ソシオネットワーク戦略ディスカッションペーパーシリーズ ISSN 2434-9445
第 86 号　2020 年 6 月
RISS Discussion Paper Series
No.86　June, 2020

# Conducting Economic Experiments at Multiple Sites: Subjects' Cognitive Ability and Attribute Information

Kazuhito Ogawa, Yusuke Osaki, Tetsuya Kawamura

Hiromasa Takahashi, Satoshi Taguchi, Yoichiro Fujii, Naoki Watanabe

文部科学大臣認定　共同利用・共同研究拠点

関西大学ソシオネットワーク戦略研究機構

Research Institute for Socionetwork Strategies,
Kansai University
Joint Usage / Research Center, MEXT, Japan
Suita, Osaka, 564-8680, Japan
URL: http://www.kansai-u.ac.jp/riss/index.html
e-mail: riss@ml.kandai.jp
tel. 06-6368-1228
fax. 06-6330-3304

# Conducting Economic Experiments at Multiple Sites: Subjects' Cognitive Ability Score as a Covariate*


Kazuhito Ogawa† Yusuke Osaki‡ Tetsuya Kawamura§
Hiromasa Takahashi¶ Satoshi Taguchi‖ Yoichiro Fujii** Naoki Watanabe††


January 28, 2023


**Abstract**

This paper clarifies the conditions under which subjects' behavior in a bandit experiment in the context of weighted voting can be homogenized across different subject pools in Japanese universities by covariate adjustment with their cognitive ability scores. Across experimental sites located in the same region, it would not be difficult to obtain the homogeneity of subjects' behavior with their cognitive ability scores in addition to their general attribute information. However, it would be difficult to obtain the homogeneity by covariate adjustment among the experimental sites located in different regions. Thus, subjects' cognitive ability score is an important covariate in integration of data collected at multiple experimental sites or in comparison of experimental results obtained at different experimental sites, and if necessary, their regional background may serve as a blocking factor for randomized block design of economic experiments.



**Keywords**: multiple-site experiment, cognitive ability, bandit experiment, weighted voting, covariate adjustment, data integration

**JEL Classification**: C91, D72, D83


---


*This paper is the first half of the Kansai University RISS Discussion Paper Series No. 86, and it was presented at the JEA annual meeting in October 2020 and the International Workshop for Lab and Field Experiments in March 2021. The authors wish to thank Eizo Akiyama, Rei Goto, Ryuichiro Ishikawa, Toshiji Kawagoe, Tatsuhiro Shichijo, Kan Takeuchi, Takeshi Nishimura, Naoko Nishimura, Nobuyuki Hanaki, Yasuyo Hamaguchi, Yoichi Hizen, and Takuma Wakayama for their comments on this research.



†Kansai University, Faculty of Sociology, Suita, Japan.

‡Waseda University, School of Commerce, Tokyo, Japan.

§Tezukayama University, Faculty of Economics and Business Management, Nara, Japan.

¶Hiroshima City University, Faculty of International Studies, Hiroshima, Japan.

‖Doshisha University, Faculty of Commerce, Kyoto, Japan.

**Meiji University, School of Commerce, Tokyo, Japan.

††(corresponding author) Keio University, Graduate School of Business Administration, 4-1-1 Hiyoshi Kohoku, Yokohama, Kanagawa 223-8526, Japan. Phone: +81-45-564-2039, E-mail: naoki50@keio.jp, ORCID 0000-0002-6140-2762.

# 1 Introduction

For conducting an experiment with common specifications at multiple sites, Snowberg and Yariv (2021) showed that the student population offers bounds on their behaviors in comparison to those observed in a representative sample of the U.S. population and participants in Amazon Mechanical Turk.[1] This paper aims to further clarify the conditions under which we can homogenize subjects' behavior in a bandit experiment across subject pools of undergraduate students at Japanese universities. We shed light on the effectiveness of subject's cognitive ability score as a potential covariate to obtain the homogeneity, which is desirable in integration of data collected at multiple experimental sites.

## 1.1 Background of This Research

Over the past two decades, technological advances and methodological developments have made it possible to conduct experiments in which more subjects can participate than before. Along with this potential, the number of subjects required for an experimental study has been increasing remarkably in recent years.[2] In laboratory experiments, about 20 years ago, the numbers of subjects in papers published in major international journals were often fewer than 100. By 2015, however, it was difficult to find experimental studies that did not involve at least 300 subjects, and recently, many papers have been published containing data collected from more than 500 subjects.[3] Nonparametric tests which are often used for data analysis in economic experiments need large number of samples to obtain clear results. This point may also be a factor in the increase in the number of subjects per paper.[4]

[1]Pioneering research in multiple-site experiments includes Roth et al. (1991), Cason, et al. (2002), and Henrich et al. (2004), which provide precautions to be taken into account for the international comparison of experimental results. Falk et al. (2018) is a recent literature we mention at the end of Section 4.

[2]During this period, the development of subject management tools, such as ORSEE (Greiner, 2003), has made it possible to recruit subjects more efficiently, and the spread of softwares for economic experiments, such as zTree (Fischbacher, 2007) and oTree (Chen et al., 2016), has made it easier to conduct economic experiments even with limited programming skills. The Journal of Economic Behavior and Organization edited a special issue (Rosser and Eckel, eds., 2010) on experimental methodologies, noting, e.g., the necessity of not leaking information that would hinder the random recruitment of subjects at each site. Azrieli et al. (2018) discussed the consistency of experimental objectives and payment schemes for subjects.

[3]This point was mentioned in an invited lecture given by Kan Takeuchi at the Kyoto University Economic Laboratory Opening Conference in 2013. His talk motivated the authors to start this research.

[4]Some nonparametric tests can calculate more accurate p-values than before even in the case of small samples; the permutation test and the permuted Brunner-Munzel test (Ernst, 2004; Neubert and Brunner, 2007), e.g., are generally based on Fisher (1935) in regard to the idea of reordering, but reordering seemed impractical until 20 years ago because of the huge amount of computations required.

With this increasing tendency in the number of subjects per paper, some researchers recently face difficulties in collecting a sufficient amount of data for each experiment. In Japan, for instance, public universities and colleges have much fewer students than do large private universities, and thus it is becoming increasingly difficult to recruit a sufficient number of college students as subjects for not only research, but also education. In experiments under these recent circumstances, the sessions with common specifications are sometimes conducted at multiple sites and the data collected at each site are merged according to certain criteria experimenters adopt.

In experimental economics, however, no standardized methods regarding how to analyze the data collected at multiple sites have been established up until the present day. When subjects' characteristics differ at different experimental sites, some values will be commonly observed but others may be missing from the data on latent subject attributes. In such a case, kernel matching (Heckman et al., 1998) is, truly, one of the simplest methods for data integration. In kernel matching, however, the data are integrated with large weights for a limited number of important covariates that are related to subject's decision-making. We should thus clarify what covariates are effective under certain conditions. If we find it difficult to merge the data for some clear reason or evidence, then we need to know appropriate blocking factors for a randomized block design of experiments prior to conducting the sessions at multiple sites. (See, e.g., Hinkelmann and Kempthone (2005).)

Online experiments, the number of those is increasing currently, truly mitigate some difficulties in recruiting a sufficient number of subjects; the location for conducting online sessions are not limited to particular physical sites.[5] It would be, however, important to guarantee the homogeneity of behavior of subjects who participate in each online sessions, because, as implied in Snowberg and Yariv (2021), participants in online experiments have various backgrounds that affect their decision-making. Therefore, it is a methodological problem we should consider to find appropriate indicators of subjects' characteristics that affect their behavior or decision making at each experimental site and share the information among researchers, regardless of whether the experiments are conducted onsite or online.

[5] Arechar et al. (2018) conducted an online experiment on public-goods games and reported that few subjects withdrew from participation in sessions for the experiment. As a software, zTree unleashed (Duch et al., 2020) was recently released for online experiments.

## 1.2 Research Questions

As a potential covariate in data integration, this paper examines the effectiveness of subjects' cognitive ability score on the Raven's test (Raven scores) which measures subjects' ability for visual pattern recognition; some economists recently reported differences in experimental results due to differences in subjects' Raven scores, as noted later in subsection 2.3.[6] We follow the specifications of an experiment conducted by Guerci et al. (2017), because the experiment was designed for investigating the subjects' pattern recognition (learning) of their payoffs in individual choice problems without strategic interaction among others' choice, the specifications of which thus fit the purpose of our examination.

In the experiment, subjects are provided with binary choice problems and asked to choose one of two weighted voting games many times, and they are given their payoffs which are stochastically generated for each choice according to a voting theory. The payoff-generating function remains the same but hidden from subjects throughout the experiment. There is a correct choice for the binary choice problems in the sense that it generates higher expected payoff. This type of study is called a bandit experiment. We compare the rates of correct choices made by subjects to the problems observed at different sites.

It is known that the underlying structure of weighted voting is so complicated that it is difficult for many people to precisely recognize it.[7] We thus expect that the average rates of correct answers are different at different experimental sites, according to the differences of the average scores of the Raven's test at different experimental sites. Does covariate adjustment with subjects' cognitive ability scores homogenize their choice behavior across different subject pools, when there are differences in those average rates of correct answers subjects among different experimental sites? If the answer is negative, what is then a potential blocking factor we should take into account in designing an experiment that will be conducted at multiple sites? These are research questions we investigate in this paper. Once data are taken, the negative answer to the first question implies that it is difficult to obtain unbiased estimates with those data.

[6]Dohmen et al. (2010) used the Wechsler Adult Intelligence Scale (WAIS) as a measure of cognitive ability and considered the relationship of subjects' risk aversion and impatience with their cognitive ability. Hanaki (2020) also considered the relationship between subjects' cognitive ability and their behavior.

[7]Felsenthal and Machover (1998, pp.164-165) noted that it might have been difficult even for the policy makers and officials who designed and re-designed the system to see the underlying relationship between the nominal voting weights and the actual voting powers.

### 1.3 Outline and Results

Our experiment was conducted at Kansai University (Kansai U-A, U-B), Osaka Sangyo University (Osaka SU), Doshisha University (Doshisha), and Hiroshima City University (Hiroshima CU). In our analysis, the data taken at Kansai U-A and Osaka SU are compared, while those obtained at Kansai U-B, Doshisha, and Hiroshima CU are compared in pairs, according to how the instructions were provided to subjects and how the reward for their participation was determined. The subjects in Kansai U-A and U-B were recruited from the same subject pool at Kansai University. Between subject pools at Kansai university and Osaka SU, there are significant differences in not only the participants' score on the Raven's test but also their gender ratio. No science and engineering students participated in the subject pool at Doshisha, and Hiroshima CU does not have the department of economics. Hiroshima CU is located in a different region from the region in which Kansai University, Osaka SU, and Doshisha are located.

Each session had 40 periods for the first part and 20 periods for the second part. We compared the rates of correct answers to the binary choice problems observed in 6 blocks of 10 periods. To obtain as unbiased a comparison as possible, we conducted propensity score matching by calculating the probability of assignment to the two experimental sites in terms of subjects' scores on the Raven's test and attribute information. Subjects' attribute information refers to their age, gender, whether they were economics majors, and whether they are science or engineering majors. Note that we do not use dichotomous logistic regression with dummy variables, because our question is not to estimate fitted values but to confirm the homogeneity of subjects' behavior.

Our main results are as follows: (1) Between Kansai U-A and Osaka SU, many significant differences were detected in the rates of correct answers to the binary choice problems when controlling only for subjects' attribute information, but those differences disappeared when adding their scores on the Raven's test to those variables. (2) Between Kansai U-B and Doshisha, there were some significant differences in the rates of correct answers but many of them disappeared by adding subjects' Raven scores to their attribute information and dropping the data taken from science or engineering majors at Kansai U-B. (3) Between Kansai U-B and Hiroshima CU, significant differences in the rates of correct answers did not disappear even by adding subjects' Raven scores to their attribute information and dropping the data taken from economics majors at Kansai U-B, and between Doshisha and Hiroshima CU, significant differences in those rates did not disappear by adding subjects' Raven scores without dropping data taken from science or engineering majors at Hiroshima City University and economics majors at Doshisha University.

In summary, across experimental sites located in the same region, subjects' behavior could be homogenized by covariate adjustment with their Raven scores, whereas it would be difficult to obtain the homogeneity among experimental sites located in different region. According to these results, subjects' cognitive ability score would be an important covariate in integration of data collected at multiple experimental sites or in comparison of experimental results obtained at different sites, and if necessary as mentioned in subsection 1.1, their regional background might serve as a blocking factor for randomized block design of experiments, where subjects are divided into homogeneous blocks and treatments are randomly assigned to the blocks.

The remainder of the paper is organized as follows. Section 2 describes the experimental design. We first explain the binary choice problems examined in this experiment, next show how the session proceeds on subjects' monitor, and note the cognitive ability test. Section 3 describes the session details for each experimental site. Section 4 describes the experimental results, after explaining propensity score matching. Section 5 summarizes this paper and discusses some points for future research.

# 2 Experimental Design

## 2.1 Binary Choice Problem

This experiment has three treatments, which are explained at the end of this subsection. A treatment involves sessions, each consisting of 60 periods. In each period, subjects are asked to choose one of two four-member committees (weighted voting games) that will divide 120 points among the members. Let $N = \{1, 2, 3, 4\}$ be the set of the members (players) to join. A committee is represented by $[q; v_1, v_2, v_3, v_4]$, where $v_i$ is the number of votes (voting weight) allocated to member $i \in N$ and $q$ is the minimum number of votes required for an allocation to be adopted (quota). Every subject acts as Member 1, who has $v_1$.

The two committees from which the subjects are asked to chose one have the same total number of votes, the same quota, and the same number of votes for Member 1. In each session, the subjects face one binary choice problem for the first 40 periods, and in the following 20 periods they face a similar but different binary choice problem. For example, in the first 40 periods, subjects face a choice between $[14; 5, 3, 7, 7]$ and $[14; 5, 4, 6, 7]$, while in the following 20 periods, they face a choice between $[6; 1, 2, 3, 4]$ and $[6; 1, 1, 4, 4]$. The subjects are not asked to play the weighted voting games that they choose.[8] Before the start of each session, the subjects are clearly informed that the other three members of the committees are all fictitious. The experiment is thus regarded as a two-armed bandit experiment with contextual information on voting situations.

The payoff each subject obtains from his or her choice is generated by a function based on a voting power index called the Deegan-Packel index (DPI) (Deegan and Packel, 1978). Given a weighted voting game, a non-empty subset $S$ of $N$ is called a coalition, and a coalition is called a winning coalition if $\sum_{i \in S} v_i \geq q$; otherwise, it is called a losing coalition. A minimum winning coalition (MWC) is a winning coalition such that deviation by any member of the coalition alters its status from winning to losing. In the experiment, for each period, one MWC is drawn with equal probability from all possible MWCs for the committee that the subject chooses. If the subject is a member of the drawn MWC, then he or she receives an equal share of the total payoff with the other members.

We here denote each MWC by the votes apportioned to its members; e.g., the MWCs in a committee $[14; 5, 3, 7, 7]$ are written as $(5, 3, 7)$, $(5, 3, 7)$, and $(7, 7)$. Member 1 belongs to two MWCs out of three, each having three members. The DPI of Member 1 is thus $2/3 \times 1/3 = 2/9$ and the expected payoff for Member 1 is $120 \times 2/3 \times 1/3$. Subjects are simply told that payoffs are determined based on a voting theory. Note that we use binary choice problems in which the better committees for the subjects are the same regardless of whether we employ DPI or other power indices such as the Banzhaf index (Banzhaf, 1965) and the Shapley-Shubik index (Shapley and Shubik, 1954).[9] Thus, this experiment does not intend to verify whether subjects learn the DPI as a payoff-generating function.

[8]The point is to avoid the complexities mentioned at the beginning of the Introduction; subjects are playing a weighted voting game with the other subjects, while simultaneously learning about the underlying relationship between their nominal voting weights and expected payoffs.

[9]Montero et al. (2008) Aleskerov et al. (2009), Esposito et al. (2012), Guerci et al. (2014), and Watanabe (2014) reported that the most frequently observed winning coalitions were MWCs in their experiments.

The binary choice problems we use are shown in Table 1.[10] Subjects are faced with one of the following sequences of binary choice problems: $A \to B$, $B \to A$, $C \to D$, or $D \to C$ (the order is indicated by the arrow), where the first problem is used in the first 40 periods, and the second in the subsequent 20 periods. The committee that generates a higher expected payoff for subjects (correct option) is Choice 2 for all problems listed there.[11] Table 1 also lists the payoff vector for each choice in each binary choice problem. For any choice, subjects obtain 40 points, if a MWC to which Member 1 belongs is selected; otherwise nothing. Subjects are not informed of what binary choice problems being given before those problems are shown on their monitors; they are not shown the list of binary choice problems in the instructions for this experiment.

Table 1: Binary choice problems, MWCs, and payoff vectors.

| | | | | |
|---|---|---|---|---|
| Problem A | Choice 1 | [14; 5, 3, 7, 7] | Choice 2 | [14; 5, 4, 6, 7] |
| | $(7_1, 7_2)$ | (0, 0, 60, 60) | (5, 4, 6) | (40, 40, 40, 0) |
| | $(5, 3, 7_1)$ | (40, 40, 40, 0) | (5, 4, 7) | (40, 40, 0, 40) |
| | $(5, 3, 7_2)$ | (40, 40, 0, 40) | (5, 6, 7) | (40, 0, 40, 40) |
| | | | (4, 6, 7) | (0, 40, 40, 40) |
| Problem B | Choice 1 | [6; 1, 2, 3, 4] | Choice 2 | [6; 1, 1, 4, 4] |
| | (2, 4) | (0, 60, 0, 60) | $(4_1, 4_2)$ | (0, 0, 60, 60) |
| | (3, 4) | (0, 0, 60, 60) | $(1_1, 1_2, 4_1)$ | (40, 40, 40, 0) |
| | (1, 2, 3) | (40 40, 40, 0) | $(1_1, 1_2, 4_2)$ | (40, 40, 0, 40) |
| Problem C | Choice 1 | [14; 3, 5, 6, 8] | Choice 2 | [14; 3, 6, 6, 7] |
| | (6, 8) | (0, 0, 60, 60) | $(6_1, 6_2)$ | (40, 40, 40, 0) |
| | (3, 5, 6) | (40, 40, 40, 0) | $(1, 2, 6_1)$ | (40, 40, 0, 40) |
| | (3, 5, 8) | (40, 40, 0, 40) | $(1, 2, 6_2)$ | (40, 0, 40, 40) |
| | | | $(6_1, 6_2, 7)$ | (0, 40, 40, 40) |
| Problem D | Choice 1 | [9; 1, 3, 5, 6] | Choice 2 | [9; 1, 2, 6, 6] |
| | (3, 6) | (0, 60, 0, 60) | $(6_1, 6_2)$ | (0,0, 60, 60) |
| | (5, 6) | (0, 0, 60, 60) | $(1, 2, 6_1)$ | (40, 40, 40, 0) |
| | (1, 3, 5) | (40, 40, 40, 0) | $(1, 2, 6_2)$ | (40, 40, 0, 40) |

Note: Subjects are all assigned to Member 1 in each committee. The MWCs in $[14; 5, 3, 7, 7]$ are, e.g., written here as $(5, 3, 7_1)$, $(5, 3, 7_2)$, and $(7_1, 7_2)$ by the votes apportioned to the members, not with references to the specific members. For each choice, MWCs are equally likely to appear.

[10] In terms of both expected payoffs and the set of MWCs in each choice, Problems A and C are identical, as are Problems B and D. However, there is a crucial difference between the two options in Problems A and C. In each of these two problems, one option has two "large" voters who can form an MWC on their own, whereas the other option does not. In Problems B and D, there is no such clear difference between the two options, as there are two large voters who can form an MWC by themselves in both.

[11] In Watanabe (2018), the correct option was changed by changing the order of the binary choice problems shown on subjects' monitors.

As noted at the beginning of this subsection, this experiment has three feedback treatments: (1) no feedback, (2) partial feedback, and (3) full feedback. For the no-feedback treatment, subjects are not informed of any payoffs they receive as the result of their committee choice until the session ends. For the partial-feedback treatment, each subject is informed of his or her own payoff in the committee he or she chose. For the full-feedback treatment, after each choice, subjects are informed of the payoffs of all four members in the committee they chose. We impose a 30-second time limit for the choice stage and a 10-second limit for the feedback stage, regardless of the amount of feedback information. If a subject does not choose a committee within 30 seconds of the choice stage, then he or she obtains zero points for that period. In this case, regardless of the treatment, in the feedback stage, the subject receives special feedback that he or she has obtained zero points for the period. This rule is clearly stated in the written instructions and explained aloud.

If a subject makes an early choice before the time limit of the choice stage, then a waiting screen is shown until all subjects in the session have made their decisions. If all the subjects make their choices within the 30-second time limit, then they all enter the feedback stage. For the no-feedback treatment, during the 10-second feedback stage, subjects are shown a screen conveying the message "Please wait until the experiment continues." For the partial-feedback and full-feedback treatments, the relevant payoff information is displayed during these 10 seconds. Subjects are prohibited from taking notes during the sessions.

This experiment is computerized by zTree (Fischbacher, 2007); during each session, the binary choice problems are provided to the subjects and their answers were chosen through a computer network. All messages instructing the subjects to wait and feedback information on their payoffs are displayed on the computer monitors. When all subjects participating in the same session have completed their choice before the 30-second time limit or the time limit in the choice stage has elapsed, the information browsing time of 10 seconds proceeds at once. For the no-feedback treatment, an indication to wait until the next alternative is presented is displayed on the subject's computer monitor, as described above. The points earned by the subjects throughout the 60 rounds are then converted to 1 JPY per point and added to the participation fee (for details, see the next subsection).

## 2.2 Subject's Monitor

Below is what subjects actually see on their monitors in sessions. In each period, subjects are provided a binary choice problem on their monitors, for instance, as follows. For each vote apportionment and the payoff distribution (in the case of the full-feedback treatment), YOU, Member 2, Member 3, and Member 4 are displayed on their monitors but they are here omitted.

**Please choose one out of the following two committees (Choice 1 or Choice 2). Each committee decides a distribution of 120 points among four members. You are Member 1. In both committees, 22 votes are apportioned to those members and you have 5 votes. Any proposals of point distributions need 14 votes in favor to be adopted.**

$$\text{Choice 1} \quad [14; \mathbf{5}, 3, 7, 7], \quad \text{Choice 2} \quad [14; \mathbf{5}, 4, 6, 7]$$

When subjects choose Choice 2 and MWC $(5, 6, 7)$ appears, they see, for instance, the following results on their monitor, regardless of any treatments.

**You chose the following committee.**

$$\text{Choice 2: } [14; \mathbf{5}, 4, 6, 7].$$

Next, in the full-feedback treatment, subjects see

**The committee decided to distribute 120 points this time as follows.**
**You obtained 40 points this time.**

$$(\mathbf{40},\ 0,\ 40,\ 40)$$

on their monitors. In the partial-feedback treatment, the payoff distribution is not shown, but rather the following note is shown on their monitors:

**You have obtained 40 points this time.**

In the no-feedback treatment, the payoff distribution is not shown and simply

**Please wait for a while.**

is shown on the subjects' monitors.

### 2.3 The Raven's Advanced Progressive Matrices Test

The Raven's test is one of the well-known tests that measure subject's cognitive ability. In each question of the test, eight patterns are drawn, and the subject selects a pattern that matches those visual patterns from the options (Carpenter et al., 1990). There are three versions of the Raven's test, Colored Progressive Matrices (CPM), Standard Progressive Matrices (SPM), and Advanced Progressive Matrices (APM), in ascending order of difficulty.

We use the APM version.[12] The APM version is composed of 48 questions in total (in Set I and Set II), and it takes 30 minutes to complete those 48 questions. In this experiment, we selected 16 of 48 questions that the subjects were asked to complete in 10 minutes after answering all of the problems in the bandit experiment. In this paper, the Raven score refers to the number of correct answers to the 16 questions on the APM version. We included questions 1, 4, 7, and 10 from Set I and questions 1, 4, 7, 10, 13, 16, 19, 22, 25, 28, 31, and 34 from Set II. These 16 questions were also used in the studies by Hanaki et al. (2016a), Hanaki et al. (2016b), Guerci et al. (2017), Watanabe (2018), Kawamura and Ogawa (2019), and Watanabe et al. (2020).[13] Different answers could be given to questions that had already been answered, when subjects were answering .

## 3 Session Details

The experiment was conducted from March 2, 2018 to October 17, 2019 at Kansai University (Senriyama campus, in Osaka), Osaka Sangyo University, Doshisha University (Imadegawa campus, in Kyoto), and Hiroshima City University.[14] The subjects were undergraduate students recruited from all over the campus at each site. There was no subject who had previously participated in this experiment, and no subject participated in sessions twice. According to syllabi at those universities, we could not find classes in which subjects could learn indices of voting power. In total, 816 subjects participated in the experiment, and the average amount paid as a reward was 2534 JPY (1 USD was about 110 JPY).

[12] Gill and Prowse (2016) and Basteck and Mantovani (2018) used the SPM version, and Proto et al. (2019) used the APM version.

[13] Proto et al. (2019) used 30 questions selected from Set II of the APM version. Set II contains 36 questions.

