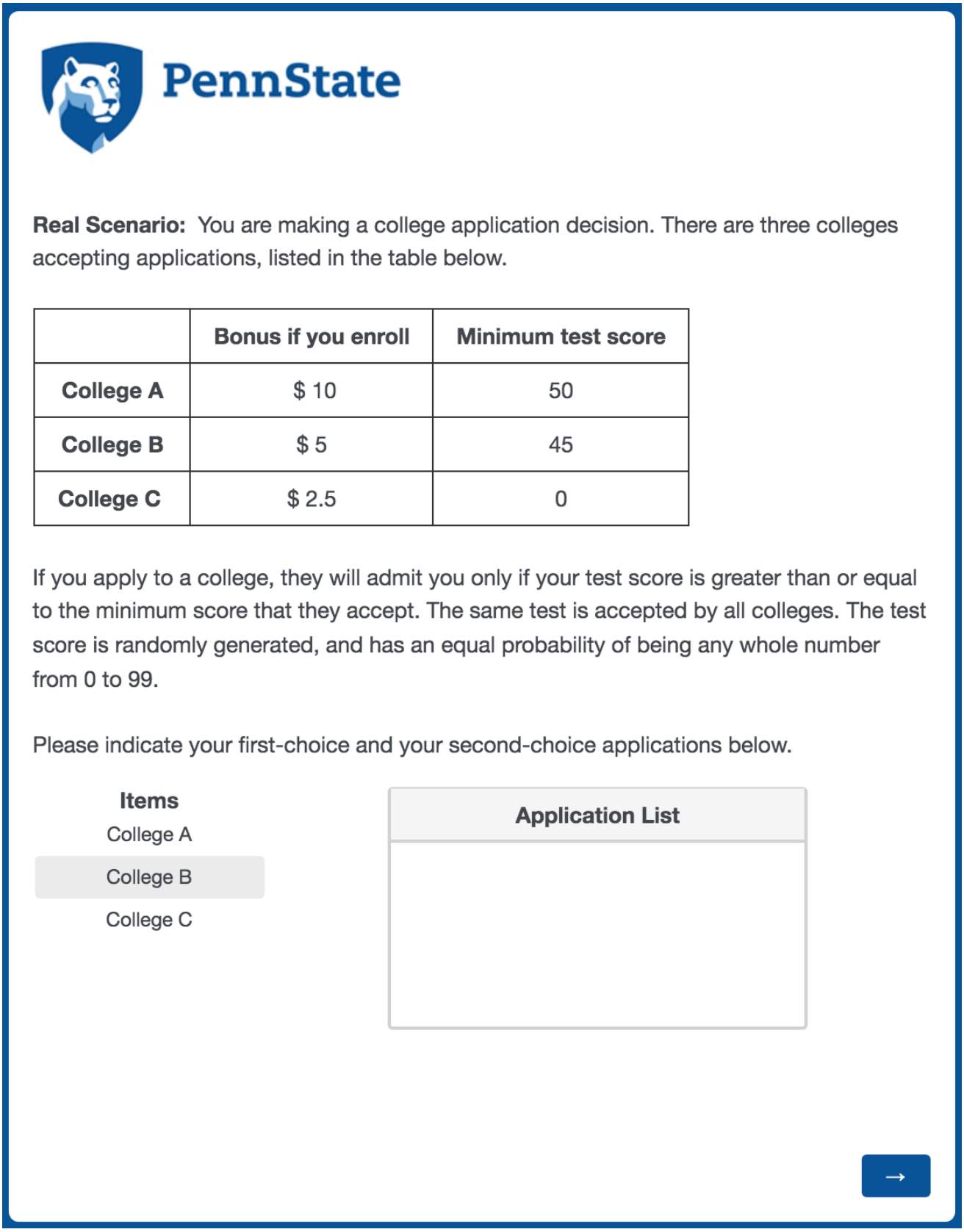


FIGURE 10. Example ("real") scenario in correlated arm.