[14] Doshisha University has the faculty of science and engineering, but it is located at another campus which is near to Nara. There was no subjects who major in science and engineering in the subject pool at Imadegawa campus of Doshisha University.

There are two groups of experimental sites, Group A and Group B. At Kansai University, subjects were recruited for both groups but each subject was assigned to only one group; the set of those subjects are abbreviated as Kansai U-A and Kansai U-B. Hereafter, the sets of subjects who participated in this experiment at Osaka Sangyo University, Doshisha University, and Hiroshima City University are abbreviated as Osaka SU, Doshisha, and Hiroshima CU, respectively.

The same experimenter gave the instructions to the subjects of Group A (Kansai U-A, Osaka SU) and paid a uniform participation reward of 500 JPY regardless of whether the subject correctly answered the questions of the Raven test conducted after their choice in the bandit experiment on weighted voting, whereas the instructions were read out using text-to-speech software to subjects of Group B (Kansai U-B, Doshisha, Hiroshima CU) and a monetary reward of 50 JPY for each correct answer was paid. In both groups, subjects were paid a reward for the points they earned in the bandit experiment as well. Kansai University, Osaka Sangyo University, and Doshisha University are located in the same area of the Kansai region, while Hiroshima City University is located in the Chugoku region.[15] The dates, feedback, order of binary choice problems, and other information of the sessions are summarized in Appendix A (Tables 80 to 84) for each site.

Table 2 lists the subjects' attributes at each experimental site. The male-to-female ratios of subjects were almost the same at Kansai U-A and U-B, while the subjects at Osaka SU and Hiroshima CU were overwhelmingly male and female, respectively. The $p$-values for the Fisher exact test for male-to-female ratio were computed in comparison with Kansai University (Kansai U-A and Kansai U-B). The one-sided test was applied to comparison between Kansai U-A and Osaka SU and between Kansai U-B and Hiroshima CU, while the two-sided test was applied to the comparison between Kansai U-B and Doshisha. The male-to-female ratio of subjects at Osaka SU (Hiroshima CU) is significantly higher (lower) than that of the subjects at Kansai U-A (Kansai U-B), while there was no significant difference in the ratio between Kansai U-B and Doshisha. Doshisha (Imadegawa campus) does not have science and engineering departments (sci-eng) and there are no economics major students (econ) at Hiroshima CU.

[15] It takes less than 1 hour by train from Kansai University to Osaka Sangyo University by local train and bus, while it takes approximately 3.5 hours from Kansai University to Hiroshima City University even by the super express (Shinkansen) between the Osaka station and the Hiroshima station.

Table 2: subjects' attribute information

| site | # of subj. | male | female | p-value | econ | sci-eng | others |
|---|---|---|---|---|---|---|---|
| Kansai U-A | 240 | 128 | 112 | | 21 | 53 | 166 |
| Osaka SU | 120 | 99 | 21 | <**0.001** | 47 | 38 | 35 |
| Kansai U-B | 197 | 98 | 99 | | 24 | 41 | 132 |
| Doshisha | 135 | 76 | 56 | 0.828 | 21 | 0 | 114 |
| Hiroshima CU | 124 | 49 | 75 | **0.047** | 0 | 24 | 100 |
| Hiroshima Shudo | 120 | 59 | 61 | 0.920 | 33 | 0 | 87 |

Note: The *p*-values for the Fisher exact test for male-to-female ratio were computed in comparison with Kansai University (Kansai U-A and Kansai U-B). The one-sided test was applied to comparison between Kansai U-A and between Kansai U-B and Hiroshima CU, where the null hypothesis is that male-to-female ratio at OSU (Hiroshima CU) was equal to or lower (higher) than that ratio at Kansai U-A (Kansai U-B). The two sided test was applied to comparison between Kansai U-B and Doshisha, where the null hypothesis is that male-to-female ratios was the same between Kansai U-B and Doshisha (Hiroshima Shudo). Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

Table 3: Raven scores of subjects: basic statistics

| site | # of subj. | mean | std.dev. | p-value | min | max |
|---|---|---|---|---|---|---|
| Kansai U-A | 240 | 11.208 | 2.170 | | 3 | 16 |
| Osaka SU | 120 | 10.625 | 3.041 | **0.027** | 3 | 16 |
| Kansai U-B | 197 | 11.518 | 2.398 | | 2 | 15 |
| Doshisha | 135 | 11.578 | 2.300 | 0.890 | 5 | 16 |
| Hiroshima CU | 124 | 10.976 | 2.441 | **0.038** | 3 | 15 |
| Hiroshima Shudo | 120 | 10.375 | 2.432 | <**0.001** | 4 | 15 |

Note: The p-values for the Brunner-Munzel test were computed in comparison to the data taken at Kansai University (Kansai U-A and Kansai U-B). The p-value for a comparison between Hiroshima CU and Hiroshima Shudo was **0.049**. The null hypothesis is that Raven scores of subjects at an experimental site are, on average, the same as those at Kansai University. Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

Table 4 shows the basic statistics for the Raven scores at each experimental site.[16] The table also shows the p-values for Brunner-Munzel test. We have (1) that subjects at Kansai U-A scored significantly higher on average than those at Osaka SU; (2) that no significant difference existed between subjects at Kansai U-B and those at Doshisha; and (3) that subjects at Kansai U-B scored significantly higher on average than those at Hiroshima CU.

Table 4: Raven scores of subjects: basic statistics

| site | # of subj. | mean | std.dev. | p-value | min | max |
|---|---|---|---|---|---|---|
| Kansai U-A | 240 | 11.208 | 2.170 | | 3 | 16 |
| Osaka SU | 120 | 10.625 | 3.041 | **0.027** | 3 | 16 |
| Kansai U-B | 197 | 11.518 | 2.398 | | 2 | 15 |
| Doshisha | 135 | 11.578 | 2.300 | 0.890 | 5 | 16 |
| Hiroshima CU | 124 | 10.976 | 2.441 | **0.038** | 3 | 15 |
| Hiroshima Shudo | 120 | 10.375 | 2.432 | <**0.001** | 4 | 15 |

Note: The p-values for the Brunner-Munzel test were computed in comparison to the data taken at Kansai University (Kansai U-A and Kansai U-B). The p-value for a comparison between Hiroshima CU and Hiroshima Shudo was **0.049**, and p-values in comparisons between Doshisha and Hiroshima CU and between Doshisha and Shudo were **0.038** and <**0.001**, respectively. The null hypothesis is that Raven scores of subjects at an experimental site are, on average, the same as those at Kansai University. Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

Thus, differences between subjects at Kansai U-A and those at Osaka SU lie in Raven scores (Raven) and male-to-female ratios (gender), there is a difference in sci-eng between Kansai U-B and Doshisha, and differences between subjects at Kansai U-B and those at Hiroshima CU lie in Raven, gender, econ, and regions.

The Economic Experiment Laboratory (EEL) at Kansai University had constructed a subject pool of non-student general public living in the northern part of Osaka Prefecture. Table 5 shows basic statistics of their Raven scores and Figure 1 depicts the histogram. Among 1057 subjects in total, 488 subjects participated in the sessions for Kawamura and Ogawa (2019), the average age of whom was around 55. The Raven scores of the university students who participated in this experiment were, on average, clearly higher than those of the general public.

[16]Those average scores do not represent the cognitive ability of representative students or participants in subject pools at the experimental sites but as characteristics of subjects who participated in the experiment. Prior to proceeding to the experiment, subjects agreed to take the cognitive ability test and the experimenter announced that they could withdraw their participation at any time during the session. After the experiment ends, the Raven scores of individual subjects are disclosed to the individuals concerned (and only those individuals) upon their request.

Table 5: Raven scores of general public: basic statistics

| test site | # of subj. | mean | std. dev. | min | max |
|---|---|---|---|---|---|
| EEL at Kansai U | 1,057 | 7.986 | 3.326 | 0 | 16 |

Note: The Raven scores of non-students noted here were measured from April 2015 to March 2018.

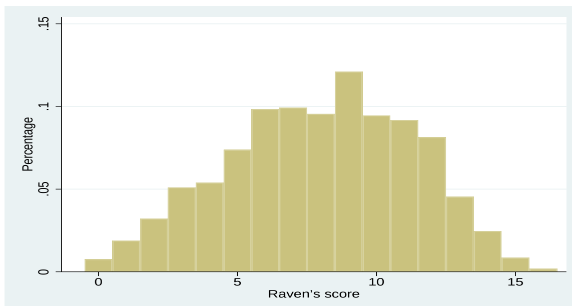



Figure 1: Histogram of Raven scores of non-student general public

# 4 Analysis

In this section, we first compare the rates of correct answers subjects chose for each block of consecutive 10 periods among experimental sites. In what follows, we employ a 5% significance level in rejecting the null hypotheses. See Appendix B for diagrams that present the time series plots of the average rates of correct answers for each period in each of the four sequences of binary choice problems at each experimental site.

## 4.1 Propensity Score Matching

When we conduct an experiment at multiple sites, the subjects are not assigned to the experimental sites randomly. In many cases, subjects in economic experiments are college students, and before their enrollment, they choose their schools and schools select their students through entrance examinations. Accordingly, in experiments at multiple experimental sites a bias arises for the selection of those sites. The rates of correct answers observed in those sites cannot be directly compared, unless some correction is made. We thus presume that the assignment of subjects to the experimental sites is determined by several covariates related to subjects' cognitive ability and attribute information. In the following analysis, we do not use kernel matching but propensity score matching, because kernel matching needs to choose some important covariates in advance but propensity score summarizes a number of potential covariates into a single score, as explained below.

Let $z$ denote the variable that represents experimental sites, say $z = 1$ for site 1 and $z = 2$ for site 2. Suppose that the assignment of subjects to experimental sites depends only on those covariates $\boldsymbol{x}$ and that the rates of correct answers, $y_1$ and $y_2$, observed at those sites are stochastically independent of the assignment of subjects to each experimental site. Under this assumption, it holds that $E(y_1|z = 1, \boldsymbol{x}) = E(y_1|\boldsymbol{x})$ and $E(y_2|z = 2, \boldsymbol{x}) = E(y_2|\boldsymbol{x})$, where $E(y|\boldsymbol{x})$ represents the conditional expected value of a random variable $y$ given $\boldsymbol{x}$. When the experimental sites have the same covariates, we have $E(y_1 - y_2|\boldsymbol{x}) = E(y_1|\boldsymbol{x}) - E(y_2|\boldsymbol{x}) = E(y|z = 1, \boldsymbol{x}) - E(y|z = 2, \boldsymbol{x})$. Thus, the causal effect can be estimated as the average treatment effect $E(y_1 - y_2) = E_{\boldsymbol{x}}E(y_1 - y_2|\boldsymbol{x}) = E_{\boldsymbol{x}}E(y|z = 1, \boldsymbol{x}) - E_{\boldsymbol{x}}E(y|z = 2, \boldsymbol{x})$, where $E_x$ represents the expectation operator with respect to $x$.

**Definition**: Denote by $\boldsymbol{x}_i$ the vector of covariates for subject $i$ and by $z_i$ the variable that stands for assignment to an experimental site. The probability of subject $i$ with his or her covariate vector $\boldsymbol{x}_i$ being assigned to site 1, $e_i = p(z_i = 1|\boldsymbol{x}_i)$, is called the propensity score of subject $i$.

In practice, many researchers estimate the propensity score of subject $i$ by using binomial logistic regression:

$$e_i = \frac{1}{1 + exp(-(b_0 + b_1 x_{i1} + \cdots + b_p x_{ip}))}.$$

Let subject $i$ at site 1 be paired with the subject $j$ at site 2 who has the closest propensity score to his or her propensity score. Their cognitive ability scores and attribute information do not differ greatly, because $e_i$ and $e_j$ have similar values. The paired subjects are thus considered to be assigned randomly to one of the two experimental sites in a pseudo manner. Consider $e$ as a random variable. By using the distribution of $e$, we have

$$E(y_1 - y_2) = E_e E(y_1|z = 1, e) - E_e E(y|z = 2, e),$$

where $z$ denotes the assignment variable. This statistical method is called propensity score matching (Rosenbaum and Rubin, 1983). Note that $\boldsymbol{x}_i$ is a covariate vector for subject $i$ but his or her propensity score $e_i$ is a scaler.

Subject pools are paired in the following analysis; between site 1 and site 2, we estimate the difference in correct answer rate adjusted with covariates using propensity score matching and verify whether the difference is statistically significant using the z-test.[17] The covariates are chosen from subjects cognitive ability score (Raven), age, gender, and affiliated faculty (econ, sci-eng, and others), and the correct answer rates are compared for each of the consecutive 10 periods.

Note that the sample size of each session at each site is small. At Osaka SU, e.g., only 10 subjects participated in each session. We thus use the data merged for all binary choice problems in the analysis. In what follows, we say that subjects' behavior was homogenized by introducing subjects' Raven scores in addition to their general attribute information in covariate adjustment, if the following criteria are satisfied; otherwise, not. The null hypothesis is that the rates of correct answers are the same after the covariate adjustment.

**Criterion 1** *For the data that integrate all feedback treatments, we cannot reject the null hypothesis for every block of 10 consecutive periods when subjects' Raven scores are introduced to all covariates regarding their attribute information.*

**Criterion 2** *For every feedback treatment, (1) the number of the rejection of the null hypothesis does not increase as compared to the case where subjects' Raven scores are not introduced and (2) there are at most two treatments in which we cannot reject the null hypothesis for every block of 10 consecutive periods.*

### Kansai University and Osaka Sangyo University

As shown in Tables 2 and 4 in Section 3, Kansai U-A and Osaka SU are both located in Osaka; each has economics major subjects and science or engineering major subjects; although the male-to-female ratio is skewed at Osaka SU, it is almost the same at Kansai U-A; and the Raven scores of subjects at Kansai U-A are on average higher than those at Osaka SU. Thus, the available covariates are Raven, age, gender, econ, and sci-eng. We will examine the following hypothesis in the case of the sites being located in the same region.

[17] We used *Stata* 16 for our data analysis. In this software, the propensity score matching is accompanied with the z-test as a standard procedure.

**Hypothesis 1** *Even if subjects differed in terms of cognitive ability score and male-to-female ratio between experimental sites, their behavior can be homogenized by introducing Raven scores as a covariate in addition to their general attribute information, when both sites are located in the same region.*

Table 6 lists the p-values for the z-test to confirm Criterion 1. As shown in the table, there were many cases in which a significant difference in correct answer rate exists in periods 41–50, in particular in the case of age, gender, econ, and sci-eng being covariates. There was, however, no significant difference in the case of Raven, age, gender, econ, and sci-eng. Criterion 1 is thus satisfied.

Table 6: p-values for z-test: Kansai U-A vs. Osaka SU, integrated data

| Covariates | $1-10$ | $11-20$ | $21-30$ | $31-40$ | $41-50$ | $51-60$ |
|---|---|---|---|---|---|---|
| age | 0.095 | 0.302 | 0.728 | 0.871 | **0.015** | **0.048** |
| gender | 0.734 | 0.680 | 0.534 | 0.148 | **0.050** | 0.157 |
| age, gender | 0.640 | 0.975 | 0.492 | 0.564 | 0.069 | 0.228 |
| age, econ, sci-eng | 0.517 | 0.281 | 0.750 | 0.740 | **0.004** | 0.066 |
| gender, econ, sci-eng | 0.929 | 0.719 | 0.267 | 0.366 | 0.087 | 0.423 |
| **age, gender, econ, sci-eng** | 0.948 | 0.635 | 0.662 | 0.965 | **0.005** | **0.043** |
| Raven | 0.372 | 0.252 | 0.529 | 0.759 | **0.014** | **0.013** |
| Raven, age | 0.062 | 0.977 | 0.952 | 0.535 | **0.016** | 0.180 |
| Raven, gender | 0.483 | 0.914 | 0.695 | 0.789 | **0.025** | 0.145 |
| Raven, age, gender | 0.822 | 0.532 | 0.776 | 0.275 | 0.079 | 0.471 |
| Raven, age, econ, sci-eng | 0.591 | 0.949 | 0.957 | 0.423 | 0.065 | 0.297 |
| Raven, gender, econ, sci-eng | 0.934 | 0.646 | 0.622 | 0.229 | 0.407 | 0.209 |
| **Raven, age, gender, econ, sci-eng** | 0.937 | 0.648 | 0.532 | 0.323 | 0.295 | 0.387 |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

Tables 7 to 9 list the p-values for the z-test to check Criterion 2, in no-feedback (No-fb), partial-feedback (Part-fb), and full-feedback (Full-fb) treatments, respectively. For No-fb, in the case of Raven, age, gender, econ, and sci-eng, there are significant differences in the correct answer rates in periods 20-30 and periods 31-40. The number of such differences did not increase as compared with the case of age, gender, econ, and sci-eng in which there are two blocks in which a significant difference exists. Moreover, for Part-fb and Full-fb, there is no significant difference in the case of Raven, age, gender, econ, and sci-eng. Criterion 2 is thus satisfied. We have the following result which supports Hypothesis 1.

**Result 1** *For subjects at Kansai University and Osaka Sangyo University, who differed in terms of cognitive ability score and male-to-female ratio, their behavior was homogenized by introducing Raven score as a covariate in addition to their general attribute information.*

Note that in Tables 6 and 7, there are blocks of 10 consecutive periods in which there exists a significant difference in subjects' behavior in the case of solely Raven being used for the covariate adjustment. It would be difficult to simply merge the data even if the Raven scores of subjects are, on average, almost the same between Kansai U-A and Osaka SU.

Table 7: p-values for z-test: Kansai U-A vs. Osaka SU, No-fb

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| age | 0.223 | 0.529 | 0.717 | 0.970 | 0.272 | 0.197 |
| gender | 0.805 | 0.637 | 0.120 | **0.031** | 0.531 | 0.313 |
| age, gender | 0.710 | 0.383 | **0.001** | **<0.001** | 0.660 | 0.596 |
| age, econ, sci-eng | 0.255 | 0.791 | 0.870 | 0.587 | 0.189 | 0.169 |
| gender, econ, sci-eng | 0.979 | 0.649 | 0.081 | **0.011** | 0.424 | 0.461 |
| **age, gender, econ, sci-eng** | 0.816 | 0.567 | **0.007** | **0.002** | 0.386 | 0.323 |
| Raven | 0.483 | 0.393 | 0.893 | 0.965 | 0.072 | **0.013** |
| Raven, age | 0.061 | 0.523 | 0.920 | 0.790 | 0.111 | 0.055 |
| Raven, gender | 0.723 | 0.729 | 0.237 | 0.059 | 0.653 | 0.376 |
| Raven, age, gender | 0.842 | 0.377 | **0.001** | **<0.001** | 0.511 | 0.402 |
| Raven, age, econ, sci-eng | **0.028** | 0.268 | 0.805 | 0.506 | 0.746 | 0.582 |
| Raven, gender, econ, sci-eng | 0.952 | 0.457 | 0.458 | 0.081 | 0.319 | 0.340 |
| **Raven, age, gender, econ, sci-eng** | 0.994 | 0.665 | **0.003** | **<0.001** | 0.932 | 0.994 |

Table 8: p-values for z-test: Kansai U-A vs. Osaka SU, Part-fb

| Covariates | $1-10$ | $11-20$ | $21-30$ | $31-40$ | $41-50$ | $51-60$ |
|---|---|---|---|---|---|---|
| age | 0.074 | 0.082 | 0.493 | 0.706 | 0.201 | 0.947 |
| gender | 0.391 | 0.529 | 0.288 | 0.588 | 0.094 | 0.329 |
| age, gender | 0.322 | 0.340 | 0.560 | 0.619 | 0.120 | 0.669 |
| age, econ, sci-eng | 0.816 | 0.390 | 0.268 | 0.825 | 0.252 | 0.791 |
| gender, econ, sci-eng | 0.689 | 0.411 | 0.229 | 0.851 | 0.175 | 0.848 |
| **age, gender, econ, sci-eng** | 0.198 | 0.145 | 0.186 | 0.263 | 0.232 | 0.915 |
| Raven | 0.086 | 0.148 | 0.130 | 0.087 | 0.128 | 0.917 |
| Raven, age | 0.546 | 0.375 | 0.251 | 0.914 | 0.153 | 0.351 |
| Raven, gender | 0.380 | 0.551 | 0.233 | 0.251 | 0.152 | 0.521 |
| Raven, age, gender | 0.457 | 0.223 | 0.203 | 0.581 | **0.044** | 0.291 |
| Raven, age, econ, sci-eng | 0.962 | 0.637 | 0.202 | 0.989 | **0.024** | 0.129 |
| Raven, gender, econ, sci-eng | 0.829 | 0.758 | **0.016** | 0.678 | 0.124 | 0.502 |
| **Raven, age, gender, econ, sci-eng** | 0.817 | 0.843 | 0.207 | 0.846 | 0.303 | 0.887 |

Table 9: p-values for z-test: Kansai U-A vs. Osaka SU, Full-fb

| Covariates | $1-10$ | $11-20$ | $21-30$ | $31-40$ | $41-50$ | $51-60$ |
|---|---|---|---|---|---|---|
| age | 0.784 | 0.870 | 0.520 | 0.481 | **0.040** | 0.061 |
| gender | 0.835 | 0.571 | 0.843 | 0.890 | 0.310 | 0.611 |
| age, gender | 0.901 | 0.324 | 0.966 | 0.992 | 0.069 | 0.225 |
| age, econ, sci-eng | 0.625 | 0.966 | 0.410 | 0.170 | 0.455 | 0.296 |
| gender, econ, sci-eng | 0.788 | 0.645 | 0.734 | 0.293 | 0.294 | 0.147 |
| **age, gender, econ, sci-eng** | 0.630 | 0.950 | 0.401 | 0.164 | 0.467 | 0.307 |
| Raven | 0.759 | 0.964 | 0.475 | 0.425 | 0.173 | 0.654 |
| Raven, age | 0.452 | 0.660 | 0.615 | 0.813 | 0.284 | 0.812 |
| Raven, gender | 0.367 | 0.963 | 0.807 | 0.730 | 0.080 | 0.437 |
| Raven, age, gender | 0.490 | 0.562 | 0.997 | 0.993 | 0.101 | 0.375 |
| Raven, age, econ, sci-eng | 0.644 | 0.906 | 0.849 | 0.404 | 0.422 | 0.129 |
| Raven, gender, econ, sci-eng | 0.984 | 0.412 | 0.445 | 0.962 | 0.284 | 0.101 |
| **Raven, age, gender, econ, sci-eng** | 0.488 | 0.980 | 0.825 | 0.319 | 0.466 | 0.126 |

### Kansai University and Doshisha University

Kansai U-B and Doshisha had subjects majoring in economics, and there was no significant difference in subjects' cognitive ability score and in male-to-female ratio, while there were no subjects majoring in science or engineering at Doshisha (Tables 2 and 4). The available covariates are Raven, age, gender, and econ. We will examine the following hypothesis to complement Result 1 in which the experimental sites are located in the same region.

**Hypothesis 2** *Even if subjects differed in numbers of those who major in science or engineering between experimental sites, their behavior can be homogenized by introducing Raven scores as a covariate in addition to their general attribute information, when there is no significant difference in cognitive ability score and male-to-female ratio of subjects between those subject pools and the sites are located in the same region.*

Table 10 lists the p-values for the z-test to confirm Criterion 1, which shows that there were almost no cases with a significant difference in correct answer rate. In particular, there was no significant difference in either the case of age, gender, and econ being the covariates or the case of Raven, age, gender, and econ being the covariates. Criterion 1 is thus satisfied.

Table 10: p-values for z-test: Kansai U-B vs. Doshisha, integrated data

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| age | 0.648 | 0.364 | **0.024** | 0.339 | 0.834 | 0.432 |
| gender | 0.793 | 0.866 | 0.157 | 0.467 | 0.691 | 0.100 |
| age, gender | 0.570 | 0.745 | 0.888 | 0.649 | 0.517 | 0.741 |
| age, econ | 0.676 | 0.268 | 0.055 | 0.319 | 0.664 | 0.441 |
| gender, econ | 0.758 | 0.891 | 0.138 | 0.597 | 0.727 | 0.109 |
| **age, gender, econ** | 0.972 | 0.906 | 0.430 | 0.622 | 0.606 | 0.368 |
| Raven | 0.878 | 0.939 | 0.135 | 0.472 | 0.513 | 0.050 |
| Raven, age | 0.539 | 0.996 | 0.093 | 0.661 | 0.656 | 0.133 |
| Raven, gender | 0.509 | 0.659 | 0.134 | 0.770 | 0.753 | 0.132 |
| Raven, age, gender | 0.243 | 0.320 | 0.355 | 0.377 | 0.957 | 0.087 |
| Raven, age, econ | 0.745 | 0.887 | 0.341 | 0.731 | 0.959 | 0.122 |
| Raven, gender, econ | 0.292 | 0.773 | 0.174 | 0.961 | 0.431 | **0.044** |
| **Raven, age, gender, econ** | 0.769 | 0.924 | 0.371 | 0.697 | 0.537 | 0.051 |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

Tables 11 to 13 list the p-values for the z-test to check Criterion 2, in No-fb, Part-fb, and Full-fb, respectively. For No-fb and Part-fb, there was no significant difference in correct answer rate at all. For Full-fb, there was one block of 10 consecutive periods in which the correct answer rates are significantly different in the case of Raven, age, gender, and econ. Kansai U-B, however, has subjects majoring in science or engineering. As shown in Table 14, however, the difference disappeared when we drop the data taken from science or engineering majors at Kansai U-B who are treated as non-economics majors in the test results shown in Table 13. Criterion 2 is satisfied under this covariate adjustment. We have the following result that supports Hypothesis 2.

**Result 2** *For subjects at Kansai University and Doshisha University, their behavior was homogenized by introducing Raven scores in addition to their general attribute information, when we drop the data taken from science or engineering majors at Kansai University.*

In Appendix C, Tables 85 to 87 list the p-values for the z-test in the cases of integrated data, No-fb, and part-fb, where the null hypothesis is that subjects behavior is homogenized between Kansai U-B and Doshisha, when science or engineering major students were excluded from the Kansai U-B. As described in Section 3, there was no significant difference in the male-to-female ratio of subject between Kansai U-B and Doshisha, and both have the department of economics. This is a reason why we dropped the data taken from subjects majoring science or engineering at Kansai U-B.

Note that there was no significant difference in Raven score between subjects who did not major in science or engineering at Kansai U-B and those at Doshisha, as shown in Table 15. Moreover, no significant difference in Raven score was confirmed between science or engineering majors and others (excl. sci-eng) at Kansai U-B. This point implies that the difference in subjects' behavior indicated in Full-fb (Table 13) came not from the cognitive ability scores of subjects majoring in science or engineering at Kansai U-B but from some behavioral characteristic of subjects peculiar to science and engineering.

Table 11: p-values for z-test: Kansai U-B vs. Doshisha, No-fb

| Covariates | $1-10$ | $11-20$ | $21-30$ | $31-40$ | $41-50$ | $51-60$ |
|---|---|---|---|---|---|---|
| age | 0.693 | 0.421 | 0.470 | 0.515 | 0.765 | 0.563 |
| gender | 0.580 | 0.629 | 0.779 | 0.657 | 0.876 | 0.885 |
| age, gender | 0.775 | 0.722 | 0.791 | 0.924 | 0.496 | 0.317 |
| age, econ | 0.533 | 0.496 | 0.596 | 0.549 | 0.689 | 0.508 |
| gender, econ | 0.370 | 0.973 | 0.904 | 0.809 | 0.892 | 0.971 |
| **age, gender, econ** | 0.745 | 0.657 | 0.860 | 0.752 | 0.576 | 0.317 |
| Raven | 0.344 | 0.608 | 0.667 | 0.464 | 0.936 | 0.582 |
| Raven, age | 0.318 | 0.806 | 0.942 | 0.705 | 0.383 | 0.308 |
| Raven, gender | 0.287 | 0.993 | 0.868 | 0.901 | 0.964 | 0.889 |
| Raven, age, gender | 0.920 | 0.479 | 0.961 | 0.941 | 0.841 | 0.926 |
| Raven, age, econ | 0.154 | 0.417 | 0.665 | 0.591 | 0.801 | 0.792 |
| Raven, gender, econ | 0.277 | 0.983 | 0.959 | 0.995 | 0.813 | 0.981 |
| **Raven, age, gender, econ** | 0.307 | 0.321 | 0.993 | 0.692 | 0.774 | 0.842 |

Table 12: p-values for z-test: Kansai U-B vs. Doshisha, Partial-fb

| Covariates | $1-10$ | $11-20$ | $21-30$ | $31-40$ | $41-50$ | $51-60$ |
|---|---|---|---|---|---|---|
| age | 0.127 | 0.886 | 0.194 | 0.989 | 0.865 | 0.415 |
| gender | 0.396 | 0.236 | 0.144 | 0.582 | 0.754 | 0.274 |
| age, gender | 0.453 | 0.780 | 0.607 | 0.604 | 0.854 | 0.414 |
| age, econ | 0.097 | 0.988 | 0.119 | 0.815 | 0.879 | 0.282 |
| gender, econ | 0.259 | 0.362 | 0.065 | 0.466 | 0.671 | 0.226 |
| **age, gender, econ** | 0.574 | 0.516 | 0.548 | 0.678 | 0.898 | 0.312 |
| Raven | 0.549 | 0.235 | 0.235 | 0.880 | 0.710 | 0.272 |
| Raven, age | 0.377 | 0.488 | 0.095 | 0.338 | 0.324 | 0.100 |
| Raven, gender | 0.490 | 0.216 | 0.237 | 0.429 | 0.580 | 0.132 |
| Raven, age, gender | 0.200 | 0.134 | 0.077 | 0.717 | 0.572 | 0.201 |
| Raven, age, econ | 0.210 | 0.519 | 0.144 | 0.524 | 0.665 | 0.205 |
| Raven, gender, econ | 0.838 | 0.106 | 0.363 | 0.730 | 0.576 | 0.321 |
| **Raven, age, gender, econ** | 0.898 | 0.232 | 0.159 | 0.519 | 0.456 | 0.101 |

Table 13: p-values for z-test: Kansai U-B vs. Doshisha, Full-fb

| Covariates | $1-10$ | $11-20$ | $21-30$ | $31-40$ | $41-50$ | $51-60$ |
|---|---|---|---|---|---|---|
| age | 0.898 | 0.671 | 0.145 | 0.801 | 0.667 | 0.095 |
| gender | 0.568 | 0.810 | 0.260 | 0.768 | 0.521 | 0.060 |
| age, gender | 0.950 | 0.507 | 0.230 | 0.228 | 0.984 | 0.078 |
| age, econ | 0.398 | 0.916 | 0.132 | 0.608 | 0.637 | 0.134 |
| gender, econ | 0.376 | 0.842 | 0.257 | 0.832 | 0.514 | 0.071 |
| **age, gender, econ** | 0.526 | 0.743 | 0.452 | 0.376 | 0.716 | 0.060 |
| Raven | 0.761 | 0.871 | 0.485 | 0.980 | 0.685 | **0.033** |
| Raven, age | 0.347 | 0.406 | 0.129 | 0.803 | 0.395 | **0.049** |
| Raven, gender | 0.824 | 0.453 | 0.108 | 0.926 | 0.644 | 0.170 |
| Raven, age, gender | 0.838 | 0.461 | 0.295 | 0.892 | 0.969 | 0.120 |
| Raven, age, econ | 0.959 | 0.450 | 0.091 | 0.981 | 0.342 | **0.047** |
| Raven, gender, econ | 0.414 | 0.925 | 0.297 | 0.845 | 0.115 | **0.012** |
| **Raven, age, gender, econ** | 0.464 | 0.159 | **0.035** | 0.516 | 0.932 | 0.188 |

Table 14: p-values for z-test: Kansai U-B (excl. sci-eng) vs. Doshisha, Full-fb

| Covariates | $1-10$ | $11-20$ | $21-30$ | $31-40$ | $41-50$ | $51-60$ |
|---|---|---|---|---|---|---|
| age | 0.464 | 0.620 | 0.324 | 0.801 | 0.825 | 0.118 |
| gender | 0.396 | 0.790 | 0.582 | 0.982 | 0.594 | 0.089 |
| age, gender | 0.801 | 0.869 | 0.412 | 0.383 | 0.905 | 0.159 |
| age, econ | 0.168 | 0.231 | 0.427 | 0.883 | 0.941 | 0.081 |
| gender, econ | 0.222 | 0.732 | 0.641 | 0.831 | 0.612 | 0.096 |
| age, gender, econ | 0.589 | 0.888 | 0.560 | 0.554 | 0.449 | 0.102 |
| Raven | 0.792 | 0.777 | 0.447 | 0.757 | 0.373 | **0.024** |
| Raven, age | 0.448 | 0.912 | 0.312 | 0.968 | 0.578 | 0.979 |
| Raven, gender | 0.979 | 0.730 | 0.329 | 0.944 | 0.711 | 0.136 |
| Raven, age, gender | 0.928 | 0.802 | 0.658 | 0.611 | 0.384 | 0.272 |
| Raven, age, econ | 0.913 | 0.989 | 0.123 | 0.529 | 0.397 | 0.076 |
| Raven, gender, econ | 0.650 | 0.725 | 0.250 | 0.953 | 0.503 | 0.084 |
| **Raven, age, gender, econ** | 0.845 | 0.948 | 0.393 | 0.667 | 0.902 | 0.193 |

Table 15: Difference in Raven scores between Kansai U-B (excl. sci-eng) and Doshisha

| site | # of subj. | mean | std.dev. | p-value | min | max |
|---|---|---|---|---|---|---|
| Kansai U-B excl. sci-eng | 156 | 11.365 | 2.479 | | 2 | 15 |
| Kansai U-B sci-eng | 41 | 12.098 | 1.985 | 0.068 | 7 | 15 |
| Doshisha | 135 | 11.578 | 2.300 | 0.492 | 5 | 16 |

Note: The p-values for the Brunner-Munzel test was computed in comparison with Kansai U-B excl. sci-eng. Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

### Hiroshima City University vs. Kansai U-B and Doshisha University

First, we compare subjects' behavior between Kansai U-B and Doshisha. Note that Kansai U-B had subjects who major in economics but Hiroshima CU did not have those subjects, while both Kansai had subjects majoring in science or engineering (Tables 2 and 4). The available covariates are Raven, age, gender, and sci-eng. There was a significant difference in male-to-female ratio and in subjects' cognitive ability score on average. We will examine the following hypothesis in the case of the sites being located in different regions.

**Hypothesis 3** *If subjects differed in terms of cognitive ability score, male-to-female ratio, or their affiliation between experimental sites, then it is difficult to homogenize their behavior by introducing Raven scores as a covariate in addition to their general attribute information, when the sites are located in different regions.*

Table 16 lists the p-values for the z-test to confirm Criterion 1, where the null hypothesis is that subjects' behavior is homogenized between Kansai U-B and Hiroshima CU after the covariate adjustment. There was a significant difference in the rates of correct answers in the case of Raven, age, gender, and sci-eng. Criterion 1 is thus not satisfied. Criterion 2 was also not satisfied. Table 17 shows that for No-fb, the number of rejections of the null hypothesis increased and that there was only one treatment in which we could not reject the null hypothesis for every block of 10 consecutive periods, when we added Raven to age, gender, and sci-eng. These problems were not resolved even if we dropped the data taken from economics majors at Kansai University (Table 18) or we compared the behavior of subjects who majored neither in science and engineering nor economics (Table 20).

We thus found that subjects' behavior could not be homogenized with subjects' Raven scores in addition to their general attribute between Kansai U-B and Hiroshima CU in any ways we tried. For more details, see Tables 95 to 106 in Appendix D. Table 19 and Table 21 show that significant difference remained in Raven score between Kansai U-B and Hiroshima CU, when we dropped the data taken from science and engineering majors or economics majors.

Table 16: p-values for z-test: Kansai U-B vs. Hiroshima CU, integrated data

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| age | **0.035** | 0.747 | **0.013** | 0.123 | 0.145 | 0.201 |
| gender | 0.116 | 0.414 | **0.018** | 0.067 | 0.062 | 0.131 |
| age, gender | 0.062 | 0.838 | **0.024** | 0.105 | 0.320 | 0.811 |
| age, sci-eng | 0.136 | 0.415 | **0.004** | 0.068 | 0.173 | 0.318 |
| gender, sci-eng | 0.208 | 0.440 | 0.056 | 0.133 | 0.158 | 0.344 |
| **age, gender, sci-eng** | 0.086 | 0.953 | **0.025** | 0.378 | 0.146 | 0.970 |
| Raven | 0.160 | 0.905 | 0.154 | 0.252 | 0.162 | 0.095 |
| Raven, age | 0.392 | 0.999 | 0.179 | 0.092 | 0.114 | 0.163 |
| Raven, gender | 0.100 | 0.947 | 0.142 | 0.200 | 0.113 | 0.339 |
| Raven, age, gender | 0.091 | 0.642 | 0.102 | 0.227 | 0.550 | 0.744 |
| Raven, age, sci-eng | 0.192 | 0.621 | **0.015** | 0.098 | 0.368 | 0.815 |
| Raven, gender, sci-eng | 0.087 | 0.664 | 0.126 | 0.059 | 0.400 | 0.276 |
| **Raven, age, gender, sci-eng** | 0.367 | 0.903 | 0.104 | 0.279 | **0.030** | 0.339 |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

Table 17: p-values for z-test: Kansai U-B vs. Hiroshima CU

| | Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|---|
| No-fb | age, gender, sci-eng | 0.453 | 0.992 | 0.920 | 0.729 | 0.449 | 0.179 |
| | Raven, age, gender, sci-eng | 0.074 | 0.232 | 0.333 | **0.031** | 0.896 | 0.233 |
| Part-fb | age, gender, sci-eng | 0.566 | 0.304 | 0.197 | 0.079 | <**0.001** | **0.016** |
| | Raven, age, gender, sci-eng | 0.483 | 0.605 | 0.605 | 0.534 | **0.002** | 0.052 |
| Full-fb | age, gender, sci-eng | 0.481 | 0.905 | **0.041** | 0.752 | 0.474 | 0.625 |
| | Raven, age, gender, sci-eng | 0.208 | 0.754 | 0.060 | 0.481 | 0.529 | 0.098 |

Table 18: p-values for z-test: Kansai U-B (excl. econ) vs. Hiroshima CU

| | Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|---|
| No-fb | age, gender, sci-eng | 0.374 | 0.747 | 0.612 | 0.796 | 0.471 | 0.325 |
| | Raven, age, gender, sci-eng | 0.106 | 0.260 | 0.222 | **0.040** | 0.911 | 0.902 |
| Part-fb | age, gender, sci-eng | 0.782 | 0.682 | 0.103 | 0.251 | **0.002** | 0.141 |
| | Raven, age, gender, sci-eng | 0.135 | 0.943 | 0.600 | 0.777 | **0.028** | 0.421 |
| Full-fb | age, gender, sci-eng | 0.638 | 0.829 | **0.004** | 0.542 | 0.102 | 0.500 |
| | Raven, age, gender, sci-eng | 0.539 | 0.633 | 0.058 | 0.591 | 0.500 | 0.343 |

Table 19: Difference in Raven score between Kansai U-B (excl. econ) and Hiroshima CU

| site | # of subj. | mean | std.dev. | p-value | min | max |
|---|---|---|---|---|---|---|
| Hiroshima CU | 124 | 10.976 | 2.441 | | 3 | 15 |
| Kansai U-B excl. econ | 173 | 11.497 | 2.465 | **0.043** | 2 | 15 |

Note: The p-value for the Brunner-Munzel test indicates rejection of the null hypothesis at the 5% significance level.

Table 20: p-values for z-test: Kansai U-B (excl. sci-eng, econ) vs. Hiroshima CU (excl. sci-eng)

| | Covariates | $1-10$ | $11-20$ | $21-30$ | $31-40$ | $41-50$ | $51-60$ |
|---|---|---|---|---|---|---|---|
| No-fb | age, gender | 0.394 | 0.869 | 0.808 | 0.687 | 0.369 | 0.279 |
| | Raven, age, gender | 0.251 | 0.789 | 0.643 | 0.735 | 0.594 | **0.033** |
| Part-fb | age, gender | 0.798 | 0.522 | 0.069 | **0.027** | **0.013** | **0.046** |
| | Raven, age, gender | 0.447 | 0.444 | **0.035** | 0.082 | **0.019** | 0.142 |
| Full-fb | age, gender | 0.956 | 0.493 | 0.066 | 0.221 | 0.629 | 0.128 |
| | Raven, age, gender | 0.154 | 0.296 | 0.098 | 0.436 | 0.733 | 0.317 |

Table 21: Difference in Raven score between Kansai U-B (excl. sci-eng, econ) and Hiroshima CU (excl. sci-eng)

| site | # of subj. | mean | std.dev. | p-value | min | max |
|---|---|---|---|---|---|---|
| Hiroshima CU excl. sci-eng | 100 | 10.590 | 2.425 | | 3 | 15 |
| Kansai U-B excl. sci-eng, econ | 132 | 11.311 | 2.574 | **0.012** | 2 | 15 |

Note: The p-value for the Brunner-Munzel test indicates rejection of the null hypothesis at the 5% significance level.

Next, we compare subjects' behavior between Hiroshima CU and Doshisha. As noted in Section 3, Hiroshima CU does not have the department of economics, while Doshisha did not have science and engineering majors who participated in this experiment. Accordingly, the covariates are Raven, age, and gender. Table 2 shows the difference in male-to-female ratio between Hiroshima CU and Doshisha; Hiroshima CU had more female subjects while Doshisha had more male subjects. The difference was significant. (The p-value for the two-sided Fisher exact test was 0.004.) Table 4 shows that subjects at Kansai U-B scored in the Raven test significantly higher on average than those at Hiroshima CU. The average Raven score of subjects at Doshisha is slightly higher than those at Kansai U-B. Thus, it is easy to see that subjects at Kansai U-B scored significantly higher on average than those at Hiroshima CU. (The p-value for the Fisher exact test was 0.038.)

Table 22 lists the p-values for the z-test in the case of the integrated data to confirm Criterion 1, where the null hypothesis is that subjects' behavior is homogenized between Hiroshima CU and Doshisha after the covariate adjustment. There was no significant difference in the rates of correct answers in the case of Raven, age, and gender. Criterion 1 is thus satisfied. Criterion 2 was, however, not satisfied in Part-fb, as shown in Table 24; the number of rejections of the null hypothesis increased when Raven was introduced as a covariate to the case of age and gender.

When we dropped the data taken from science or engineering majors at Hiroshima CU and economics majors at Doshisha, we found that the number of rejections of the null hypothesis did not increase for Part-fb, while there are two treatments in which we cannot reject the null hypothesis for every block of 10 consecutive periods, as shown in Table 24. Criterion 2 is then satisfied. (Tables 107 to 112 in Appendix D list the p-values for the z-test with other selections of covariates for No-fb, Part-fb, and Full-fb, respectively.) Table 25 shows the differences in Raven score between Hiroshima CU and Doshisha, when some data were dropped in analysis.

Table 22: p-values for z-test: Hiroshima CU vs. Doshisha, integrated data

| Covariates | $1-10$ | $11-20$ | $21-30$ | $31-40$ | $41-50$ | $51-60$ |
|---|---|---|---|---|---|---|
| age | 0.090 | 0.408 | 0.371 | 0.475 | 0.374 | 0.996 |
| gender | 0.103 | 0.281 | 0.459 | 0.465 | 0.113 | 0.998 |
| **age, gender** | 0.060 | 0.218 | 0.308 | 0.494 | 0.228 | 0.780 |
| Raven | 0.054 | 0.189 | 0.354 | 0.414 | 0.066 | 0.678 |
| Raven, age | 0.080 | 0.092 | 0.404 | 0.266 | 0.319 | 0.616 |
| Raven, gender | 0.082 | 0.173 | 0.463 | 0.424 | 0.106 | 0.694 |
| **Raven, age, gender** | 0.112 | 0.224 | 0.742 | 0.476 | 0.804 | 0.493 |

Table 23: p-values for z-test: Hiroshima CU vs. Doshisha

| | Covariates | $1-10$ | $11-20$ | $21-30$ | $31-40$ | $41-50$ | $51-60$ |
|---|---|---|---|---|---|---|---|
| No-fb | age, gender | 0.170 | 0.984 | 0.474 | 0.610 | 0.245 | 0.626 |
| | Raven, age, gender | 0.270 | 0.471 | 0.977 | 0.933 | 0.554 | 0.391 |
| Part-fb | age, gender | 0.474 | 0.058 | 0.599 | 0.169 | **0.022** | 0.183 |
| | Raven, age, gender | 0.652 | **0.017** | 0.333 | 0.097 | **0.013** | 0.857 |
| Full-fb | age, gender | 0.475 | 0.142 | 0.902 | 0.281 | 0.199 | 0.464 |
| | Raven, age, gender | 0.522 | 0.697 | 0.192 | 0.907 | 0.737 | 0.457 |

Table 24: p-values for z-test: Hiroshima CU (excl. sci-eng) vs. Doshisha (excl. econ)

| | Covariates | $1-10$ | $11-20$ | $21-30$ | $31-40$ | $41-50$ | $51-60$ |
|---|---|---|---|---|---|---|---|
| No-fb | age, gender | 0.573 | 0.626 | 0.620 | 0.981 | 0.100 | 0.369 |
| | Raven, age, gender | 0.116 | 0.645 | 0.363 | 0.323 | 0.944 | 0.242 |
| Part-fb | age, gender | 0.416 | **0.003** | 0.470 | 0.101 | 0.195 | 0.483 |
| | Raven, age, gender | 0.415 | **0.009** | 0.547 | 0.231 | 0.152 | 0.560 |
| Full-fb | age, gender | 0.971 | 0.240 | 0.797 | 0.380 | 0.999 | 0.854 |
| | Raven, age, gender | 0.138 | 0.518 | 0.263 | 0.551 | 0.680 | 0.677 |

Table 25: Difference in Raven score between Hiroshima CU and Doshisha

| site | # of subj. | mean | std.dev. | p-value | min | max |
|---|---|---|---|---|---|---|
| Hiroshima CU excl. sci-eng | 100 | 10.590 | 2.425 | | 3 | 15 |
| Hiroshima CU sci-eng | 24 | 12.583 | 1.792 | **$<$0.001** | 8 | 15 |
| Doshisha excl. econ | 114 | 11.404 | 2.356 | | 3 | 15 |
| Doshisha econ | 21 | 12.524 | 1.721 | **0.038** | 9 | 16 |
| Hiroshima CU excl. sci-eng | 100 | 10.590 | 2.425 | | 3 | 15 |
| Doshisha excl. econ | 114 | 11.404 | 2.356 | **0.007** | 5 | 15 |

Note: The p-values for the Brunner-Munzel test were computed in comparison to the data taken at Hiroshima CU excl. sci-eng, Doshisha excl. econ, and Hiroshima CU excl. sci-eng, respectively. The null hypothesis is that Raven scores of subjects at an experimental site are, on average, the same. Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

We have now the following result that supports Hypothesis 3.

**Result 3** *(1) For subjects at Kansai University and Hiroshima City University, who differed in terms of cognitive ability scores and male-to-female ratios, as well as affiliation with the economics department, their behavior could not be homogenized with subjects' Raven scores and their general attribute information. (2) For subjects at Hiroshima City University and Doshisha University, who differed in terms of cognitive ability scores, male-to-female ratios, and their affiliation between experimental sites, subjects' behavior was not homogenized by introducing Raven scores without dropping the data taken from science or engineering majors at Hiroshima City University and economics majors at Doshisha University.*

Falk et al. (2018) conducted an international survey and showed that preferences vary with age, gender, and cognitive ability across individuals, yet these relationships appear to be partly country specific. Subjects' preferences or behaviors are affected by some regional aspects. This paper obtained similar results, but we further showed that subjects' behavior can be homogenized with their Raven scores and general attribute information when the experimental sites are located in the same region. The homogeneity of subjects' behavior is desirable in integration of data collected at multiple experimental sites.

## Kansai U-B (excl. sci-eng) vs. Doshisha: remaining

Table 26: p-values for z-test: Kandai-B (excl. eng.) vs. Doshisha, No-info

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| age | 0.376 | 0.259 | 0.375 | 0.346 | 0.726 | 0.485 |
| gender | 0.446 | 0.305 | 0.657 | 0.454 | 0.925 | 0.768 |
| age, gender | 0.779 | 0.633 | 0.789 | 0.492 | 0.461 | 0.177 |
| age, econ | 0.423 | 0.313 | 0.513 | 0.336 | 0.650 | 0.485 |
| gender, econ | 0.266 | 0.605 | 0.994 | 0.595 | 0.922 | 0.865 |
| **age, gender, econ** | 0.796 | 0.649 | 0.868 | 0.467 | 0.557 | 0.210 |
| Raven | 0.316 | 0.532 | 0.588 | 0.290 | 0.843 | 0.786 |
| Raven, age | 0.352 | 0.657 | 0.954 | 0.900 | 0.575 | 0.813 |
| Raven, gender | 0.223 | 0.384 | 0.913 | 0.681 | 0.668 | 0.404 |
| Raven, age, gender | 0.632 | 0.993 | 0.844 | 0.878 | 0.414 | 0.235 |
| Raven, age, econ | 0.192 | 0.391 | 0.825 | 0.597 | 0.580 | 0.696 |
| Raven, gender, econ | 0.115 | 0.761 | 0.917 | 0.808 | 0.903 | 0.687 |
| **Raven, age, gender, econ** | 0.707 | 0.548 | 0.910 | 0.627 | 0.530 | 0.150 |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

Table 27: p-values for z-test: Kansai U-B (excl. eng.) vs. Doshisha, Partial-info

| Covariates | $1-10$ | $11-20$ | $21-30$ | $31-40$ | $41-50$ | $51-60$ |
|---|---|---|---|---|---|---|
| age | 0.105 | 0.539 | 0.102 | 0.670 | 0.691 | 0.504 |
| gender | 0.409 | 0.112 | 0.118 | 0.401 | 0.701 | 0.475 |
| age, gender | 0.445 | 0.236 | 0.582 | 0.968 | 0.943 | 0.685 |
| age, econ | 0.089 | 0.661 | 0.080 | 0.612 | 0.650 | 0.433 |
| gender, econ | 0.359 | 0.142 | 0.087 | 0.402 | 0.519 | 0.413 |
| **age, gender, econ** | 0.282 | 0.202 | 0.492 | 0.777 | 0.926 | 0.577 |
| Raven | 0.554 | 0.078 | 0.096 | 0.795 | 0.765 | 0.620 |
| Raven, age | 0.169 | 0.834 | **0.002** | 0.070 | 0.152 | 0.053 |
| Raven, gender | 0.571 | 0.135 | 0.152 | 0.382 | 0.869 | 0.553 |
| Raven, age, gender | 0.175 | 0.360 | 0.093 | 0.270 | 0.800 | 0.174 |
| Raven, age, econ | 0.217 | 0.676 | **0.005** | 0.105 | 0.232 | 0.102 |
| Raven, gender, econ | 0.859 | **0.033** | 0.257 | 0.438 | 0.595 | 0.520 |
| **Raven, age, gender, econ** | 0.233 | 0.118 | 0.108 | 0.260 | 0.859 | 0.433 |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

## 5 Concluding Remarks

In a statistical analysis of subject's behavior observed in sessions conducted at multiple experimental sites, it might have been enough to control subjects' attribute information with dummy variables, but it would not be the best way to adopt; at Osaka Sangyo University and Hiroshima City University, there were subjects who had higher Raven scores than those at Kansai University had on average. We asked all subjects to note their reasoning behind their choices after they completed binary choice problems. Reading the notes, we could know many high score holders made some deep inference on the underlying structure of weighted voting, regardless of the experimental sites at which they participated in this experiment. If we simply used a dummy variable representing experimental sites without taking into account their cognitive ability scores, then we would not be able to precisely evaluate the difference in subjects' behavior observed at different sites.