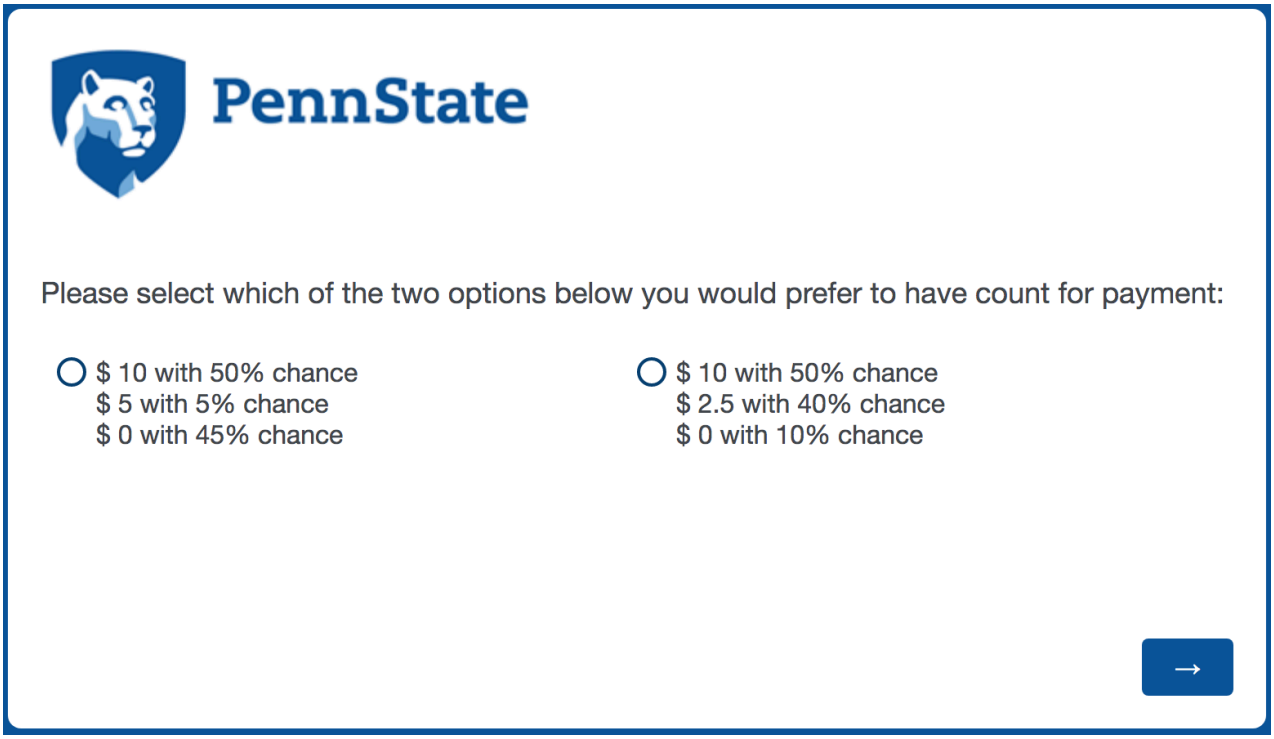
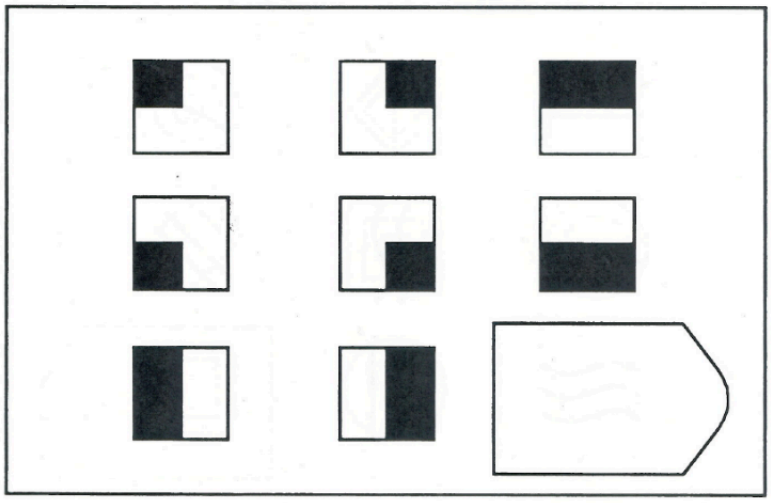