In this paper, we found that across experimental sites located in the same region, it would not be difficult to obtain the homogeneity of subjects' behavior by covariate adjustment with their cognitive ability scores. Subjects' cognitive ability score would thus be an important covariate in integration of data taken at multiple experimental sites or in comparison of those data across different experimental sites. We also found that among the experimental sites located in different regions, it would be difficult to obtain the homogeneity. As mentioned in subsection 1.1, if necessary, subjects' regional background might serve as a blocking factor for randomized block design of experiments, where subjects are divided into homogeneous blocks and treatments are randomly assigned to the blocks.

Let us review what we have shown from another viewpoint. When subjects generalized what they had learned in a binary choice problem to a similar but different one, this higher order concept of learning is called "meaningful learning" (Weber, 2003; Rick and Weber, 2010), "transfer of learning" (Cooper and Kagel, 2003, 2008), or "epiphany" (Dufwenberg et al., 2010). According to Observation 2 shown in Ogawa et al. (2020), it was more plausible that subjects' meaningful learning was observed at the experimental site (Kansai U-A) where their cognitive ability scores were higher on average.[18] Results 1, 2, and 3 thus jointly imply that even if sessions were conducted at multiple experimental sites, subjects' behavior would not differ remarkably when there was no significant difference in their cognitive ability scores in the case where the experimental sites were located in the same region.

Regarding subjects' regional backgrounds, for example, Henrich et al. (2004) found ethnographic evidence for ultimatum games, Engel (2011) conducted a meta-analysis for dictator games, and Shimomura and Yamato (2012) estimated impacts of ethnicities on market outcomes.[19] In the situations these three papers dealt with, payoffs are determined not only by a subject's own decision but also by the others' decisions, and thus different cultural backgrounds result in different attitudes toward others. In order to mitigate this effect, this paper intentionally examined a situation for individual decision-making.

In this paper, as noted at the beginning of Section 1, we confined our attention to subjects' behavior across subject pools of undergraduate students in Japanese universities. We referred to the cognitive ability scores of non-student general public living in the northern part of Osaka Prefecture. (See Table 5 and Figure 1.) In games of prisoner's dilemma, chicken, battle-of-the-sexes, Holm et al. (2020) recently showed how CEOs' strategic behavior and their belief on others' strategic decision-making are different from those of general public. They adopted a stratified random sampling conditioned with subjects' income, age,

[18]This paper is the first half of Ogawa et al. (2020). Observation 2 is mentioned in the second half, where the definition of meaningful learning was strengthened with the concept of win-stay-lose-shift strategy (Nowak and Sigmund, 1993). Meaningful learning was observed also at Doshisha according to the definition by Guerci et al. (2017).

[19]Sutton et al. (2000) is a standard textbook on meta-analysis in medical research. Governments and academic societies systematically review evidence from multiple sources and conduct routine meta-analyses on intervention studies to formulate clinical treatment guidelines.

gender, residential area (as a regional characteristic) and so on, instead of using propensity score matching with those covariates, and suggested that we should take some data on subjects' professional backgrounds as well, when we conduct experiments for examining behavior of the non-student general public.

Finally, we refer to a cognitive ability test that is recently being started to share by researchers across disciplines in social sciences. As noted in subsection 2.3, the 16 questions excerpted from the full version of the Raven's APM test were shared among some laboratories in Japan and overseas. Many papers were written on experiments for weighted voting, asset markets, school choice, etc., by authors who shared the information on the subjects' cognitive ability scores taken with the same test questions. Raven's test kits and answer sheets are copyrighted and must be purchased from a publisher. The Matrix Reasoning Quiz, however, has been recently developed for the same kind of cognitive ability test as the Raven's test, and some researchers have started to cooperate with experiments in not only economics, but also other social sciences (`https://icar-project.com/`).

## Shudo

Table 28: p-values for z-test: Kansai U-B vs. Shudo

| Covariates | $1-10$ | $11-20$ | $21-30$ | $31-40$ | $41-50$ | $51-60$ |
|---|---|---|---|---|---|---|
| age | 0.391 | 0.865 | 0.455 | **0.020** | 0.354 | 0.113 |
| gender | 0.665 | 0.489 | 0.118 | **0.007** | 0.070 | **0.005** |
| age, gender | 0.727 | 0.771 | 0.589 | **0.038** | 0.361 | 0.075 |
| age, econ | 0.531 | 0.553 | 0.442 | **0.041** | 0.356 | 0.112 |
| gender, econ | 0.544 | 0.750 | 0.206 | **0.010** | **0.038** | **0.013** |
| **age, gender, econ** | 0.720 | 0.982 | 0.418 | 0.056 | 0.430 | 0.115 |
| Raven | 0.960 | 0.114 | 0.133 | **0.001** | **0.033** | **0.005** |
| Raven, age | 0.828 | 0.281 | 0.458 | **0.031** | 0.238 | **0.046** |
| Raven, gender | 0.851 | 0.057 | 0.124 | **0.014** | 0.069 | **0.017** |
| Raven, age, gender | 0.715 | 0.663 | 0.705 | 0.145 | 0.075 | **0.012** |
| Raven, age, econ | 0.800 | 0.948 | 0.712 | 0.079 | 0.172 | **0.041** |
| Raven, gender, econ | 0.713 | 0.277 | 0.103 | **0.005** | 0.223 | 0.231 |
| **Raven, age, gender, econ** | 0.857 | 0.429 | 0.204 | **0.005** | 0.272 | 0.230 |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

Table 29: p-values for z-test: HCU vs. Shudo

| Covariates | $1-10$ | $11-20$ | $21-30$ | $31-40$ | $41-50$ | $51-60$ |
|---|---|---|---|---|---|---|
| age | 0.077 | 0.811 | 0.145 | 0.917 | 0.680 | 0.500 |
| gender | 0.122 | 0.741 | 0.238 | 0.586 | 0.828 | 0.322 |
| age, gender | 0.277 | 0.518 | 0.419 | 0.876 | 0.771 | 0.176 |
| Raven | 0.343 | 0.762 | 0.979 | 0.281 | 0.899 | 0.255 |
| Raven, age | 0.107 | 0.733 | 0.575 | 0.218 | 0.599 | 0.257 |
| Raven, gender | 0.211 | 0.720 | 0.261 | 0.722 | 0.727 | 0.177 |
| Raven, age, gender | 0.353 | 0.915 | 0.377 | 0.675 | 0.474 | 0.708 |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

Table 30: p-values for z-test: Doshisha vs. Shudo

| Covariates | $1-10$ | $11-20$ | $21-30$ | $31-40$ | $41-50$ | $51-60$ |
|---|---|---|---|---|---|---|
| age | 0.553 | **0.042** | 0.483 | **0.003** | 0.078 | 0.225 |
| gender | 0.879 | 0.214 | 0.840 | 0.058 | 0.132 | 0.223 |
| age, gender | 0.734 | 0.065 | 0.933 | 0.059 | 0.307 | 0.485 |
| age, econ | 0.673 | **0.047** | 0.531 | **0.014** | **0.028** | 0.096 |
| gender, econ | 0.607 | 0.332 | 0.669 | 0.256 | 0.172 | 0.365 |
| **age, gender, econ** | 0.515 | 0.490 | 0.408 | 0.126 | **0.040** | 0.214 |
| Raven | 0.505 | 0.141 | 0.767 | 0.120 | 0.188 | 0.320 |
| Raven, age | 0.348 | 0.463 | 0.848 | 0.283 | 0.180 | 0.146 |
| Raven, gender | 0.825 | 0.383 | 0.613 | 0.149 | 0.285 | 0.101 |
| Raven, age, gender | 0.904 | 0.242 | 0.868 | 0.261 | 0.298 | 0.480 |
| Raven, age, econ | 0.886 | 0.715 | 0.894 | 0.194 | 0.165 | **0.047** |
| Raven, gender, econ | 0.582 | 0.535 | 0.646 | 0.117 | 0.078 | 0.240 |
| **Raven, age, gender, econ** | 0.953 | 0.676 | 0.446 | 0.486 | 0.065 | 0.310 |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

## Doshisha vs. Shudo

Table 31: p-values for z-test: Doshisha vs. Shudo, NO-info

| Covariates | $1-10$ | $11-20$ | $21-30$ | $31-40$ | $41-50$ | $51-60$ |
|---|---|---|---|---|---|---|
| age | 0.878 | 0.234 | 0.133 | **0.011** | 0.746 | 0.115 |
| gender | 0.568 | 0.650 | 0.241 | 0.050 | 0.445 | 0.053 |
| age, gender | 0.943 | 0.143 | 0.050 | **0.008** | 0.703 | 0.124 |
| age, econ | 0.746 | 0.312 | 0.210 | 0.169 | 0.383 | **0.030** |
| gender, econ | 0.998 | 0.354 | 0.358 | 0.235 | 0.565 | 0.119 |
| **age, gender, econ** | 0.732 | 0.750 | 0.443 | 0.598 | 0.588 | **0.045** |
| Raven | 0.807 | 0.463 | 0.212 | 0.068 | 0.746 | 0.066 |
| Raven, age | 0.781 | 0.574 | 0.101 | **0.033** | 0.296 | **0.024** |
| Raven, gender | 0.956 | 0.601 | 0.186 | **0.039** | 0.693 | 0.108 |
| Raven, age, gender | 0.716 | 0.904 | 0.594 | 0.224 | 0.935 | 0.386 |
| Raven, age, econ | 0.897 | 0.470 | 0.297 | 0.309 | 0.422 | 0.056 |
| Raven, gender, econ | 0.680 | 0.311 | 0.286 | 0.212 | 0.676 | **0.043** |
| **Raven, age, gender, econ** | 0.373 | 0.824 | 0.359 | 0.681 | 0.061 | **0.003** |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

Table 32: p-values for z-test: Doshisha vs. Shudo, Partial-info

| Covariates | $1-10$ | $11-20$ | $21-30$ | $31-40$ | $41-50$ | $51-60$ |
|---|---|---|---|---|---|---|
| age | 0.542 | **0.037** | 0.447 | 0.157 | **0.009** | 0.555 |
| gender | 0.762 | **0.046** | 0.923 | 0.523 | 0.133 | 0.991 |
| age, gender | 0.311 | 0.298 | 0.482 | 0.768 | **0.019** | 0.861 |
| age, econ | 0.763 | 0.053 | 0.338 | 0.140 | **0.007** | 0.228 |
| gender, econ | 0.470 | 0.100 | 0.554 | 0.902 | 0.243 | 0.991 |
| **age, gender, econ** | 0.524 | **0.031** | 0.668 | 0.110 | **0.023** | 0.997 |
| Raven | 0.948 | 0.151 | 0.942 | 0.428 | 0.072 | 0.514 |
| Raven, age | 0.368 | 0.404 | 0.896 | 0.746 | 0.383 | 0.722 |
| Raven, gender | 0.893 | **0.023** | 0.665 | 0.325 | 0.250 | 0.703 |
| Raven, age, gender | 0.824 | **0.016** | 0.502 | **0.039** | 0.297 | 0.879 |
| Raven, age, econ | 0.371 | 0.415 | 0.824 | 0.528 | 0.266 | 0.510 |
| Raven, gender, econ | 0.757 | 0.058 | 0.998 | 0.660 | 0.156 | 0.964 |
| **Raven, age, gender, econ** | 0.958 | **0.003** | 0.344 | **0.016** | 0.323 | 0.899 |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

Table 33: p-values for z-test: Doshisha vs. Shudo, Full-info

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| age | 0.878 | 0.234 | 0.133 | **0.011** | 0.746 | 0.115 |
| gender | 0.568 | 0.650 | 0.241 | 0.050 | 0.445 | 0.053 |
| age, gender | 0.943 | 0.143 | 0.050 | **0.008** | 0.703 | 0.124 |
| age, econ | 0.746 | 0.312 | 0.210 | 0.169 | 0.383 | **0.030** |
| gender, econ | 0.998 | 0.354 | 0.358 | 0.235 | 0.565 | 0.119 |
| **age, gender, econ** | 0.732 | 0.750 | 0.443 | 0.598 | 0.588 | **0.045** |
| Raven | 0.807 | 0.463 | 0.212 | 0.068 | 0.746 | 0.066 |
| Raven, age | 0.781 | 0.574 | 0.101 | **0.033** | 0.296 | **0.024** |
| Raven, gender | 0.956 | 0.601 | 0.186 | **0.039** | 0.693 | 0.108 |
| Raven, age, gender | 0.716 | 0.904 | 0.594 | 0.224 | 0.935 | 0.386 |
| Raven, age, econ | 0.897 | 0.470 | 0.297 | 0.309 | 0.422 | 0.056 |
| Raven, gender, econ | 0.680 | 0.311 | 0.286 | 0.212 | 0.676 | **0.043** |
| **Raven, age, gender, econ** | 0.373 | 0.824 | 0.359 | 0.681 | 0.061 | **0.003** |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

## HCU vs. Shudo

Table 34: p-values for z-test: HCU vs. Shudo, Full-info

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| age | 0.505 | 0.719 | **0.010** | 0.824 | 0.339 | 0.976 |
| gender | 0.916 | 0.819 | **0.025** | 0.531 | 0.416 | 0.755 |
| age, gender | 0.475 | 0.210 | **0.011** | 0.558 | 0.393 | 0.703 |
| Raven | 0.627 | 0.775 | **0.034** | 0.414 | 0.832 | 0.819 |
| Raven, age | 0.424 | 0.638 | **0.048** | 0.106 | 0.846 | 0.817 |
| Raven, gender | 0.425 | 0.749 | 0.055 | 0.485 | 0.866 | 0.656 |
| Raven, age, gender | 0.966 | 0.538 | **0.000** | 0.977 | 0.684 | 0.764 |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

Table 35: p-values for z-test: HCU vs. Shudo, Partial-info

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| age | 0.750 | 0.356 | 0.421 | 0.239 | 0.661 | 0.695 |
| gender | 0.735 | 0.309 | 0.493 | 0.551 | 0.354 | 0.368 |
| age, gender | 0.873 | 0.195 | 0.219 | 0.134 | 0.810 | 0.855 |
| Raven | 0.861 | 0.730 | 0.726 | 0.665 | 0.626 | 0.804 |
| Raven, age | 0.890 | 0.644 | 0.632 | 0.494 | 0.601 | 0.746 |
| Raven, gender | 0.612 | 0.522 | 0.549 | 0.817 | 0.895 | 0.907 |
| Raven, age, gender | 0.842 | 0.197 | 0.444 | 0.416 | 0.501 | 0.501 |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

Table 36: p-values for z-test: HCU vs. Shudo, No-info

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| age | **0.049** | 0.451 | 0.297 | 0.182 | 0.332 | 0.102 |
| gender | **0.034** | 0.619 | 0.633 | 0.414 | 0.836 | **0.041** |
| age, gender | 0.100 | 0.667 | 0.703 | 0.519 | 0.164 | 0.101 |
| Raven | 0.140 | 0.351 | 0.127 | 0.077 | 0.825 | 0.057 |
| Raven, age | 0.114 | 0.261 | **0.033** | **0.046** | 0.568 | 0.195 |
| Raven, gender | 0.103 | 0.276 | 0.484 | 0.271 | 0.823 | 0.079 |
| Raven, age, gender | **0.001** | 0.785 | 0.868 | 0.957 | 0.564 | **0.008** |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

## Kansai U-B vs. Shudo

Table 37: p-values for z-test: Kansai U-B vs. Shudo, Full-info

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| age | 0.719 | 0.601 | 0.591 | 0.643 | 0.977 | 0.426 |
| gender | 0.647 | 0.830 | 0.710 | 0.351 | 0.394 | 0.136 |
| age, gender | 0.526 | 0.253 | 0.496 | 0.759 | 0.921 | 0.357 |
| age, econ | 0.744 | 0.962 | 0.282 | 0.641 | 0.977 | 0.707 |
| gender, econ | 0.578 | 0.720 | 0.656 | 0.257 | 0.417 | 0.234 |
| **age, gender, econ** | 0.864 | 0.720 | 0.151 | 0.824 | 0.727 | 0.832 |
| Raven | 0.563 | 0.866 | 0.383 | 0.587 | 0.308 | 0.115 |
| Raven, age | 0.733 | 0.737 | 0.309 | 0.509 | 0.592 | 0.475 |
| Raven, gender | 0.571 | 0.881 | 0.280 | 0.549 | 0.343 | 0.185 |
| Raven, age, gender | 0.460 | 0.589 | 0.482 | 0.887 | 0.586 | 0.085 |
| Raven, age, econ | 0.563 | 0.396 | 0.719 | 0.226 | 0.254 | 0.215 |
| Raven, gender, econ | 0.727 | 0.805 | 0.351 | 0.725 | 0.467 | 0.112 |
| **Raven, age, gender, econ** | 0.828 | 0.796 | 0.342 | 0.171 | 0.231 | 0.283 |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

Table 38: p-values for z-test: Kansai U-B vs. Shudo, Partial-info

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| age | 0.415 | 0.221 | 0.091 | 0.116 | **0.010** | 0.179 |
| gender | 0.686 | 0.647 | 0.192 | 0.314 | **0.038** | 0.231 |
| age, gender | 0.194 | 0.119 | 0.068 | **0.025** | 0.364 | 0.660 |
| age, econ | 0.369 | 0.325 | 0.120 | 0.172 | **0.011** | 0.197 |
| gender, econ | 0.717 | 0.667 | 0.210 | 0.324 | 0.081 | 0.337 |
| **age, gender, econ** | 0.333 | 0.393 | 0.251 | 0.263 | 0.476 | 0.679 |
| Raven | 0.307 | 0.248 | 0.126 | 0.187 | **0.006** | 0.087 |
| Raven, age | 0.972 | 0.767 | 0.912 | 0.902 | **0.019** | 0.098 |
| Raven, gender | 0.221 | 0.253 | 0.215 | 0.182 | **0.036** | 0.363 |
| Raven, age, gender | 0.381 | 0.171 | 0.147 | 0.150 | 0.069 | 0.689 |
| Raven, age, econ | 0.823 | 0.540 | 0.957 | 0.959 | **0.011** | 0.143 |
| Raven, gender, econ | 0.203 | 0.235 | 0.105 | 0.151 | 0.054 | 0.163 |
| **Raven, age, gender, econ** | 0.226 | 0.232 | 0.246 | 0.111 | 0.357 | 0.916 |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

Table 39: p-values for z-test: Kansa Ui-B vs. Shudo, No-info

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| age | 0.231 | 0.215 | 0.081 | **0.003** | 0.837 | 0.071 |
| gender | 0.174 | 0.408 | 0.113 | **0.014** | 0.649 | **0.037** |
| age, gender | 0.727 | 0.138 | **0.048** | **0.003** | 0.990 | 0.251 |
| age, econ | 0.113 | 0.712 | 0.330 | **0.038** | 0.871 | 0.209 |
| gender, econ | 0.102 | 0.945 | 0.184 | **0.021** | 0.243 | 0.117 |
| **age, gender, econ** | 0.090 | 0.810 | 0.351 | 0.051 | 0.592 | 0.183 |
| Raven | 0.573 | 0.113 | **0.017** | **0.001** | 0.948 | **0.044** |
| Raven, age | 0.210 | 0.722 | 0.148 | **0.010** | 0.475 | **0.009** |
| Raven, gender | 0.211 | 0.261 | 0.059 | **0.008** | 0.607 | **0.035** |
| Raven, age, gender | 0.188 | 0.830 | 0.364 | 0.097 | 0.775 | 0.150 |
| Raven, age, econ | 0.212 | 0.662 | 0.255 | **0.015** | 0.682 | 0.070 |
| Raven, gender, econ | 0.123 | 0.835 | 0.196 | **0.041** | 0.589 | 0.462 |
| **Raven, age, gender, econ** | 0.141 | 0.836 | 0.484 | **0.044** | 0.825 | 0.291 |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

## Kansai U-B (excl. sci-eng) vs. Shudo

Table 40: p-values for z-test: Kansai U-B (excl. eng.) vs. Shudo

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| age | 0.795 | 0.655 | 0.362 | **0.008** | 0.417 | 0.069 |
| gender | 0.573 | 0.483 | 0.146 | **0.008** | 0.084 | **0.006** |
| age, gender | 0.858 | 0.609 | 0.181 | **0.006** | 0.219 | **0.022** |
| age, econ | 0.999 | 0.687 | 0.432 | **0.036** | 0.700 | 0.375 |
| gender, econ | 0.522 | 0.809 | 0.195 | **0.009** | **0.033** | **0.011** |
| **age, gender, econ** | 0.623 | 0.732 | 0.351 | **0.018** | 0.871 | 0.266 |
| Raven | 0.858 | 0.221 | 0.094 | **0.006** | 0.113 | **0.024** |
| Raven, age | 0.573 | 0.163 | 0.080 | **0.004** | 0.213 | **0.012** |
| Raven, gender | 0.835 | 0.115 | 0.076 | **0.002** | 0.095 | **0.027** |
| Raven, age, gender | 0.759 | 0.332 | 0.219 | **0.005** | 0.146 | **0.016** |
| Raven, age, econ | 0.548 | 0.873 | 0.517 | 0.089 | 0.116 | **0.033** |
| Raven, gender, econ | 0.666 | 0.541 | 0.133 | **0.028** | 0.317 | 0.089 |
| **Raven, age, gender, econ** | 0.921 | 0.311 | 0.347 | 0.094 | **0.037** | **0.049** |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

Table 41: p-values for z-test: Kansai U-B (excl. eng.) vs. Shudo, No-info

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| age | 0.155 | 0.160 | 0.093 | **0.003** | 0.963 | 0.117 |
| gender | 0.112 | 0.220 | 0.107 | **0.010** | 0.681 | **0.038** |
| age, gender | 0.531 | 0.229 | 0.114 | **0.014** | 0.725 | 0.287 |
| age, econ | 0.058 | 0.641 | 0.341 | **0.048** | 0.781 | 0.125 |
| gender, econ | 0.087 | 0.843 | 0.142 | **0.011** | 0.267 | 0.127 |
| **age, gender, econ** | 0.086 | 0.499 | 0.696 | 0.117 | 0.527 | 0.315 |
| Raven | 0.270 | 0.085 | **0.020** | **p<0.001** | 0.976 | **0.028** |
| Raven, age | 0.195 | 0.138 | **0.037** | **0.002** | 0.546 | 0.227 |
| Raven, gender | 0.081 | 0.116 | **0.015** | **p<0.001** | 0.395 | **0.008** |
| Raven, age, gender | 0.484 | 0.556 | 0.498 | 0.124 | 0.428 | 0.345 |
| Raven, age, econ | 0.096 | 0.847 | 0.952 | 0.191 | 0.699 | 0.073 |
| Raven, gender, econ | 0.141 | 0.945 | 0.081 | **0.011** | 0.598 | 0.298 |
| **Raven, age, gender, econ** | **0.046** | 0.568 | 0.772 | 0.225 | 0.870 | 0.441 |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