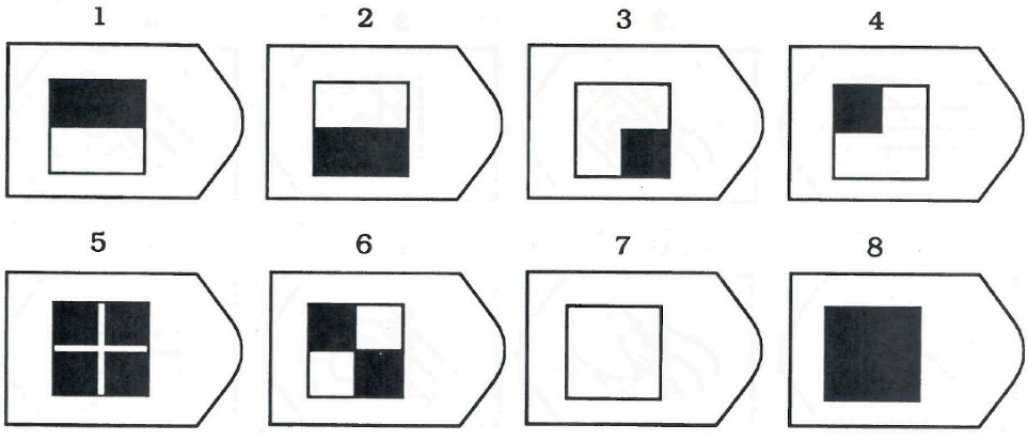