Table 42: p-values for z-test: Kansa Ui-B (excl. eng.) vs. Shudo, Partial-info

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| age | 0.351 | 0.426 | 0.089 | 0.209 | **0.039** | 0.260 |
| gender | 0.437 | 0.821 | 0.108 | 0.282 | **0.027** | 0.232 |
| age, gender | 0.208 | 0.591 | 0.059 | 0.066 | **0.044** | 0.209 |
| age, econ | 0.325 | 0.556 | 0.087 | 0.203 | 0.136 | 0.438 |
| gender, econ | 0.420 | 0.782 | 0.090 | 0.247 | **0.029** | 0.251 |
| **age, gender, econ** | 0.548 | 0.813 | 0.116 | 0.131 | **0.042** | 0.194 |
| Raven | 0.238 | 0.614 | 0.198 | 0.401 | **0.004** | 0.309 |
| Raven, age | 0.451 | 0.836 | 0.232 | 0.265 | 0.057 | 0.102 |
| Raven, gender | 0.256 | 0.747 | 0.230 | 0.445 | **0.007** | 0.301 |
| Raven, age, gender | 0.360 | 0.672 | 0.570 | 0.820 | **0.001** | 0.140 |
| Raven, age, econ | 0.457 | 0.870 | 0.351 | 0.540 | **0.031** | 0.076 |
| Raven, gender, econ | 0.116 | 0.515 | 0.090 | 0.306 | **0.003** | 0.194 |
| **Raven, age, gender, econ** | 0.188 | 0.553 | 0.085 | 0.340 | **p<0.001** | 0.068 |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

Table 43: p-values for z-test: Kansai U-B (excl. eng.) vs. Shudo, Full-info

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| age | 0.554 | 0.365 | 0.296 | 0.945 | 0.952 | 0.522 |
| gender | 0.809 | 0.587 | 0.362 | 0.496 | 0.536 | 0.166 |
| age, gender | 0.454 | 0.459 | 0.722 | 0.528 | 0.668 | 0.229 |
| age, econ | 0.714 | 0.828 | 0.156 | 0.724 | 0.986 | 0.727 |
| gender, econ | 0.767 | 0.533 | 0.400 | 0.356 | 0.507 | 0.221 |
| **age, gender, econ** | 0.536 | 0.647 | 0.555 | 0.453 | 0.674 | 0.239 |
| Raven | 0.896 | 0.879 | 0.126 | 0.435 | 0.326 | 0.126 |
| Raven, age | 0.584 | 0.913 | 0.612 | 0.733 | 0.322 | 0.147 |
| Raven, gender | 0.653 | 0.491 | 0.191 | 0.798 | 0.281 | 0.161 |
| Raven, age, gender | 0.536 | 0.358 | 0.729 | 0.940 | 0.547 | 0.310 |
| Raven, age, econ | 0.499 | 0.871 | 0.490 | 0.978 | 0.328 | 0.184 |
| Raven, gender, econ | 0.789 | 0.151 | 0.095 | 0.664 | 0.470 | 0.317 |
| **Raven, age, gender, econ** | 0.366 | 0.811 | 0.481 | 0.807 | 0.548 | 0.143 |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

## Kandai (excl. male) vs. HCU (excl. male)

Table 44: p-values for z-test: Kansai U-B (excl. male) vs. HCU (excl. male)

| Covariates | 1 – 10 | 11 – 20 | 21 – 30 | 31 – 40 | 41 – 50 | 51 – 60 |
|---|---|---|---|---|---|---|
| age | 0.061 | 0.533 | 0.381 | 0.143 | **0.036** | 0.302 |
| sci-eng | 0.075 | 0.641 | 0.396 | 0.128 | **0.026** | 0.148 |
| **age, sci-eng** | 0.065 | 0.679 | 0.254 | 0.062 | **0.022** | 0.458 |
| Raven | 0.094 | 0.376 | 0.633 | 0.404 | 0.185 | 0.256 |
| Raven, age | 0.374 | 0.229 | 0.740 | 0.210 | 0.143 | 0.232 |
| Raven, sci-eng | 0.083 | 0.713 | 0.441 | 0.141 | **0.039** | 0.187 |
| **Raven, age, sci-eng** | 0.112 | 0.367 | 0.279 | 0.255 | 0.139 | 0.468 |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

Table 45: p-values for z-test: Kansa Ui-B (excl. male) vs. HCU (excl. male), No-info

| Covariates | 1 – 10 | 11 – 20 | 21 – 30 | 31 – 40 | 41 – 50 | 51 – 60 |
|---|---|---|---|---|---|---|
| age | 0.134 | 0.193 | 0.062 | 0.803 | 0.715 | 0.817 |
| Raven | 0.608 | 0.153 | **0.018** | 0.634 | 0.778 | 0.993 |
| Raven, age | 0.358 | 0.109 | **0.001** | 0.242 | 0.756 | 0.608 |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

. 理系ダミーを入れると ote: rikei dummy != 0 predicts success perfectly rikei dummy dropped and 1 obs not used perfect predictors are not allowed; try estimating your logit model before calling teffects psmatch というエラーが出て計算できない。

Table 46: p-values for z-test: Kansai U-B (excl. male) vs. HCU (excl. male), Partial-info

| Covariates | 1 – 10 | 11 – 20 | 21 – 30 | 31 – 40 | 41 – 50 | 51 – 60 |
|---|---|---|---|---|---|---|
| age | 0.358 | 0.771 | 0.094 | 0.097 | **0.006** | 0.138 |
| sci-eng | 0.243 | 0.340 | **0.024** | 0.085 | **0.001** | **0.047** |
| **age, sci-eng** | 0.358 | 0.771 | 0.094 | 0.097 | **0.006** | 0.138 |
| Raven | 0.144 | 0.413 | **0.048** | 0.202 | **0.007** | 0.099 |
| Raven, age | 0.605 | 0.875 | 0.254 | 0.438 | **p<0.001** | 0.090 |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

理系ダミーを入れると ote: rikei dummy != 0 predicts success perfectly rikei dummy dropped and 1 obs not used perfect predictors are not allowed; try estimating your logit model before calling teffects psmatch というエラーが出て計算できない。

Table 47: p-values for z-test: Kansai u-B (excl. male) vs. HCU (excl. male), Full-info

| Covariates | $1-10$ | $11-20$ | $21-30$ | $31-40$ | $41-50$ | $51-60$ |
|---|---|---|---|---|---|---|
| age | 0.737 | 0.678 | **0.017** | 0.106 | 0.477 | 0.332 |
| sci-eng | 0.677 | 0.985 | **0.040** | **0.017** | 0.377 | 0.356 |
| **age, sci-eng** | 0.952 | 0.590 | **0.001** | 0.060 | 0.457 | 0.422 |
| Raven | 0.474 | 0.964 | **0.016** | 0.066 | 0.516 | 0.187 |
| Raven, age | 0.400 | 0.452 | **0.002** | 0.050 | 0.717 | 0.574 |
| Raven, sci-eng | 0.419 | 0.533 | **0.004** | **0.020** | 0.522 | 0.128 |
| **Raven, age, sci-eng** | 0.265 | 0.326 | **p<0.001** | **0.002** | 0.538 | 0.192 |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

## Doshisha (female) vs. HCU (female)

Table 48: p-values for z-test: Doshisha (excl. male) vs. HCU (excl. male)

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| **age** | 0.617 | 0.453 | 0.531 | 0.602 | 0.264 | 0.883 |
| Raven | 0.176 | 0.933 | 0.914 | 0.912 | 0.116 | 0.690 |
| **Raven, age** | 0.468 | 0.746 | 0.357 | 0.605 | 0.359 | 0.818 |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

Table 49: p-values for z-test: Doshisha (excl. male) vs. HCU (excl. male), Full-info

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| **age** | 0.705 | 0.288 | 0.876 | 0.537 | 0.506 | 0.453 |
| Raven | 0.825 | 0.376 | 0.254 | 0.980 | 0.302 | 0.229 |
| **Raven, age** | 0.476 | 0.404 | 0.061 | 0.490 | 0.608 | 0.280 |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

Table 50: p-values for z-test: Doshisha (excl. male) vs. HCU (excl. male), Partial-info

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| **age** | 0.602 | **0.003** | 0.570 | 0.104 | 0.054 | 0.427 |
| Raven | 0.837 | **0.015** | 0.818 | 0.640 | **0.002** | 0.413 |
| **Raven, age** | 0.415 | **0.027** | 0.593 | 0.129 | **0.014** | 0.069 |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

Table 51: p-values for z-test: Doshisha (excl. male) vs. HCU (excl. male), No-info

| Covariates | $1-10$ | $11-20$ | $21-30$ | $31-40$ | $41-50$ | $51-60$ |
|---|---|---|---|---|---|---|
| **age** | 0.334 | 0.113 | 0.450 | 0.443 | 0.351 | 0.828 |
| Raven | 0.270 | 0.082 | 0.669 | 0.568 | 0.513 | 0.208 |
| **Raven, age** | 0.389 | 0.189 | 0.683 | 0.602 | 0.691 | 0.192 |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

## Doshisha (excl. male, econ.) vs. HCU (excl. male)

Table 52: p-values for z-test: Doshisha (excl. male and econ.) vs. HCU (excl. male)

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| **age** | 0.773 | 0.502 | 0.363 | 0.701 | 0.157 | 0.648 |
| Raven | 0.494 | 0.613 | 0.629 | 0.969 | 0.205 | 0.893 |
| **Raven, age** | 0.516 | 0.090 | 0.704 | 0.519 | 0.373 | 0.815 |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

Table 53: p-values for z-test: Doshisha (excl. male and econ.) vs. HCU (excl. male), Full-info

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| **age** | 0.650 | 0.052 | 0.677 | 0.421 | 0.517 | 0.175 |
| Raven | 0.841 | 0.094 | 0.092 | 0.833 | 0.573 | 0.175 |
| **Raven, age** | 0.548 | 0.056 | 0.491 | 0.811 | 0.970 | 0.350 |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

Table 54: p-values for z-test: Doshisha (excl. male and econ.) vs. HCU (excl. male), Partial-info

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| **age** | 0.345 | **0.005** | 0.627 | 0.090 | **0.026** | 0.310 |
| Raven | 0.775 | **0.007** | 0.449 | 0.655 | **0.020** | 0.633 |
| **Raven, age** | 0.420 | **0.016** | 0.689 | 0.182 | **0.043** | 0.283 |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

Table 55: p-values for z-test: Doshisha (excl. male and econ.) vs. HCU (excl. male), No-info

| Covariates | $1-10$ | $11-20$ | $21-30$ | $31-40$ | $41-50$ | $51-60$ |
|---|---|---|---|---|---|---|
| **age** | 0.334 | 0.113 | 0.450 | 0.443 | 0.351 | 0.828 |
| Raven | 0.270 | 0.082 | 0.669 | 0.568 | 0.513 | 0.208 |
| **Raven, age** | 0.389 | 0.189 | 0.683 | 0.602 | 0.691 | 0.192 |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

## HCU vs. Shudo

Table 56: p-values for z-test: Shudo vs. Hiroshima CU (excel male), integrated.

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| **age** | 0.120 | 0.478 | 0.446 | 0.602 | 0.339 | 0.095 |
| Raven | 0.555 | 0.967 | 0.163 | 0.479 | 0.729 | 0.524 |
| **Raven, age** | 0.442 | 0.814 | 0.235 | 0.620 | 0.945 | 0.138 |

Table 57: p-values for z-test: Shudo vs. Hiroshima CU (excel male), No-info

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| **age** | 0.136 | 0.093 | 0.609 | 0.416 | 0.282 | 0.386 |
| Raven | 0.362 | **0.016** | 0.306 | 0.114 | 0.918 | 0.323 |
| **Raven, age** | **0.004** | **0.049** | 0.242 | 0.355 | 0.952 | 0.187 |

Table 58: p-values for z-test: Shudo vs. Hiroshima CU (excel male), Partial-info

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| **age** | 0.424 | 0.606 | 0.158 | 0.076 | 0.362 | 0.330 |
| Raven | 0.579 | 0.164 | 0.358 | 0.077 | 0.828 | 0.953 |
| **Raven, age** | 0.770 | 0.256 | 0.111 | 0.063 | 0.297 | 0.485 |

Table 59: p-values for z-test: Shudo vs. Hiroshima CU (excel male), Full-info

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| **age** | 0.127 | 0.246 | **p<0.001** | 0.111 | 0.996 | 0.877 |
| Raven | 0.908 | 0.230 | 0.243 | 0.337 | 0.253 | 0.153 |
| **Raven, age** | 0.252 | 0.288 | **0.004** | 0.058 | 0.523 | 0.453 |

Table 60: p-values for z-test: Shudo vs. Hiroshima CU (excel male, sci-eng.), integrated

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| **age** | 0.344 | 0.988 | **0.049** | 0.241 | 0.683 | 0.631 |
| Raven | 0.118 | 0.804 | 0.425 | 0.965 | 0.876 | 0.235 |
| **Raven, age** | 0.335 | 0.712 | 0.504 | 0.985 | 0.618 | 0.329 |

Table 61: p-values for z-test: Shudo vs. Hiroshima CU (excel male, sci-eng.), No-info

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| **age** | 0.136 | 0.093 | 0.609 | 0.416 | 0.282 | 0.386 |
| Raven | 0.362 | **0.016** | 0.306 | 0.114 | 0.918 | 0.323 |
| **Raven, age** | **0.004** | **0.049** | 0.242 | 0.355 | 0.952 | 0.187 |

Table 62: p-values for z-test: Shudo vs. Hiroshima CU (excel male, sci-eng.), Partial-info

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| **age** | 0.424 | 0.606 | 0.158 | 0.076 | 0.362 | 0.330 |
| Raven | 0.579 | 0.164 | 0.358 | 0.077 | 0.828 | 0.953 |
| **Raven, age** | 0.770 | 0.256 | 0.111 | 0.063 | 0.297 | 0.485 |

Table 63: p-values for z-test: Shudo vs. Hiroshima CU (excel male, sci-eng.), Full-info

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| **age** | **0.004** | **0.023** | **p<0.001** | **0.006** | 0.757 | 0.774 |
| Raven | 0.443 | 0.360 | **0.003** | 0.230 | 0.888 | 0.482 |
| **Raven, age** | 0.095 | **0.030** | **p<0.001** | 0.**p<0.001** | 0.618 | 1.000 |

Table 64: p-values for z-test: Shudo vs. Hiroshima CU (excel male, econ.), integrated

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| **age** | 0.553 | 0.662 | 0.155 | 0.535 | 0.206 | 1.000 |
| Raven | 0.301 | 0.604 | 0.597 | 0.602 | 0.561 | 0.229 |
| **Raven, age** | 0.977 | 0.721 | 0.746 | 0.990 | 0.661 | 0.767 |

Table 65: p-values for z-test: Shudo vs. Hiroshima CU (excel male, econ.), No-info

| Covariates | 1 – 10 | 11 – 20 | 21 – 30 | 31 – 40 | 41 – 50 | 51 – 60 |
|---|---|---|---|---|---|---|
| **age** | 0.471 | 0.487 | 0.845 | 0.977 | 0.089 | 0.995 |
| Raven | 0.940 | **0.024** | 0.254 | **0.046** | 0.518 | 0.904 |
| **Raven, age** | 0.069 | **0.038** | 0.327 | 0.245 | 0.538 | **0.049** |

Table 66: p-values for z-test: Shudo vs. Hiroshima CU (excel male, econ.), Partial-info

| Covariates | 1 – 10 | 11 – 20 | 21 – 30 | 31 – 40 | 41 – 50 | 51 – 60 |
|---|---|---|---|---|---|---|
| **age** | 0.793 | 0.380 | 0.471 | 0.246 | 0.562 | 0.483 |
| Raven | 0.314 | 0.060 | 0.278 | 0.068 | 0.951 | 0.855 |
| **Raven, age** | 0.774 | 0.643 | 0.632 | 0.286 | 0.612 | 0.802 |

Table 67: p-values for z-test: Shudo vs. Hiroshima CU (excel male, econ.), Full-info

| Covariates | 1 – 10 | 11 – 20 | 21 – 30 | 31 – 40 | 41 – 50 | 51 – 60 |
|---|---|---|---|---|---|---|
| **age** | 0.541 | 0.144 | **0.047** | 0.382 | 0.681 | 0.917 |
| Raven | 0.815 | 0.278 | 0.332 | 0.196 | 0.203 | 0.096 |
| **Raven, age** | 0.143 | 0.054 | **0.014** | 0.343 | 0.333 | 0.782 |

Table 68: p-values for z-test: Shudo vs. Hiroshima CU (excel male, econ. sci-eng), integrated

| Covariates | 1 – 10 | 11 – 20 | 21 – 30 | 31 – 40 | 41 – 50 | 51 – 60 |
|---|---|---|---|---|---|---|
| **age** | 0.650 | 0.684 | 0.248 | 0.714 | 0.079 | 0.485 |
| Raven | 0.497 | 0.958 | 0.662 | 0.810 | 0.699 | 0.841 |
| **Raven, age** | 0.977 | 0.849 | 0.232 | 0.636 | 0.255 | 0.823 |

Table 69: p-values for z-test: Shudo vs. Hiroshima CU (excel male, econ.), No-info

| Covariates | 1 – 10 | 11 – 20 | 21 – 30 | 31 – 40 | 41 – 50 | 51 – 60 |
|---|---|---|---|---|---|---|
| **age** | 0.471 | 0.487 | 0.845 | 0.977 | 0.089 | 0.995 |
| Raven | 0.940 | **0.024** | 0.254 | **0.046** | 0.518 | 0.904 |
| **Raven, age** | 0.069 | **0.038** | 0.327 | 0.245 | 0.538 | **0.049** |

Table 70: p-values for z-test: Shudo vs. Hiroshima CU (excel male, econ.), Partial-info

| Covariates | 1 – 10 | 11 – 20 | 21 – 30 | 31 – 40 | 41 – 50 | 51 – 60 |
|---|---|---|---|---|---|---|
| **age** | 0.793 | 0.380 | 0.471 | 0.246 | 0.562 | 0.483 |
| Raven | 0.314 | 0.060 | 0.278 | 0.068 | 0.951 | 0.855 |
| **Raven, age** | 0.774 | 0.643 | 0.632 | 0.286 | 0.612 | 0.802 |

Table 71: p-values for z-test: Shudo vs. Hiroshima CU (excel male, econ.), Full-info

| Covariates | $1-10$ | $11-20$ | $21-30$ | $31-40$ | $41-50$ | $51-60$ |
|---|---|---|---|---|---|---|
| **age** | 0.200 | **0.037** | **0.000** | 0.100 | 0.897 | 0.903 |
| Raven | 0.611 | 0.371 | **0.007** | 0.555 | 0.894 | 0.161 |
| **Raven, age** | 0.183 | **0.006** | **0.020** | 0.081 | 0.318 | 0.611 |

## Kansai U-A vs. OSU: excl. female

Table 72: p-values for z-test: Kansai U-A vs. Osaka Sangyo. (excl female)

| Covariates | $1-10$ | $11-20$ | $21-30$ | $31-40$ | $41-50$ | $51-60$ |
|---|---|---|---|---|---|---|
| age | 0.145 | 0.526 | 0.914 | 0.680 | **0.001** | **0.008** |
| age, econ | 0.099 | 0.510 | 0.443 | 0.884 | **0.020** | **0.013** |
| age, sci-eng. | **0.033** | 0.226 | 0.505 | 0.935 | **0.006** | **0.017** |
| sci-eng., econ | 0.090 | 0.253 | 0.177 | 0.557 | 0.125 | 0.086 |
| **age, sci-eng., econ** | 0.375 | 0.460 | 0.436 | 0.872 | **0.005** | **0.030** |
| Raven | 0.094 | 0.966 | 0.979 | 0.452 | **0.034** | 0.121 |
| Raven, age | **0.038** | 0.345 | 0.638 | 0.811 | **0.002** | **0.014** |
| Raven, age, econ | 0.221 | 0.420 | 0.412 | 0.579 | **0.034** | 0.058 |
| Raven, age, sci-eng. | **0.001** | 0.157 | 0.336 | 0.969 | 0.138 | 0.267 |
| Raven, sci-eng., econ | 0.277 | 0.682 | 0.658 | 0.260 | **0.048** | 0.121 |
| **Raven, age, sci-eng., econ** | 0.379 | 0.401 | 0.293 | 0.875 | 0.168 | 0.602 |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

Table 73: p-values for z-test: Kansai U-A vs. Osaka Sangyo. (excl female): No-info.

| Covariates | $1-10$ | $11-20$ | $21-30$ | $31-40$ | $41-50$ | $51-60$ |
|---|---|---|---|---|---|---|
| age | 0.288 | 0.864 | 0.608 | 0.478 | 0.231 | 0.205 |
| age, econ | 0.063 | 0.458 | 0.483 | 0.827 | 0.418 | 0.269 |
| age, sci-eng. | 0.264 | 0.920 | 0.790 | 0.731 | 0.341 | 0.266 |
| sci-eng., econ | 0.071 | 0.491 | 0.420 | 0.861 | 0.232 | 0.155 |
| **age, sci-eng., econ** | 0.134 | 0.214 | 0.239 | 0.806 | 0.080 | **0.039** |
| Raven | **0.035** | 0.479 | 0.668 | 0.927 | 0.312 | 0.128 |
| Raven, age | **0.027** | 0.245 | 0.480 | 0.811 | 0.061 | **0.028** |
| Raven, age, econ | **0.005** | 0.111 | 0.054 | 0.294 | 0.511 | 0.221 |
| Raven, age, sci-eng. | 0.057 | 0.477 | 0.740 | 0.873 | 0.077 | **0.027** |
| Raven, sci-eng., econ | **0.006** | 0.200 | 0.289 | 0.980 | 0.135 | 0.126 |
| **Raven, age, sci-eng., econ** | **0.011** | **0.038** | **0.031** | 0.165 | 0.249 | 0.093 |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

Table 74: p-values for z-test: Kansai U-A vs. Osaka Sangyo. (excl female):Partial-info

| Covariates | $1-10$ | $11-20$ | $21-30$ | $31-40$ | $41-50$ | $51-60$ |
|---|---|---|---|---|---|---|
| age | 0.104 | 0.242 | 0.360 | 0.519 | 0.073 | 0.733 |
| age, econ | 0.337 | 0.504 | 0.265 | 0.650 | **0.043** | 0.232 |
| age, sci-eng. | 0.127 | 0.099 | **0.049** | 0.193 | 0.411 | 0.896 |
| sci-eng., econ | 0.284 | 0.476 | 0.093 | 0.661 | 0.060 | 0.472 |
| **age, sci-eng., econ** | 0.655 | 0.841 | 0.652 | 0.902 | 0.094 | 0.511 |
| Raven | 0.052 | 0.618 | **0.030** | 0.177 | **0.045** | 0.492 |
| Raven, age | 0.240 | 0.268 | **0.046** | 0.655 | 0.147 | 0.252 |
| Raven, age, sci-eng. | **0.020** | 0.120 | **0.009** | 0.321 | **0.025** | **0.035** |
| Raven, age, econ | 0.775 | 0.439 | 0.194 | 0.811 | 0.062 | 0.220 |
| Raven, sci-eng., econ | 0.090 | 0.472 | 0.070 | 0.455 | 0.055 | 0.085 |
| **Raven, age, sci-eng., econ** | 0.701 | 0.954 | 0.439 | 0.814 | **0.028** | 0.089 |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

Table 75: p-values for z-test: Kansai U-A vs. Osaka Sangyo. (excl female): Full-info.

| Covariates | $1-10$ | $11-20$ | $21-30$ | $31-40$ | $41-50$ | $51-60$ |
|---|---|---|---|---|---|---|
| age | 0.617 | 0.697 | 0.252 | 0.286 | **0.020** | **0.045** |
| age, econ | 0.760 | 0.867 | 0.454 | 0.412 | 0.064 | 0.115 |
| age, sci-eng. | 0.642 | 0.876 | 0.662 | 0.238 | 0.239 | 0.121 |
| sci-eng., econ | 0.287 | 0.663 | 0.605 | 0.303 | 0.479 | 0.149 |
| **age, sci-eng., econ** | 0.952 | 0.628 | 0.438 | 0.060 | 0.732 | 0.942 |
| Raven | 0.733 | 0.851 | 0.511 | 0.320 | 0.140 | 0.473 |
| Raven, age | 0.074 | 0.643 | 0.357 | 0.404 | 0.148 | **0.045** |
| Raven, age, econ | 0.262 | 0.675 | 0.827 | 0.723 | 0.305 | 0.120 |
| Raven, age, sci-eng. | 0.189 | 0.914 | 0.231 | 0.568 | 0.389 | 0.103 |
| Raven, sci-eng., econ | 0.979 | 0.535 | 0.910 | 0.172 | 0.865 | 0.556 |
| **Raven, age, sci-eng., econ** | 0.839 | 0.585 | 0.610 | 0.059 | 0.947 | 0.585 |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

## **HCU (excl. sci-eng) vs. Shudo (excel. econ)**

Table 76: p-values for z-test: HCU (excl. sci-eng) vs. Shudo (excel. econ)

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| age | 0.118 | 0.498 | 0.260 | 0.678 | 0.514 | 0.690 |
| gender | 0.146 | 0.291 | **0.047** | 0.537 | 0.552 | 0.528 |
| **age, gender** | 0.192 | 0.306 | 0.096 | 0.499 | 0.266 | 0.849 |
| Raven | 0.705 | 0.964 | 0.309 | 0.424 | 0.760 | 0.238 |
| Raven, age | 0.567 | 0.743 | 0.509 | 0.724 | 0.675 | 0.544 |
| Raven, gender | 0.218 | 0.188 | **0.024** | 0.681 | 0.985 | 0.441 |
| **Raven, age, gender** | 0.480 | 0.786 | 0.188 | 0.805 | 0.279 | 0.926 |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

Table 77: p-values for z-test: HCU (excl. sci-eng) vs. Shudo (excel. econ), No-info

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| age | **0.019** | 0.583 | 0.773 | 0.572 | 0.840 | 0.266 |
| gender | **0.036** | 0.581 | 0.964 | 0.866 | 0.893 | 0.064 |
| **age, gender** | 0.160 | 0.811 | 0.453 | 0.717 | 0.249 | 0.421 |
| Raven | 0.574 | 0.074 | 0.143 | 0.058 | 0.952 | 0.365 |
| Raven, age | **0.003** | 0.600 | 0.305 | 0.444 | 0.446 | **0.023** |
| Raven, gender | 0.379 | 0.374 | 0.786 | 0.277 | 0.797 | 0.182 |
| **Raven, age, gender** | 0.172 | 0.078 | 0.290 | 0.312 | 0.187 | 0.087 |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

Table 78: p-values for z-test: HCU (excl. sci-eng) vs. Shudo (excel. econ), Partial-info

| Covariates | $1-10$ | $11-20$ | $21-30$ | $31-40$ | $41-50$ | $51-60$ |
|---|---|---|---|---|---|---|
| age | 0.522 | 0.133 | 0.260 | 0.084 | 0.171 | 0.310 |
| gender | 0.383 | **0.048** | 0.146 | 0.169 | 0.239 | 0.219 |
| **age, gender** | 0.540 | 0.061 | 0.124 | 0.159 | 0.355 | 0.133 |
| Raven | 0.894 | 0.133 | 0.578 | 0.256 | 0.737 | 0.629 |
| Raven, age | 0.861 | 0.188 | 0.658 | 0.501 | 0.575 | 0.817 |
| Raven, gender | 0.324 | 0.141 | 0.311 | 0.256 | 0.785 | 0.964 |
| **Raven, age, gender** | 0.562 | 0.211 | 0.488 | 0.115 | 0.254 | 0.379 |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

Table 79: p-values for z-test: HCU (excl. sci-eng) vs. Shudo (excel. econ), Full-info

| Covariates | $1-10$ | $11-20$ | $21-30$ | $31-40$ | $41-50$ | $51-60$ |
|---|---|---|---|---|---|---|
| age | 0.906 | 0.365 | **0.004** | 0.563 | 0.819 | 0.718 |
| gender | 0.343 | 0.433 | **0.008** | 0.927 | 0.903 | 0.974 |
| **age, gender** | 0.875 | 0.318 | **0.007** | 0.685 | 0.985 | 0.860 |
| Raven | 0.398 | 0.827 | **0.003** | 0.400 | 0.796 | 0.982 |
| Raven, age | 0.591 | 0.408 | **0.005** | 0.819 | 0.759 | 0.901 |
| Raven, gender | 0.540 | 0.498 | **0.026** | 0.318 | 0.908 | 0.617 |
| **Raven, age, gender** | 0.679 | 0.464 | **0.018** | 0.995 | 0.749 | 0.896 |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

## Declarations

### Funding

This research was supported by a MEXT Grant-in-Aid (17K18573), Joint Usage/Research of RISS at Kansai University, and the Kansai University Fund for Supporting Young Scholars 2017 "Conducting Economic Experiments at Multiple Sites."

### Conflict of Interest

The authors declare that they have no conflict of interest.

### Availability of Data and Material

All raw and processed data are available upon requests.

### Code Availability

All data were processed with STATA and any codes were saved as ado files. Those codes are available upon requests.

### Ethical Approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committees and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This article does not contain any studies with animals performed by any of the authors.

## References


Aleskerov, F., Beliani, A., Pogorelskiy, K. 2009. Power and Preferences: An Experimental Approach. National Research University, Higher School of Economics, mimeo.

Arechar, A.A., Gächter, S., Molleman, L. 2018. Conducting Interactive Experiments Online *Experimental Economics* 21, 99-131

Azrieli, Y., Chambers, C.P., Healy, P.J. 2018. Incentives in Experiments: A Theoretical Analysis. *Journal Political Economics* 126, 1472-1503.

Banzhaf, J.F. 1965. Weighted Voting Doesn't Work: A Mathematical Analysis. *Rutgers Law Review* 19, 317-343.

Basteck, C., Mantovani, M. 2018. Cognitive Ability and Games of School Choice. *Games and Economic Behavior* 109, 156-183.

Brunner, E., Munzel, U. 2000. The Nonparametric Behrens-Fisher Problem: Asymptotic Theory and a Small-Sample Approximation. *Biometrical Journal* 42, 17-25.

Carpenter, P.A., Just, M.A., Shell, P. 1990. What One Intelligence Test Measures: A Theoretical Account of the Processing in the Raven Progressive Matrices Test. *Psychological Review* 97, 404-413.

Cason, T.N., Saijo, T., Yamato, T. 2002. Voluntary Participation and Spite in Public Good Provision Experiments: An International Comparison. *Experimental Economics* 5, 133-153.

Chen, D.L., Schnger, M., Wickens, C. 2016. oTree-An Open-Source Platform for Laboratory, Online, and Field Experiments. *Journal of Behavioral and Experimental Finance* 9, 88-97.

Cooper, D.J., Kagel, J.H. 2003. Lessons Learned: Generalized Learning across Games. *American Economic Review* 93, 202-207.

——— 2008. Learning and Transfer in Signaling Games. *Economic Theory* 34, 415-439.

Deegan, J., and Packel, E. 1978. A New Index of Power for Simple n-Person Games. *International Journal of Game Theory* 7, 113-123.

Dohmen, T., Falk, A., Huffman, D., Sunde, U. 2010. Are Risk Aversion and Impatience Related to Cognitive Ability? *American Economic Review* 100, 1238-1260.

Duch, M., Grossmann, M., Lauer, T. 2020. *z-Tree Unleashed: A Novel Client-Integrating Architecture for Conducting z-Tree Experiments over the Internet.* doi: 10.13140/RG.2.2.11801.36961

Dufwenberg, M., Sundaram, R., Butler, D. J. 2010. Epiphany in the Game of 21. *Journal of Economic Behavior and Organization* 75, 132-143.

Engel, C. 2011. Dictator Games: a Meta Study. *Experimental Economics* 14, 583-610.

Ernst, M.D. 2004. Permutation Methods: A Basis for Exact Inference. *Statistical Science* 19, 676-685.

Esposito, G., Guerci, E., Lu, X., Hanaki, N., Watanabe, N. 2012. An Experimental Study on "Meaningful Learning" in Weighted Voting Games. mimeo., Aix-Marseille University.

Falk, A., Becker, A. Dohmen, T., Enke, B., Huffman, D., Sunde, U. 2018. Global Evidence on Economic Preferences. *Quarterly Journal of Economics* 133, 1645-1692.

Felsenthal, D.S., Machover, M. 1998. *The Measurement of Voting Power : Theory and Practice, Problems and Paradoxes.* Edward Elgar, London.

Fischbacher, U. 2007. z-Tree: Zurich Toolbox for Ready-made Economic Experiments. *Experrimental Econimics* 10, 171-178.

Fisher, R.A. 1935. *The Design of Experiments*, 1st ed., (Hafner Pub. Co., 1971).

Gill, D., Prowse, V. 2016. Cognitive Ability, Character Skills, and Learning to Play Equilibrium: A Level-k Analysis. *Journal of Political Economics* 124, 1619-1676.

Greiner, B. 2003. Online Recruitment System for Economic Experiments (ORSEE). `http://www.orsee.org/web`

Guerci, E., Hanaki, N., Watanabe, N. 2017. Meaningful Learning in Weighted Voting Games: An Experiment. *Theory and Decision* 83, 131-153.

Guerci, E., Hanaki, N., Watanabe, N., Esposito, G., Lu, X. 2014. A Methodological Note on a Weighted Voting Experiment. *Social Choice and Welfare* 43, 827-850.

Hanaki, N. Cognitive Ability and Observed Behavior in Laboratory Experiments: Implications for Macroeconomic Theory. *Japanese Economic Review* 71, 355-378.

Hanaki, N., Jacquemet, N., Luchini, S., Zylbersztejn, A. 2016. Cognitive Ability and the Effect of Strategic Uncertainty. *Theory and Decision* 81, 101-121.

Hanaki, N., Jacquemet, N., Luchini, S., Zylbersztejn, A. 2016. Fluid Intelligence and Cognitive Reflection in a Strategic Environment: Evidence from Dominance-Solvable Game. *Frontiers in Psychology: Personality and Social Psychology* 10 August 2016. http://dx.doi.org/10.3389/fpsyg.2016.01188

Heckman, J.J., Ichimura, H., Smith, J., Todd, P. 1998. Characterizing Selection Bias Using Experimental Data. *Econometrica* 66, 1017-1098.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., 2004. *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*. Oxford University Press.

Hinkelmann, K., Kempthone, O., 2005. *Design and Analysis of Experiments, Volume 2, Advanced Experimental Design*. John Wiley & Sons, Inc.

Holm, H.J., Nee, V., Opper, S. 2020, Strategic Decisions: Behavioral Differences between CEOs and Others. *Experimental Economics* 23, 154-180.

Kawamura, T., Ogawa, K. 2019. Cognitive Ability and Human Behavior in Experimental Ultimatum Games. *Research in Economics* 73, 97-106.

Montero, M., Sefton, M., Zhang, P. 2008. Enlargement and the Balance of Power: An Experimental Study. *Social Choice and Welfare* 30, 69-87.

Neubert, K., Brunner, E. 2007. A Studentized Permutation Test for the Non-Parametric Behrens-Fisher Problem. *Computational Statistics* & *Data Analysis* 51, 5192-5204.

Nowak, M,A., Sigmund, K. 1993. A Strategy of Win Stay, Lose Shift that Outperforms Tit-for-Tat in the Prisoner's-Dilemma Game. *Nature* 364, 56-58.

Ogawa, K., Osaki, Y., Kawamura, T., Takahashi, H., Taguchi, S., Fujii, Y., Watanabe, N. 2020. Conducting Economic Experiments at Multiple Sites: Subjects' Cognitive Ability and Attribute Information. the Kansai University RISS Discussion Paper Series No. 86.

Proto, E., Rustichini, A., Sofianos, A. 2019. Intelligence, Personality and Gains from Cooperation in Repeated Interactions. *Journal Political Economy* 127; 1351-1390.

Rick, S., Weber, R.A. 2010. Meaningful Learning and Transfer of Learning in Games Played Repeatedly without Feedback. *Games and Economic Behavior* 68, 716-730.

Rosenbaum, P.R., Rubin, D.B. 1983. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrica* 70; 41-55.

Rosser, J.B. Jr., Eckel, C. 2010. Special Issue on the Methodology of Experimental Economics. Rosser, J.B. Jr., Eckel, C. eds., *Journal of Economic Behavior and Organization* 73, 1-132.

Roth, A. E., Prasnikar, V., Okuno-Fujiwara, M., Zamir, S. 1991. Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study. *American Economic Review* 81, 1068-1095.

Snowberg, E., Yariv, L. 2021. Testing the Waters: Behavior across Participant Pools. *American Economic Review* 11, 687-719.

Shapley, L.S., Shubik, M. 1954. A Method for Evaluating the Distribution of Power in a Committee System. *American Political Science Review* 48, 787-792.

Shimomura, K., Yamato, T. 2012. Impact of Ethnicities on Market Outcome. In Hino, H., J. Lonsdale, J., Ranis, G., Stewart, F. eds., *Ethnic Diversity and Economic Instability in Africa: Interdisciplinary Perspectives* 286-313. Cambridge, Cambridge University Press.

Sutton, A.J, Abrams, K.R., Jones, D.R., Sheldon, T.A., Song, F. 2000. *Methods for Meta-Analysis in Medical Research*. Chichester, Wiley.

Watanabe, N. 2014. Coalition Formation in a Weighted Voting Experiment. *Japanese Journal of Electral Studies* 30, 56-67.

Watanabe, N. 2018. Supplementary Results for “Meaningful Learning in Weighted Voting Games: An Experiment”. Kansai University RISS Discussion Paper Series No. 62.

Watanabe, N., Kawamura, T., Ogawa, K. 2020. A School Choice Experiment: Cognitive Ability and Information. mimeo.

Weber, R.A. 2003. Learning with No Feedback in a Competitive Guessing Game. *Games and Economic Behavior* 44, 134-144.

## Appendix A: Session Details

Dates and other details at each site are listed below. The tables also list feedback, the order of binary choice problems (order), the number of subjects (# of subj.), the average, maximum, and minimum amounts of payment (avg. pay, max, and min) in Japanese yen (JPY) and the standard deviation of payment (std.dev.), where decimals are rounded off.

Table 80: Sessions at Kansai U-A

| date | feedback | order | # of subj. | avg. pay | max | min | std.dev. |
|---|---|---|---|---|---|---|---|
| 2019/02/16 | Part-fb | A→B | 20 | 2654 | 2840 | 2400 | 132 |
| 2019/02/16 | Part-fb | B→A | 20 | 2470 | 2840 | 2120 | 213 |
| 2019/02/26 | Full-fb | A→B | 20 | 2572 | 3040 | 2160 | 211 |
| 2019/02/26 | Full-fb | B→A | 20 | 2452 | 2800 | 2200 | 158 |
| 2019/03/08 | No-fb | A→B | 20 | 2574 | 2840 | 2240 | 151 |
| 2019/03/08 | No-fb | B→A | 20 | 2454 | 3080 | 2080 | 259 |
| 2019/06/13 | No-fb | C→D | 20 | 2532 | 2960 | 2320 | 176 |
| 2019/06/13 | No-fb | D→C | 20 | 2358 | 2920 | 2000 | 257 |
| 2019/06/15 | Part-fb | C→D | 20 | 2520 | 2840 | 2240 | 138 |
| 2019/06/15 | Part-fb | D→C | 20 | 2438 | 2680 | 2160 | 157 |
| 2019/07/11 | Full-fb | C→D | 20 | 2526 | 2840 | 2160 | 193 |
| 2019/07/11 | Full-fb | D→C | 20 | 2384 | 2720 | 2120 | 164 |
| Total | | | 240 | 2495 | 3080 | 2000 | 205 |

Table 81: Sessions at Osaka SU

| date | feedback | order | # of subj. | avg. | max | min | std.dev. |
|---|---|---|---|---|---|---|---|
| 2018/11/19 | No-fb | A→B | 10 | 2456 | 2800 | 2080 | 167 |
| 2018/11/19 | Part-fb | | 10 | 2464 | 2560 | 2360 | 57 |
| 2018/11/19 | Full-fb | | 10 | 2596 | 2840 | 2280 | 187 |
| 2018/11/20 | No-fb | B→A | 10 | 2336 | 2600 | 2000 | 174 |
| 2018/11/20 | Part-fb | | 10 | 2516 | 2840 | 2320 | 185 |
| 2018/11/20 | Full-fb | | 10 | 2444 | 2720 | 2120 | 164 |
| 2018/12/03 | No-fb | C→D | 10 | 2516 | 2840 | 2120 | 203 |
| 2018/12/03 | Part-fb | | 10 | 2596 | 2760 | 2360 | 117 |
| 2018/12/03 | Full-fb | | 10 | 2540 | 2920 | 2120 | 224 |
| 2018/12/10 | No-fb | D→C | 10 | 2372 | 2760 | 2000 | 230 |
| 2018/12/10 | Part-fb | | 10 | 2448 | 2640 | 2280 | 130 |
| 2018/12/10 | Full-fb | | 10 | 2488 | 2680 | 2200 | 160 |
| Total | | | 120 | 2481 | 2920 | 2000 | 188 |

Table 82: Sessions at Kansai U-B

| date | feedback | order | # of subj. | avg. | max | min | std.dev. |
|---|---|---|---|---|---|---|---|
| 2018/03/02 | Full-fb | A→B | 17 | 2755 | 3050 | 2390 | 156 |
| 2018/03/02 | Full-fb | | 12 | 2664 | 2950 | 2370 | 211 |
| 2018/03/05 | No-fb | A→B | 9 | 2599 | 3000 | 2250 | 242 |
| 2018/03/05 | No-fb | | 13 | 2718 | 3010 | 2260 | 203 |
| 2018/05/16 | Part-fb | A→B | 14 | 2592 | 2860 | 2250 | 171 |
| 2018/05/16 | Part-fb | | 13 | 2635 | 3040 | 2180 | 237 |
| 2018/05/21 | Full-fb | B→A | 12 | 2558 | 2750 | 2250 | 155 |
| 2018/05/21 | Full-fb | | 11 | 2541 | 2850 | 2200 | 207 |
| 2018/06/20 | No-fb | B→A | 8 | 2450 | 2600 | 2230 | 139 |
| 2018/10/03 | Part-fb | B→A | 13 | 2517 | 2780 | 2230 | 144 |
| 2018/10/03 | Part-fb | | 13 | 2539 | 3030 | 2310 | 198 |
| 2019/05/27 | Full-fb | B→A | 19 | 2419 | 2770 | 1920 | 243 |
| 2019/11/22 | No-fb | A→B | 23 | 2128 | 2420 | 1820 | 158 |
| Total | | | 197 | 2586 | 3050 | 1900 | 222 |

Table 83: Sessions at Doshisha Univ.

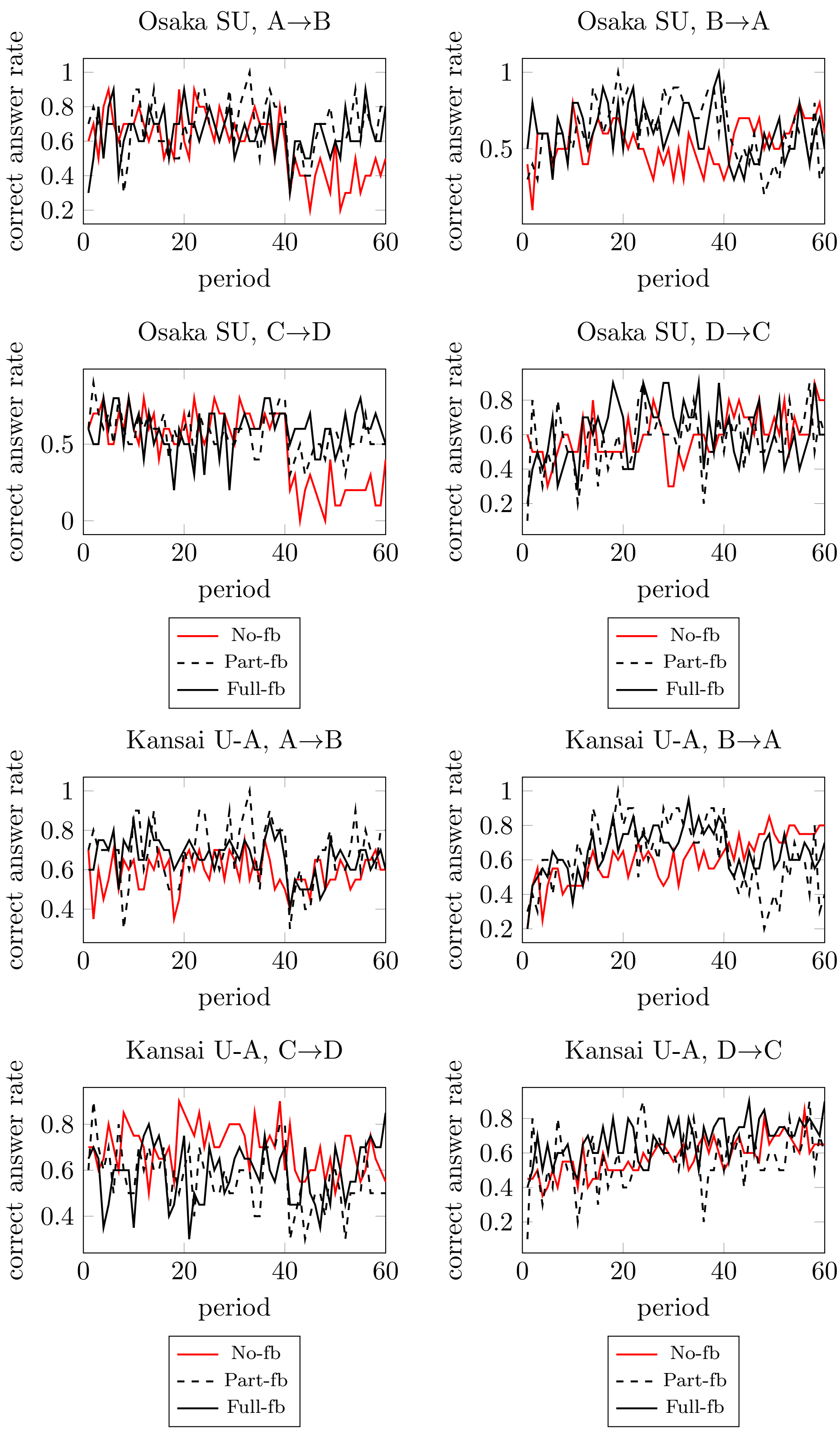
| date | feedback | order | # of subj. | avg. | max | min | std.dev. |
|---|---|---|---|---|---|---|---|
| 2018/07/13 | Full-fb | A→B | 9 | 2606 | 2960 | 2060 | 283 |
| 2018/07/13 | Full-fb | | 11 | 2734 | 2980 | 2440 | 159 |
| 2018/11/02 | No-fb | | 12 | 2721 | 2960 | 2380 | 151 |
| 2018/11/02 | No-fb | | 15 | 2532 | 2840 | 2110 | 230 |
| 2018/11/01 | Part-fb | | 15 | 2714 | 2960 | 2450 | 134 |
| 2018/11/01 | Part-fb | | 13 | 2687 | 3080 | 2460 | 198 |
| 2019/07/12 | Full-fb | B→A | 10 | 2372 | 2790 | 1750 | 278 |
| 2019/07/12 | Full-fb | | 10 | 2460 | 2670 | 2250 | 122 |
| 2019/07/12 | No-fb | | 10 | 2406 | 2680 | 2110 | 171 |
| 2019/07/18 | No-fb | | 10 | 2596 | 2960 | 2260 | 232 |
| 2019/10/17 | Part-fb | | 10 | 2589 | 2730 | 2170 | 164 |
| 2019/10/17 | Part-fb | | 10 | 2479 | 2710 | 2180 | 169 |
| Total | | | 135 | 2584 | 3080 | 1750 | 229 |

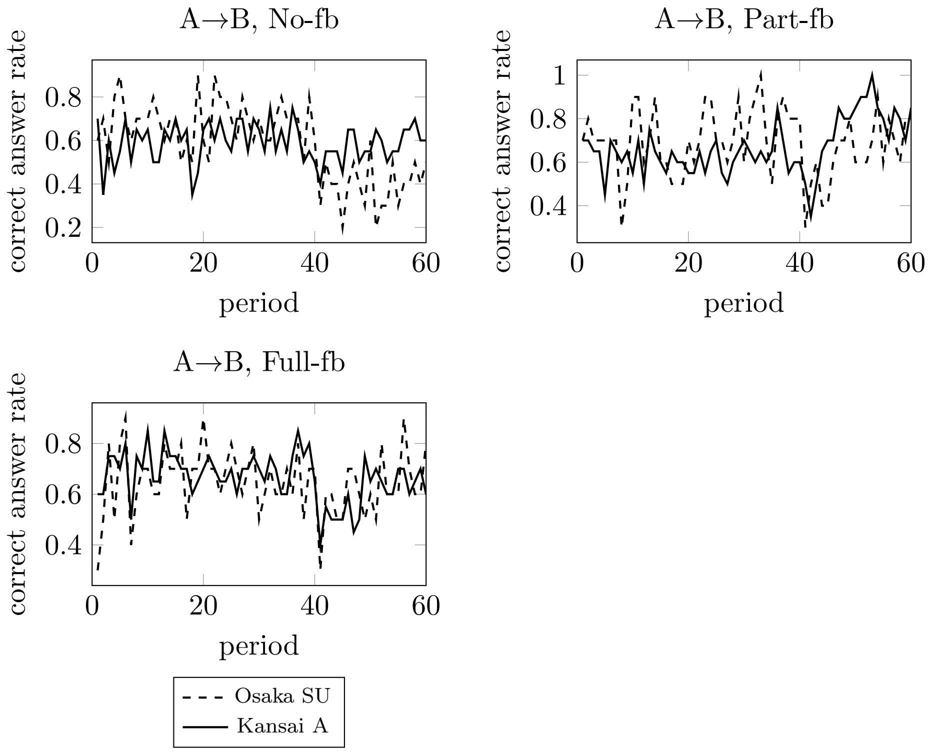
Table 84: Sessions at Hiroshima CU

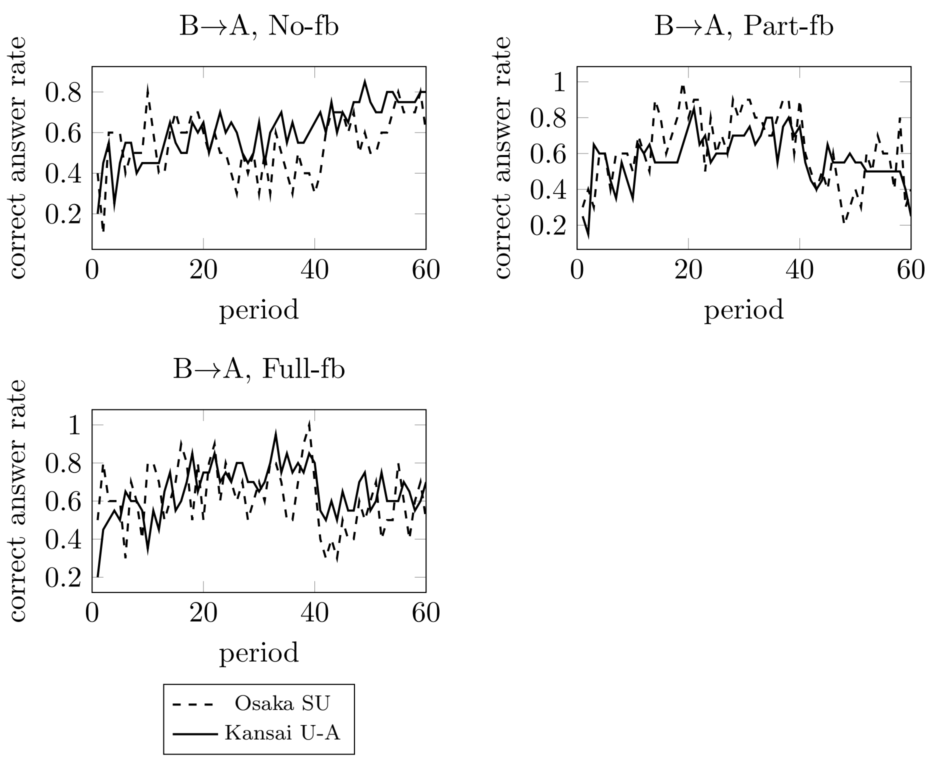
| date | feedback | order | # of subj. | avg. | max | min | std.dev. |
|---|---|---|---|---|---|---|---|
| 2018/05/29 | Full-fb | A→B | 10 | 2709 | 2930 | 2460 | 172 |
| 2018/05/30 | Full-fb | | 11 | 2628 | 3030 | 2070 | 241 |
| 2018/07/04 | No-fb | A→B | 10 | 2602 | 3080 | 2230 | 246 |
| 2018/07/25 | No-fb | | 10 | 2484 | 2760 | 2310 | 140 |
| 2018/10/31 | Part-fb | A→B | 14 | 2688 | 2970 | 2220 | 194 |
| 2018/10/31 | Part-fb | | 9 | 2508 | 3060 | 2200 | 231 |
| 2018/11/28 | Part-fb | B→A | 10 | 2364 | 2630 | 2080 | 181 |
| 2018/11/28 | Part-fb | | 10 | 2554 | 2770 | 2240 | 178 |
| 2019/05/29 | Full-fb | B→A | 10 | 2448 | 2700 | 1930 | 216 |
| 2019/05/29 | Full-fb | | 10 | 2584 | 3010 | 2140 | 232 |
| 2019/06/12 | No-fb | B→A | 10 | 2337 | 3040 | 1810 | 350 |
| 2019/06/12 | No-fb | | 10 | 2272 | 2520 | 2040 | 179 |
| Total | | | 124 | 2521 | 3080 | 1810 | 257 |