FIGURE 11. Example of matching risky gamble





Select your answer to the question above:

Choice 1 Choice 2 Choice 3 Choice 4 Choice 5 Choice 6 Choice 7 Choice 8

FIGURE 12. Example of Raven question

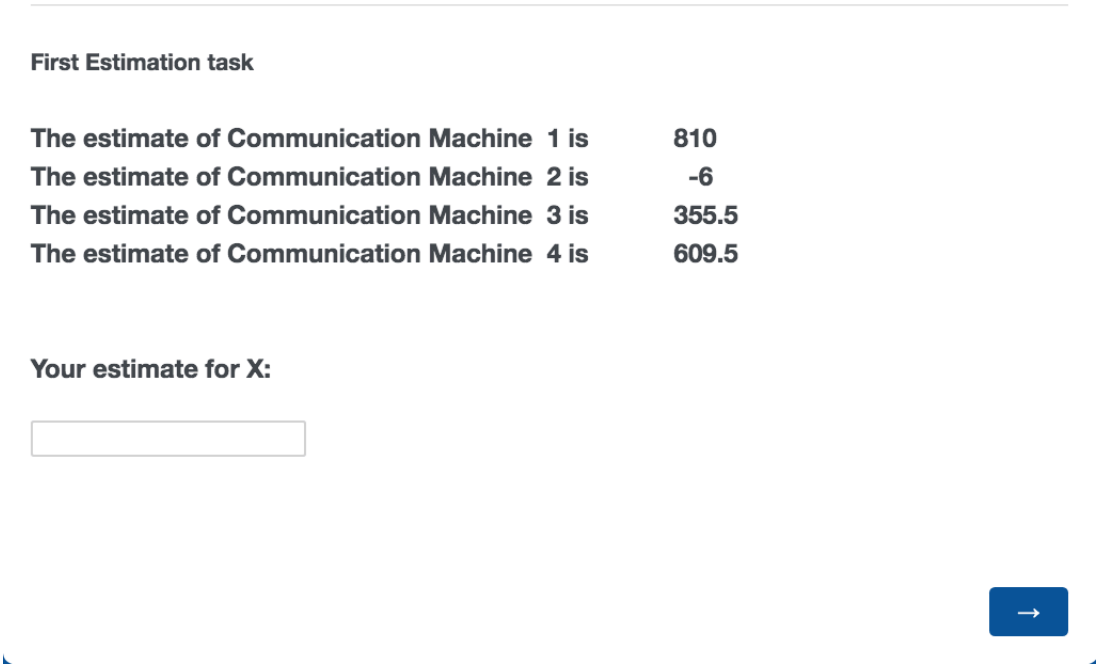


FIGURE 13. Example of Enke-Zimmermann correlation neglect task.

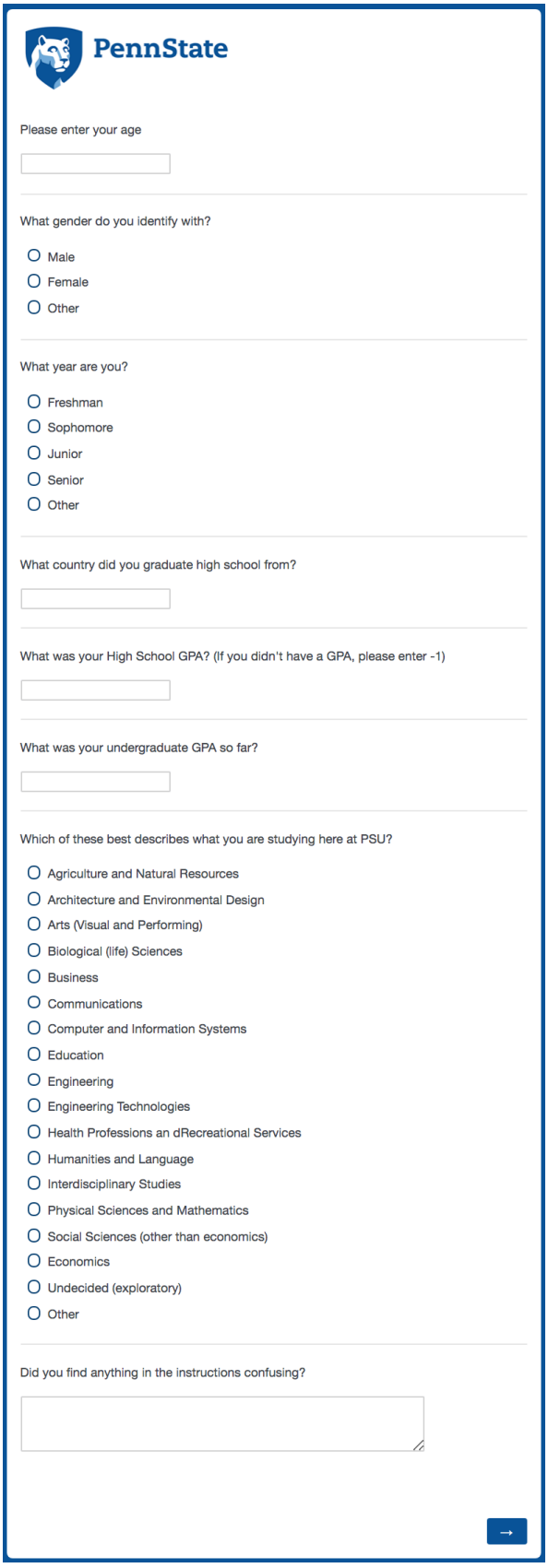


FIGURE 14. Demographic questionnaire.