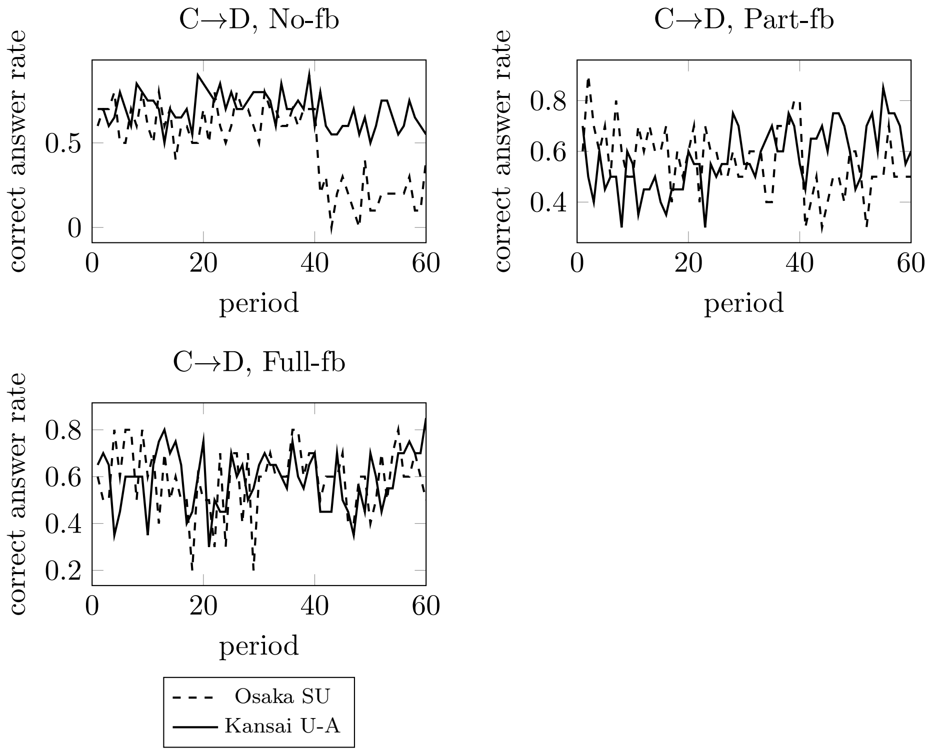
## Appendix B: Time Series Plots

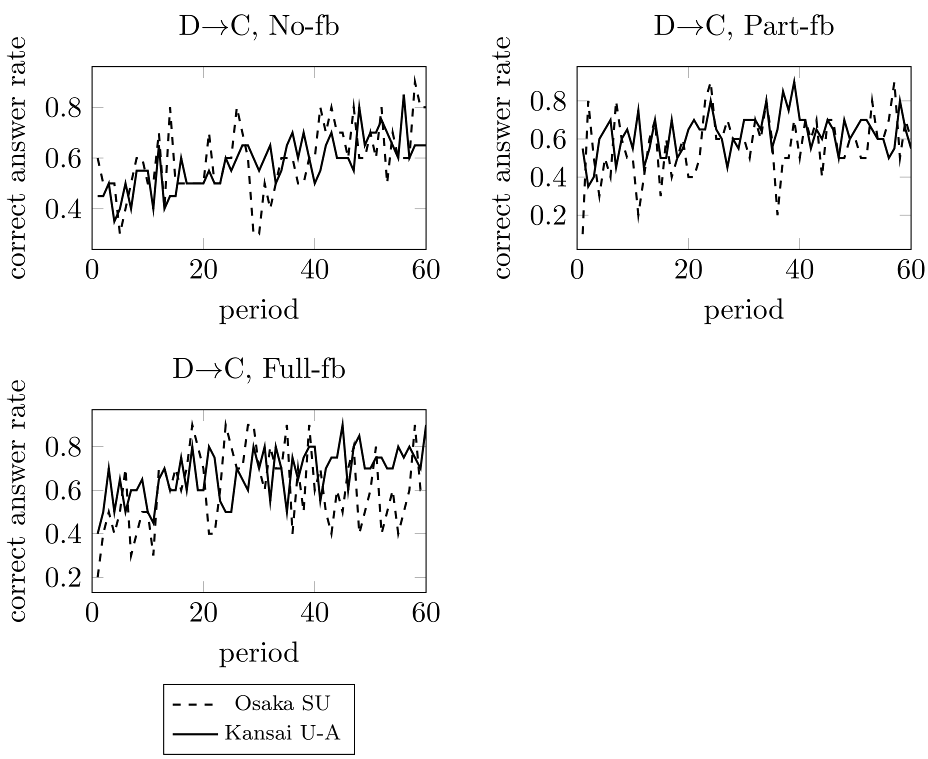
Below are the time series plots of the rates of correct answers subjects chose for each treatment at each site. In each figure, the horizontal axis stands for the periods from the 1st period to the 60th period, and the vertical axis represents the percentage of correct answers in each period.

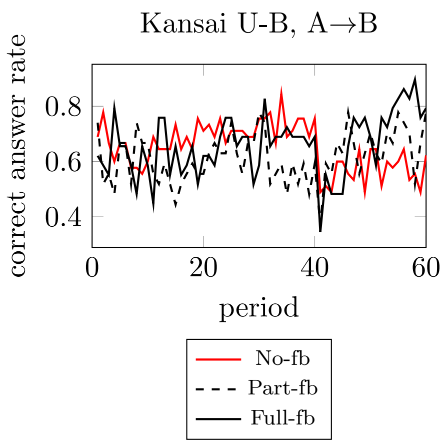

Osaka SU, A→B
Osaka SU, B→A
Osaka SU, C→D
Osaka SU, D→C
Kansai U-A, A→B
Kansai U-A, B→A
Kansai U-A, C→D
Kansai U-A, D→C
correct answer rate
period
No-fb
Part-fb
Full-fb

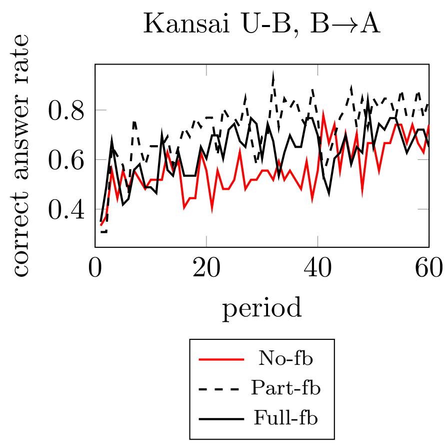

A→B, No-fb
A→B, Part-fb
A→B, Full-fb
correct answer rate
period
Osaka SU
Kansai A

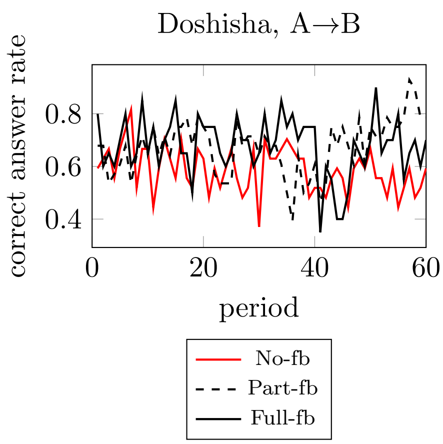

B→A, No-fb
B→A, Part-fb
B→A, Full-fb
correct answer rate
period
Osaka SU
Kansai U-A

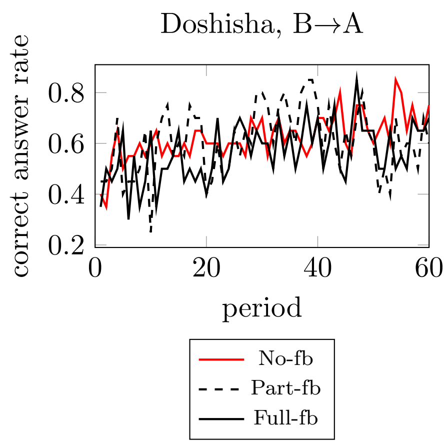

C→D, No-fb
C→D, Part-fb
C→D, Full-fb
correct answer rate
period
0
20
40
60
0.5
0.2
0.4
0.6
0.8
Osaka SU
Kansai U-A

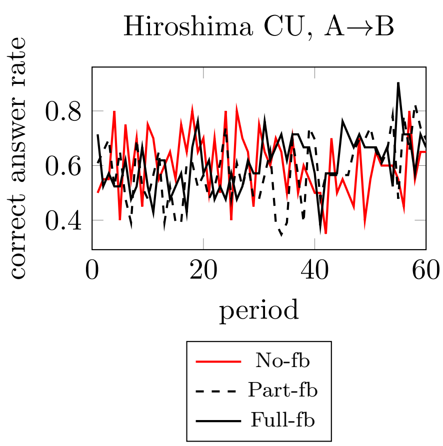

D→C, No-fb
D→C, Part-fb
D→C, Full-fb
correct answer rate
period
0
20
40
60
0.2
0.4
0.6
0.8
Osaka SU
Kansai U-A

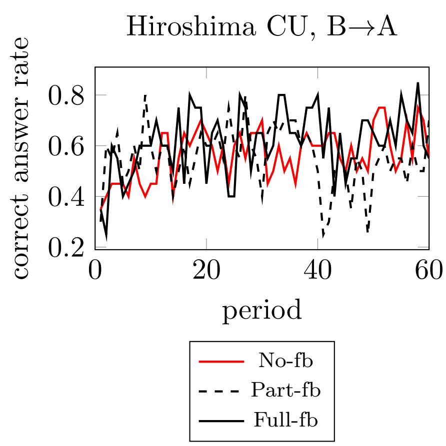

Kansai U-B, A→B
correct answer rate
0.4
0.6
0.8
0
20
40
60
period
No-fb
Part-fb
Full-fb

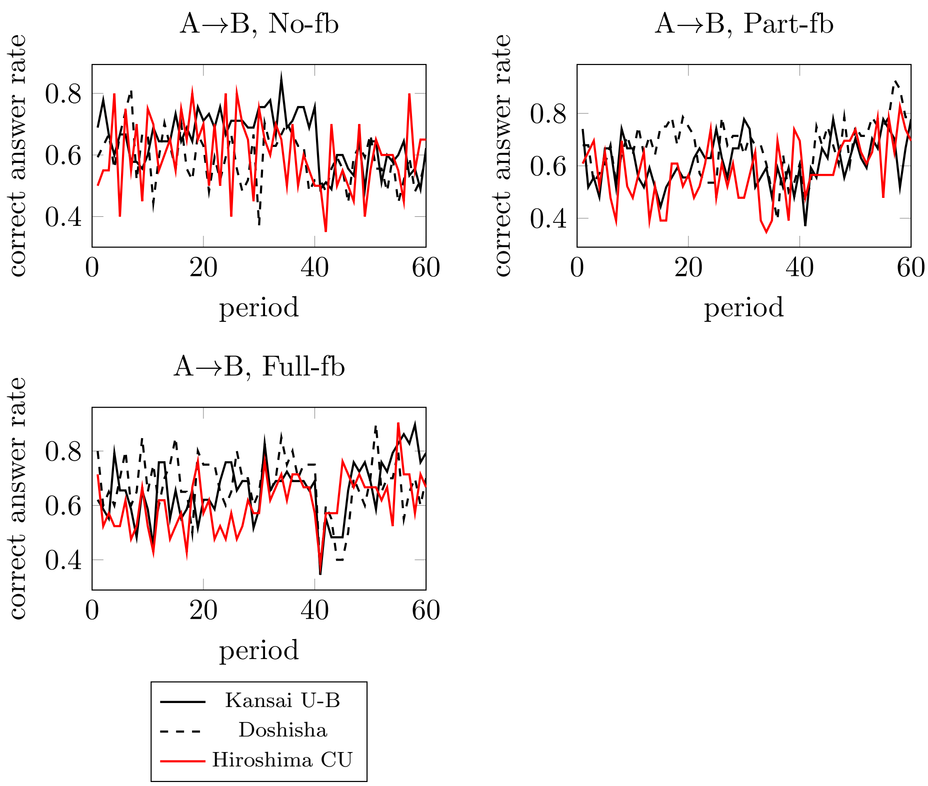

Kansai U-B, B→A
correct answer rate
0.4
0.6
0.8
0
20
40
60
period
No-fb
Part-fb
Full-fb

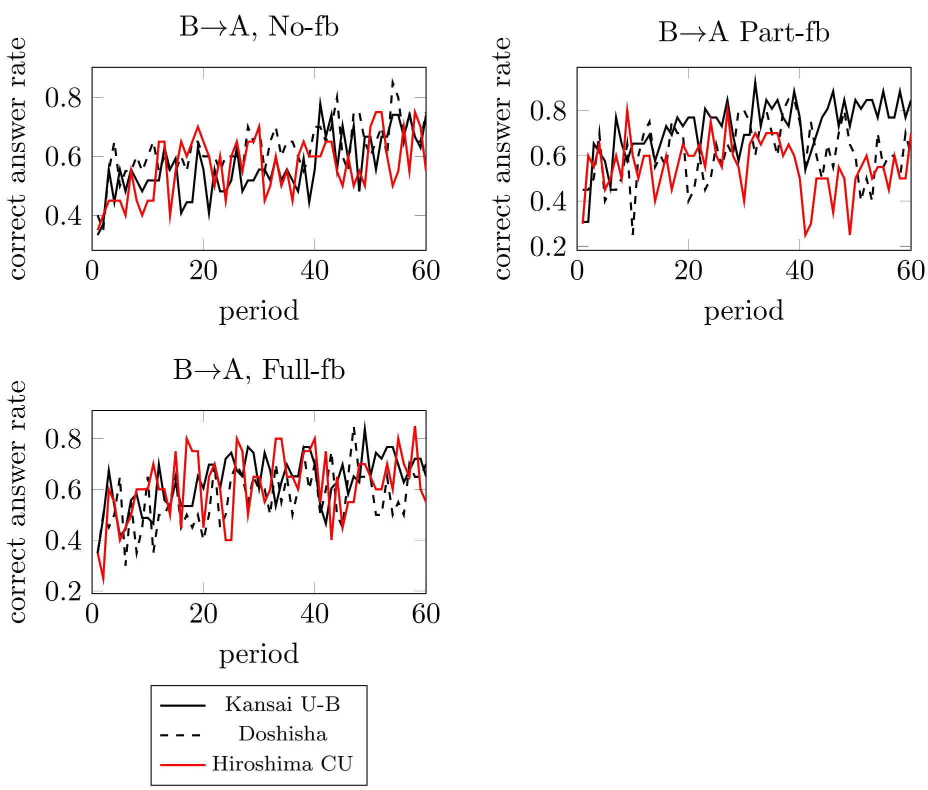

Doshisha, A→B
correct answer rate
0.4
0.6
0.8
0
20
40
60
period
No-fb
Part-fb
Full-fb


Doshisha, B→A
correct answer rate
0.2
0.4
0.6
0.8
0
20
40
60
period
No-fb
Part-fb
Full-fb


Hiroshima CU, A→B
correct answer rate
0.4
0.6
0.8
0
20
40
60
period
No-fb
Part-fb
Full-fb


Hiroshima CU, B→A
correct answer rate
0.2
0.4
0.6
0.8
0
20
40
60
period
No-fb
Part-fb
Full-fb


A→B, No-fb
A→B, Part-fb
A→B, Full-fb
correct answer rate
period
0.4
0.6
0.8
0
20
40
60
Kansai U-B
Doshisha
Hiroshima CU


B→A, No-fb
B→A Part-fb
B→A, Full-fb
correct answer rate
period
0.2
0.4
0.6
0.8
0
20
40
60
Kansai U-B
Doshisha
Hiroshima CU

## Appendix C: Propensity Score Matching: Kansai U-B (excl. sci-eng) vs. Doshisha

Some tables in this Appendix list the p-values for the z-test, where the null hypothesis is that subjects behavior is homogenized between Kansai U-B and Doshisha, when science or engineering major students were excluded from the Kansai U-B. Difference in Raven score is also noted for reference in other tables.

Table 85: p-values for z-test: Kansai U-B (excl. sci-eng) vs. Doshisha, integrated data

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| age | 0.758 | 0.967 | 0.275 | 0.252 | 0.792 | 0.627 |
| gender | 0.708 | 0.673 | 0.148 | 0.231 | 0.941 | 0.220 |
| age, gender | 0.622 | 0.833 | 0.713 | 0.218 | 0.770 | 0.874 |
| age, econ | 0.980 | 0.733 | 0.280 | 0.203 | 0.898 | 0.596 |
| gender, econ | 0.726 | 0.793 | 0.150 | 0.213 | 0.946 | 0.213 |
| **age, gender, econ** | 0.503 | 0.942 | 0.954 | 0.359 | 0.732 | 0.931 |
| Raven | 0.841 | 0.575 | 0.185 | 0.422 | 0.955 | 0.267 |
| Raven, age | 0.691 | 0.765 | 0.323 | 0.925 | 0.910 | 0.668 |
| Raven, gender | 0.710 | 0.724 | 0.133 | 0.294 | 0.965 | 0.234 |
| Raven, age, gender | 0.679 | 0.634 | 0.864 | 0.703 | 0.905 | 0.336 |
| Raven, age, econ | 0.873 | 0.756 | 0.362 | 0.655 | 0.730 | 0.617 |
| Raven, gender, econ | 0.906 | 0.951 | 0.156 | 0.286 | 0.920 | 0.081 |
| **Raven, age, gender, econ** | 0.867 | 0.807 | 0.738 | 0.451 | 0.758 | 0.324 |

Table 86: p-values for z-test: Kansai U-B (excl. sci-eng) vs. Doshisha, No-fb

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| age | 0.376 | 0.259 | 0.375 | 0.346 | 0.726 | 0.485 |
| gender | 0.446 | 0.305 | 0.657 | 0.454 | 0.925 | 0.768 |
| age, gender | 0.779 | 0.633 | 0.789 | 0.492 | 0.461 | 0.177 |
| age, econ | 0.423 | 0.313 | 0.513 | 0.336 | 0.650 | 0.485 |
| gender, econ | 0.266 | 0.605 | 0.994 | 0.595 | 0.922 | 0.865 |
| **age, gender, econ** | 0.796 | 0.649 | 0.868 | 0.467 | 0.557 | 0.210 |
| Raven | 0.316 | 0.532 | 0.588 | 0.290 | 0.843 | 0.786 |
| Raven, age | 0.352 | 0.657 | 0.954 | 0.900 | 0.575 | 0.813 |
| Raven, gender | 0.223 | 0.384 | 0.913 | 0.681 | 0.668 | 0.404 |
| Raven, age, gender | 0.632 | 0.993 | 0.844 | 0.878 | 0.414 | 0.235 |
| Raven, age, econ | 0.192 | 0.391 | 0.825 | 0.597 | 0.580 | 0.696 |
| Raven, gender, econ | 0.115 | 0.761 | 0.917 | 0.808 | 0.903 | 0.687 |
| **Raven, age, gender, econ** | 0.707 | 0.548 | 0.910 | 0.627 | 0.530 | 0.150 |

Table 87: p-values for z-test: Kansai U-B (excl. sci-eng) vs. Doshisha, Part-fb

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| age | 0.105 | 0.539 | 0.102 | 0.670 | 0.691 | 0.504 |
| gender | 0.409 | 0.112 | 0.118 | 0.401 | 0.701 | 0.475 |
| age, gender | 0.445 | 0.236 | 0.582 | 0.968 | 0.943 | 0.685 |
| age, econ | 0.089 | 0.661 | 0.080 | 0.612 | 0.650 | 0.433 |
| gender, econ | 0.359 | 0.142 | 0.087 | 0.402 | 0.519 | 0.413 |
| **age, gender, econ** | 0.282 | 0.202 | 0.492 | 0.777 | 0.926 | 0.577 |
| Raven | 0.554 | 0.078 | 0.096 | 0.795 | 0.765 | 0.620 |
| Raven, age | 0.169 | 0.834 | **0.002** | 0.070 | 0.152 | 0.053 |
| Raven, gender | 0.571 | 0.135 | 0.152 | 0.382 | 0.869 | 0.553 |
| Raven, age, gender | 0.175 | 0.380 | 0.093 | 0.270 | 0.800 | 0.174 |
| Raven, age, econ | 0.217 | 0.676 | **0.005** | 0.105 | 0.232 | 0.102 |
| Raven, gender, econ | 0.859 | **0.033** | 0.257 | 0.438 | 0.595 | 0.520 |
| **Raven, age, gender, econ** | 0.233 | 0.118 | 0.108 | 0.260 | 0.859 | 0.433 |

Table 88: p-values for z-test: Kansai U-B (excl. sci-eng) vs. Doshisha, Full-fb

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| age | 0.464 | 0.620 | 0.324 | 0.801 | 0.825 | 0.118 |
| gender | 0.396 | 0.790 | 0.582 | 0.982 | 0.594 | 0.089 |
| age, gender | 0.801 | 0.869 | 0.412 | 0.383 | 0.905 | 0.159 |
| age, econ | 0.168 | 0.231 | 0.427 | 0.883 | 0.941 | 0.081 |
| gender, econ | 0.222 | 0.732 | 0.641 | 0.831 | 0.612 | 0.096 |
| **age, gender, econ** | 0.589 | 0.888 | 0.560 | 0.554 | 0.449 | 0.102 |
| Raven | 0.792 | 0.777 | 0.447 | 0.757 | 0.373 | **0.024** |
| Raven, age | 0.448 | 0.912 | 0.312 | 0.968 | 0.578 | 0.979 |
| Raven, gender | 0.979 | 0.730 | 0.329 | 0.944 | 0.711 | 0.136 |
| Raven, age, gender | 0.928 | 0.802 | 0.658 | 0.611 | 0.384 | 0.272 |
| Raven, age, econ | 0.913 | 0.989 | 0.123 | 0.529 | 0.397 | 0.076 |
| Raven, gender, econ | 0.650 | 0.725 | 0.250 | 0.953 | 0.503 | 0.084 |
| **Raven, age, gender, econ** | 0.845 | 0.948 | 0.393 | 0.667 | 0.902 | 0.193 |