Correlation Neglect in Student-to-School Matching (#18666)

Created: 01/17/2019 06:35 AM (PT)
Public: 10/17/2019 12:39 PM (PT)

Author(s)
Alex Rees-Jones (Wharton) - alre@wharton.upenn.edu
Ran Shorrer (Penn State University) - shorrer@psu.edu
Chloe Tergiman (Penn State University) - cjt16@psu.edu

1) Have any data been collected for this study already?
No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?
We are testing if students make more mistakes in college application decisions when schools' rankings of students are correlated vs. uncorrelated.

3) Describe the key dependent variable(s) specifying how they will be measured.
We present subjects with a series of college application scenarios. In each scenario, there are three different colleges where they could potentially enroll, each offering a different bonus payment and requiring different minimum test scores. The subject must choose two of the three to apply to, and rank order them. Call the highest paying college A, the second-highest B, and the lowest C.
Our DV is a categorical variable recording the application list submitted, indicating one of the six options: A>B, A>C, B>A, B>C, C>A, or C>B.

4) How many and which conditions will participants be assigned to?
There are two versions of the scenario set-up: one with a single test-score used by all schools (i.e., perfectly correlated evaluations) and one where each school has its own independent test score (i.e., perfectly uncorrelated evaluations). We have nine scenarios presented in each version, preserving the correlation structure but changing the test score requirements across schools. The nine versions are designed as matched pairs, so that for each correlated scenario there is a matching uncorrelated scenario that leads to the same distribution over final outcomes when the application lists of interest – A>B and A>C – are chosen.
Each subject will face both versions of the scenario set-up, allowing a within-subject comparison of responses. We randomly assign which appears first, allowing for a between-subjects comparison looking just at responses to the first correlation structure.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
We will conduct a Fisher's exact test to test for a different distribution of our DV by correlation condition. We will additionally conduct difference-of-proportions tests of the difference in the fraction of subjects submitting the application list A>B and A>C by correlation condition. We hypothesize that the list A>B will be more common under perfectly correlated evaluation. We will conduct these analyses for each of the nine matched pairs, as well as on the pooled preference data.
We will conduct this analysis in two ways: (1) using only the first scenario faced by each subject, and (2) using both scenarios from each subject. To the extent that results differ, we favor analysis (1) as the cleanest between-subject comparison.
We additionally have a battery of nine gamble choices, which each present a pure monetary gamble that is equivalent to the A>B vs A>C choice for the matched-pair scenarios. We will use these responses as a benchmark when evaluating whether responses to the correlated or uncorrelated scenario better reflects subjects' informed preferences. We predict that responses in the uncorrelated scenario will be closer to informed preference.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
We will exclude subjects who have documented experimental non-compliance in the lab, or who leave the experiment before it is completed.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.
We will run experimental sessions until we have 80 responses in each of the "correlated scenario first" and "uncorrelated scenario first" treatment arms.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)
Secondary Analyses:
We will examine the two-way tabulation of subjects' responses to both versions of the scenario. Since material incentives are the same in each scenario, we are interested in the rate at which subjects provide different responses in the two versions. We are most interested in comparing the rate at which subjects list A>C in one condition and A>B in the other. We predict that A>B will be more common in the correlated condition.
We will additionally construct a within-subject measure of susceptibility to our mistake of interest by generating a within-subject count of the number of matched-pair scenarios where the subject listed A>B in the correlated scenario but A>C in the uncorrelated scenario. We will regress this measure on our experimental elicitation of performance on Raven's matrices and on a measure of correlation neglect generated from a version of Enke and Zimmerman's (EZ's) experimental design. The Raven's measure is simply the raw count of correct answers to Raven's matrices. Our EZ measure follows EZ's approach: we present three forecasting questions that each allow the calculation of the EZ chi parameter, and we take the subject's median calculated chi as their measure of susceptibility to correlation neglect.
After examining the association between our measure, Raven's performance, and EZ's correlation neglect measure, we will rerun the analyses including a

battery of demographic covariates: dummy variables for gender, having an English-speaking country of origin, and having a major that requires mathematics, as well as continuous measures of high-school and college GPA.