## HCU (excl. sci-eng) vs. Shudo

Table 89: p-values for z-test: HCU (excl. sci-eng) vs. Shudo, No-info

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| age | **0.048** | 0.500 | 0.395 | 0.177 | 0.433 | 0.104 |
| gender | 0.112 | 0.799 | 0.877 | 0.703 | 0.952 | **0.004** |
| **age, gender** | **0.012** | 0.825 | 0.730 | 0.939 | 0.933 | **0.024** |
| Raven | 0.091 | 0.218 | 0.130 | 0.094 | 0.826 | 0.059 |
| Raven, age | **0.025** | 0.619 | 0.115 | **0.023** | 0.408 | 0.475 |
| Raven, gender | 0.135 | 0.989 | 0.995 | 0.650 | 0.821 | **0.006** |
| **Raven, age, gender** | 0.152 | 0.945 | 0.704 | 0.262 | 0.649 | **0.013** |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

Table 90: p-values for z-test: HCU (excl. sci-eng) vs. Shudo, Partial-info

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| age | 0.821 | 0.221 | 0.372 | 0.221 | 0.864 | 0.714 |
| gender | 0.652 | 0.085 | 0.237 | 0.265 | 0.426 | 0.414 |
| **age, gender** | 0.474 | 0.417 | 0.232 | 0.177 | 0.703 | 0.720 |
| Raven | 0.861 | 0.402 | 0.544 | 0.361 | 0.902 | 0.810 |
| Raven, age | 0.983 | 0.132 | 0.213 | 0.350 | 0.688 | 0.628 |
| Raven, gender | 0.644 | 0.387 | 0.450 | 0.571 | 0.931 | 0.864 |
| **Raven, age, gender** | 0.821 | 0.200 | 0.090 | 0.204 | 0.998 | 0.519 |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

Table 91: p-values for z-test: HCU (excl. sci-eng) vs. Shudo, Full-info

| Covariates | $1-10$ | $11-20$ | $21-30$ | $31-40$ | $41-50$ | $51-60$ |
|---|---|---|---|---|---|---|
| age | 0.823 | 0.785 | 0.252 | 0.729 | 0.592 | 0.929 |
| gender | 0.595 | 0.562 | **0.021** | 0.816 | 0.978 | 0.933 |
| **age, gender** | 0.623 | 0.654 | 0.091 | 0.304 | 0.964 | 0.874 |
| Raven | 0.515 | 0.847 | 0.050 | 0.706 | 0.541 | 0.725 |
| Raven, age | 0.968 | 0.082 | **0.016** | 0.786 | 0.993 | 0.987 |
| Raven, gender | 0.951 | 0.744 | **0.003** | 0.721 | 0.898 | 0.570 |
| **Raven, age, gender** | 0.816 | 0.758 | **0.023** | 0.909 | 0.934 | 0.608 |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

## HCU (excl. sci-eng) vs. Shudo (excl. econ)

Table 92: p-values for z-test: HCU (excl. sci-eng) vs. Shudo (excel. econ), No-info

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| age | **0.019** | 0.583 | 0.773 | 0.572 | 0.840 | 0.266 |
| gender | **0.036** | 0.581 | 0.964 | 0.866 | 0.893 | 0.064 |
| **age, gender** | 0.160 | 0.811 | 0.453 | 0.717 | 0.249 | 0.421 |
| Raven | 0.574 | 0.074 | 0.143 | 0.058 | 0.952 | 0.365 |
| Raven, age | **0.003** | 0.600 | 0.305 | 0.444 | 0.446 | **0.023** |
| Raven, gender | 0.379 | 0.374 | 0.786 | 0.277 | 0.797182 | 0. |
| **Raven, age, gender** | 0.172 | 0.078 | 0.290 | 0.312 | 0.187 | 0.087 |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

Table 93: p-values for z-test: HCU (excl. sci-eng) vs. Shudo (excel. econ), Partial-info

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| age | 0.522 | 0.133 | 0.260 | 0.084 | 0.171 | 0.310 |
| gender | 0.383 | **0.048** | 0.146 | 0.169 | 0.239 | 0.219 |
| **age, gender** | 0.540 | 0.061 | 0.124 | 0.159 | 0.355 | 0.133 |
| Raven | 0.894 | 0.133 | 0.578 | 0.256 | 0.737 | 0.629 |
| Raven, age | 0.861 | 0.188 | 0.658 | 0.501 | 0.575 | 0.817 |
| Raven, gender | 0.324 | 0.141 | 0.311 | 0.256 | 0.785 | 0.964 |
| **Raven, age, gender** | 0.562 | 0.211 | 0.488 | 0.115 | 0.254 | 0.379 |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

Table 94: p-values for z-test: HCU (excl. sci-eng) vs. Shudo (excel. econ), Full-info

| Covariates | $1-10$ | $11-20$ | $21-30$ | $31-40$ | $41-50$ | $51-60$ |
|---|---|---|---|---|---|---|
| age | 0.906 | 0.365 | **0.004** | 0.563 | 0.819 | 0.718 |
| gender | 0.343 | 0.433 | **0.008** | 0.927 | 0.903 | 0.974 |
| **age, gender** | 0.875 | 0.318 | **0.007** | 0.685 | 0.985 | 0.860 |
| Raven | 0.398 | 0.827 | **0.003** | 0.400 | 0.796 | 0.982 |
| Raven, age | 0.591 | 0.408 | **0.005** | 0.819 | 0.759 | 0.901 |
| Raven, gender | 0.540 | 0.498 | **0.026** | 0.318 | 0.908 | 0.617 |
| **Raven, age, gender** | 0.679 | 0.464 | **0.018** | 0.995 | 0.749 | 0.896 |

Note: Emboldened values indicate rejection of the null hypothesis at the 5% significance level.

## Appendix D: Propensity Score Matching with Hiroshima CU

Some tables in this Appendix list the p-values for the z-test, where the null hypothesis is that subjects behavior at a site is homogenized with the behavior at Hiroshima CU. Differences in Raven score are also noted for reference in other tables.

### Kansai U-B vs. Hiroshima CU

Table 95: p-values for z-test: Kansai U-B vs. Hiroshima CU, No-fb

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| age | 0.168 | 0.733 | 0.870 | 0.301 | 0.290 | 0.963 |
| gender | 0.203 | 0.847 | 0.684 | 0.233 | 0.604 | 0.739 |
| age, gender | 0.230 | 0.786 | 0.923 | 0.821 | 0.452 | 0.444 |
| age, sci-eng | 0.141 | 0.845 | 0.969 | 0.375 | 0.392 | 0.815 |
| gender, sci-eng | 0.388 | 0.984 | 0.939 | 0.513 | 0.692 | 0.365 |
| **age, gender, sci-eng** | 0.453 | 0.992 | 0.920 | 0.729 | 0.449 | 0.179 |
| Raven | 0.713 | 0.208 | 0.598 | 0.648 | 0.925 | 0.857 |
| Raven, age | 0.476 | 0.489 | 0.684 | 0.439 | 0.440 | 0.500 |
| Raven, gender | 0.612 | 0.334 | 0.647 | 0.598 | 0.979 | 0.949 |
| Raven, age, gender | 0.857 | 0.356 | 0.549 | 0.601 | 0.725 | 0.660 |
| Raven, age, sci-eng | 0.673 | 0.230 | 0.433 | 0.630 | 0.689 | 0.621 |
| Raven, gender, sci-eng | **0.025** | 0.841 | 0.404 | **0.044** | 0.908 | 0.492 |
| **Raven, age, gender, sci-eng** | 0.074 | 0.232 | 0.333 | **0.031** | 0.896 | 0.233 |

Table 96: p-values for z-test: Kansai U-B vs. Hiroshima CU, Part-fb

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| age | 0.433 | 0.288 | 0.095 | 0.085 | **0.006** | 0.243 |
| gender | 0.474 | 0.065 | **0.022** | **0.025** | **0.002** | **0.018** |
| age, gender | 0.476 | 0.340 | 0.055 | **0.011** | **0.017** | 0.197 |
| age, sci-eng | 0.685 | 0.140 | 0.161 | 0.168 | **<0.001** | **0.042** |
| gender, sci-eng | 0.437 | 0.114 | **0.018** | **0.036** | **0.007** | **0.032** |
| **age, gender, sci-eng** | 0.566 | 0.304 | 0.197 | 0.079 | **<0.001** | **0.016** |
| Raven | 0.348 | 0.256 | 0.151 | 0.115 | **0.002** | 0.055 |
| Raven, age | 0.425 | 0.618 | 0.208 | 0.179 | **0.016** | 0.225 |
| Raven, gender | 0.355 | 0.093 | 0.054 | 0.104 | **0.003** | **0.035** |
| Raven, age, gender | 0.242 | 0.221 | **0.025** | **0.025** | **0.009** | 0.195 |
| Raven, age, sci-eng | 0.568 | 0.982 | 0.512 | 0.713 | **0.001** | **0.018** |
| Raven, gender, sci-eng | 0.527 | 0.189 | **0.043** | **0.021** | **0.002** | **0.001** |
| **Raven, age, gender, sci-eng** | 0.483 | 0.605 | 0.605 | 0.534 | **0.002** | 0.052 |

Table 97: p-values for z-test: Kansai U-B vs. Hiroshima CU, Full-fb

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| age | 0.341 | 0.788 | **0.032** | 0.983 | 0.438 | 0.452 |
| gender | 0.514 | 0.617 | 0.055 | 0.989 | 0.921 | 0.282 |
| age, gender | 0.537 | 0.687 | 0.081 | 0.890 | 0.393 | 0.762 |
| age, sci-eng | 0.618 | 0.351 | 0.059 | 0.581 | 0.474 | 0.567 |
| gender, sci-eng | 0.370 | 0.728 | 0.076 | 0.756 | 0.935 | 0.330 |
| **age, gender, sci-eng** | 0.481 | 0.905 | **0.041** | 0.752 | 0.474 | 0.625 |
| Raven | 0.272 | 0.865 | 0.117 | 0.957 | 0.917 | 0.305 |
| Raven, age | 0.800 | 0.927 | 0.057 | 0.783 | 0.483 | 0.698 |
| Raven, gender | 0.235 | 0.656 | 0.051 | 0.781 | 0.731 | 0.459 |
| Raven, age, gender | 0.144 | 0.719 | **0.022** | 0.700 | 0.219 | 0.552 |
| Raven, age, sci-eng | 0.614 | 0.786 | **0.011** | 0.641 | 0.789 | 0.585 |
| Raven, gender, sci-eng | 0.236 | 0.742 | **0.014** | 0.516 | 0.603 | 0.282 |
| **Raven, age, gender, sci-eng** | 0.208 | 0.754 | 0.060 | 0.481 | 0.529 | 0.098 |

Table 98: p-values for z-test: Kansai U-B (excl. econ) vs. Hiroshima CU, No-fb

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| age | 0.267 | 0.838 | 0.996 | 0.491 | 0.412 | 0.942 |
| gender | 0.302 | 0.851 | 0.843 | 0.422 | 0.644 | 0.801 |
| age, gender | 0.312 | 0.878 | 0.878 | 0.953 | 0.386 | 0.658 |
| age, sci-eng | 0.257 | 0.867 | 0.868 | 0.535 | 0.795 | 0.913 |
| gender, sci-eng | 0.517 | 0.994 | 0.960 | 0.692 | 0.729 | 0.504 |
| **age, gender, sci-eng** | 0.374 | 0.747 | 0.612 | 0.796 | 0.471 | 0.325 |
| Raven | 0.547 | 0.329 | 0.573 | 0.694 | 0.901 | 0.729 |
| Raven, age | 0.223 | 0.846 | 0.816 | 0.263 | 0.395 | 0.761 |
| Raven, gender | 0.652 | 0.724 | 0.863 | 0.553 | 0.855 | 0.871 |
| Raven, age, gender | 0.779 | 0.514 | 0.669 | 0.692 | 0.946 | 0.772 |
| Raven, age, sci-eng | 0.222 | 0.609 | 0.728 | 0.241 | 0.544 | 0.978 |
| Raven, gender, sci-eng | 0.179 | 0.622 | 0.567 | 0.174 | 0.585 | 0.796 |
| **Raven, age, gender, sci-eng** | 0.106 | 0.260 | 0.222 | **0.040** | 0.911 | 0.902 |

Table 99: p-values for z-test: Kansai U-B (excl. econ) vs. Hiroshima CU, Part-fb

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| age | 0.481 | 0.436 | 0.126 | 0.130 | **0.002** | 0.238 |
| gender | 0.492 | 0.136 | **0.027** | 0.064 | **0.002** | **0.015** |
| age, gender | 0.938 | 0.982 | 0.194 | 0.220 | **0.001** | **0.004** |
| age, sci-eng | 0.942 | 0.194 | 0.058 | 0.114 | **0.001** | 0.054 |
| gender, sci-eng | 0.491 | 0.309 | **0.039** | 0.108 | **0.007** | **0.038** |
| **age, gender, sci-eng** | 0.782 | 0.682 | 0.103 | 0.251 | **0.002** | 0.141 |
| Raven | 0.379 | 0.404 | 0.238 | 0.260 | **0.004** | 0.080 |
| Raven, age | 0.536 | 0.990 | 0.297 | 0.534 | **0.003** | 0.150 |
| Raven, gender | 0.421 | 0.146 | 0.121 | 0.120 | **0.004** | 0.057 |
| Raven, age, gender | 0.752 | 0.740 | 0.389 | 0.994 | **0.006** | 0.094 |
| Raven, age, sci-eng | 0.979 | 0.853 | 0.260 | 0.844 | **0.031** | 0.073 |
| Raven, gender, sci-eng | 0.179 | 0.296 | 0.176 | 0.080 | **0.012** | **0.038** |
| **Raven, age, gender, sci-eng** | 0.135 | 0.943 | 0.600 | 0.777 | **0.028** | 0.421 |

Table 100: p-values for z-test: Kansai U-B (excl. econ) vs. Hiroshima CU, Full-fb

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| age | 0.328 | 0.990 | 0.054 | 0.347 | 0.431 | 0.321 |
| gender | 0.444 | 0.911 | 0.051 | 0.584 | 0.938 | 0.164 |
| age, gender | 0.470 | 0.794 | 0.093 | 0.577 | 0.276 | 0.662 |
| age, sci-eng | 0.906 | 0.738 | **0.014** | 0.176 | 0.400 | 0.255 |
| gender, sci-eng | 0.300 | 0.988 | 0.079 | 0.416 | 0.810 | 0.220 |
| **age, gender, sci-eng** | 0.638 | 0.829 | **0.004** | 0.542 | 0.102 | 0.500 |
| Raven | 0.312 | 0.974 | 0.091 | 0.693 | 0.823 | 0.157 |
| Raven, age | 0.375 | 0.822 | **0.021** | 0.258 | 0.154 | 0.058 |
| Raven, gender | 0.104 | 0.924 | 0.068 | 0.582 | 0.935 | 0.212 |
| Raven, age, gender | 0.122 | 0.789 | 0.193 | 0.505 | 0.058 | 0.458 |
| Raven, age, sci-eng | 0.725 | 0.706 | **0.042** | 0.329 | 0.355 | 0.636 |
| Raven, gender, sci-eng | 0.238 | 0.899 | 0.116 | 0.738 | 0.628 | 0.091 |
| **Raven, age, gender, sci-eng** | 0.539 | 0.633 | 0.058 | 0.591 | 0.500 | 0.343 |

Table 101: p-values for z-test: Kansai U-B (sci-eng) vs. Hiroshima CU (sci-eng)

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| age | 0.526 | 0.521 | 0.659 | 0.540 | **0.010** | 0.459 |
| gender | 0.254 | 0.661 | 0.584 | 0.318 | 0.215 | 0.588 |
| age, gender | 0.588 | 0.482 | 0.510 | 0.585 | **0.006** | 0.275 |
| Raven | **0.028** | 0.566 | 0.178 | 0.997 | 0.601 | 0.698 |
| Raven, age | 0.791 | 0.138 | 0.186 | 0.642 | 0.247 | 0.670 |
| Raven, gender | **0.017** | 0.488 | 0.321 | 0.823 | 0.431 | 0.895 |
| Raven, age, gender | 0.497 | 0.412 | 0.553 | 0.486 | <**0.001** | 0.296 |

Note: The p-value for the Brunner-Munzel test indicates rejection of the null hypothesis at the 5% significance level. According to Criterion 1, we cannot say that subjects' behavior was homogenized, although, as far as science or engineering majors are concerned, there was no significant difference in Raven scores of subjects, on average, between Hiroshima CU and Kansai U-B.

Table 102: Difference in Raven score between Kansai U-B (sci-eng) and Hiroshima CU (sci-eng and others)

| site | # of subj. | mean | std.dev. | p-value | min | max |
|---|---|---|---|---|---|---|
| Hiroshima CU sci-eng | 24 | 12.583 | 1.792 | | 8 | 15 |
| Hiroshima CU others | 100 | 10.590 | 2.425 | <**0.001** | 3 | 15 |
| Kansai U-B sci-eng | 41 | 12.098 | 1.985 | 0.341 | 7 | 15 |

Note: The p-values for the Brunner-Munzel test are provided for the comparison with Hiroshima CU sci-eng. An emboldened value indicates rejection of the null hypothesis at the 5% significance level. As far as science or engineering majors are concerned, there is no significant difference in Raven scores of subjects between Hiroshima CU and Kansai U-B. Hiroshima CU has the Faculty of Arts, but only a few students of the faculty participated in the experiment.

Table 103: p-values for z-test: Kansai-B (excl. sci-eng, econ) vs. Hiroshima CU (excl. sci-eng), integrated.

| Covariates | $1-10$ | $11-20$ | $21-30$ | $31-40$ | $41-50$ | $51-60$ |
|---|---|---|---|---|---|---|
| age | 0.575 | 0.519 | 0.054 | **0.030** | 0.059 | 0.301 |
| gender | 0.280 | 0.482 | 0.127 | 0.091 | 0.071 | 0.226 |
| **age, gender** | 0.534 | 0.564 | 0.110 | 0.074 | **0.025** | 0.238 |
| Raven | 0.153 | 0.694 | 0.217 | 0.110 | 0.073 | 0.144 |
| Raven age | 0.186 | 0.342 | 0.373 | 0.227 | **0.016** | 0.106 |
| Raven gender | 0.056 | 0.355 | 0.105 | 0.068 | 0.231 | 0.157 |
| **Raven, age, gender** | 0.359 | 0.383 | 0.610 | 0.248 | **0.038** | 0.469 |

Table 104: p-values for z-test: Kansai-B (excl. sci-eng, econ) vs. Hiroshima CU (excl. sci-eng), No-fb.

| Covariates | $1-10$ | $11-20$ | $21-30$ | $31-40$ | $41-50$ | $51-60$ |
|---|---|---|---|---|---|---|
| age | 0.164 | 0.699 | 0.892 | 0.339 | 0.742 | 0.697 |
| gender | 0.440 | 0.861 | 0.957 | 0.625 | 0.822 | 0.275 |
| **age, gender** | 0.394 | 0.869 | 0.808 | 0.687 | 0.369 | 0.279 |
| Raven | 0.753 | 0.442 | 0.334 | 0.798 | 0.869 | 0.907 |
| Raven age | 0.061 | 0.546 | 0.968 | 0.162 | 0.794 | 0.581 |
| Raven gender | 0.324 | 0.728 | 0.903 | 0.397 | 0.547 | 0.376 |
| **Raven, age, gender** | 0.251 | 0.789 | 0.643 | 0.735 | 0.594 | **0.033** |

Table 105: p-values for z-test: Kansai-B (excl. sci-eng) vs. Hiroshima CU (excl. sci-eng), Part-fb.

| Covariates | $1-10$ | $11-20$ | $21-30$ | $31-40$ | $41-50$ | $51-60$ |
|---|---|---|---|---|---|---|
| age | 0.351 | 0.252 | **0.023** | 0.050 | **0.001** | 0.231 |
| gender | 0.402 | 0.265 | **0.014** | **0.041** | **0.003** | **0.036** |
| **age, gender** | 0.798 | 0.522 | 0.069 | 0.027 | 0.013 | 0.046 |
| Raven | 0.194 | 0.447 | 0.061 | 0.237 | $<$ **0.001** | 0.134 |
| Raven age | 0.502 | 0.825 | 0.120 | 0.266 | **0.002** | 0.336 |
| Raven gender | 0.159 | 0.240 | 0.081 | 0.167 | **0.003** | 0.187 |
| **Raven, age, gender** | 0.447 | 0.444 | **0.035** | 0.082 | **0.019** | 0.142 |

Table 106: p-values for z-test: Kansai-B (excl. sci-eng, econ) vs. Hiroshima CU (excl. sci-eng), Full-fb.

| Covariates | $1-10$ | $11-20$ | $21-30$ | $31-40$ | $41-50$ | $51-60$ |
|---|---|---|---|---|---|---|
| age | 0.513 | 0.386 | 0.204 | 0.309 | 0.931 | 0.233 |
| gender | 0.735 | 0.876 | 0.123 | 0.147 | 0.452 | 0.089 |
| **age, gender** | 0.956 | 0.493 | 0.066 | 0.221 | 0.629 | 0.128 |
| Raven | 0.938 | 0.879 | 0.277 | 0.520 | 0.638 | 0.190 |
| Raven age | 0.513 | 0.386 | 0.204 | 0.309 | 0.931 | 0.233 |
| Raven gender | 0.334 | 0.716 | 0.065 | 0.156 | 0.917 | 0.358 |
| **Raven, age, gender** | 0.154 | 0.296 | 0.098 | 0.436 | 0.733 | 0.317 |

**Hiroshima CU vs. Doshisha**

Table 107: p-values for z-test: Hiroshima CU vs. Doshisha, No-fb

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| age | 0.200 | 0.572 | 0.897 | 0.725 | 0.161 | 0.774 |
| gender | 0.053 | 0.924 | 0.598 | 0.415 | 0.532 | 0.570 |
| **age, gender** | 0.170 | 0.984 | 0.474 | 0.610 | 0.245 | 0.626 |
| Raven | 0.201 | 0.159 | 0.719 | 0.920 | 0.796 | 0.808 |
| Raven, age | 0.555 | 0.098 | 0.296 | 0.258 | 0.923 | 0.814 |
| Raven, gender | 0.097 | 0.658 | 0.760 | 0.478 | 0.366 | 0.733 |
| **Raven, age, gender** | 0.270 | 0.471 | 0.977 | 0.933 | 0.554 | 0.391 |

Table 108: p-values for z-test: Hiroshima CU vs. Doshisha, Part-fb

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| age | 0.909 | **0.005** | 0.166 | **0.037** | **0.030** | 0.895 |
| gender | 0.983 | <**0.001** | 0.626 | 0.224 | **0.006** | 0.236 |
| **age, gender** | 0.474 | 0.058 | 0.599 | 0.169 | **0.022** | 0.183 |
| Raven | 0.796 | **0.005** | 0.363 | 0.145 | **0.040** | 0.599 |
| Raven, age | 0.547 | **0.006** | 0.320 | 0.183 | **0.006** | 0.211 |
| Raven, gender | 0.988 | **0.003** | 0.639 | 0.288 | **0.017** | 0.290 |
| **Raven, age, gender** | 0.652 | **0.017** | 0.333 | 0.097 | **0.013** | 0.857 |

Table 109: p-values for z-test: Hiroshima CU vs. Doshisha, Full-fb

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| age | 0.497 | 0.193 | 0.683 | 0.525 | 0.136 | 0.297 |
| gender | 0.275 | 0.912 | 0.572 | 0.908 | 0.549 | 0.666 |
| **age, gender** | 0.475 | 0.142 | 0.902 | 0.281 | 0.199 | 0.464 |
| Raven | 0.775 | 0.549 | 0.806 | 0.717 | 0.563 | 0.370 |
| Raven, age | 0.800 | 0.487 | 0.525 | 0.896 | 0.386 | 0.361 |
| Raven, gender | 0.366 | 0.741 | 0.911 | 0.957 | 0.196 | 0.359 |
| **Raven, age, gender** | 0.522 | 0.697 | 0.192 | 0.907 | 0.737 | 0.457 |

Table 110: p-values for z-test: Hiroshima CU (excl. sci-eng) vs. Doshisha (excl. econ), No-fb

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| age | 0.147 | 0.458 | 0.870 | 0.634 | 0.095 | 0.952 |
| gender | 0.142 | 0.919 | 0.694 | 0.554 | 0.659 | 0.161 |
| **age, gender** | 0.573 | 0.626 | 0.620 | 0.981 | 0.100 | 0.369 |
| Raven | 0.234 | 0.182 | 0.407 | 0.705 | 0.460 | 0.874 |
| Raven, age | 0.515 | 0.539 | 0.868 | 0.819 | 0.898 | 0.399 |
| Raven, gender | 0.150 | 0.773 | 0.368 | 0.373 | 0.769 | 0.371 |
| **Raven, age, gender** | 0.116 | 0.645 | 0.363 | 0.323 | 0.944 | 0.242 |

Table 111: p-values for z-test: Hiroshima CU (excl. sci-eng) vs. Doshisha (excl. econ), Part-fb

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| age | 0.835 | **0.001** | 0.150 | **0.024** | 0.069 | 0.615 |
| gender | 0.738 | <**0.001** | 0.556 | 0.176 | **0.016** | 0.141 |
| **age, gender** | 0.416 | **0.003** | 0.470 | 0.101 | 0.195 | 0.483 |
| Raven | 0.668 | **0.001** | 0.220 | 0.113 | **0.030** | 0.285 |
| Raven, age | 0.638 | <**0.001** | 0.190 | **0.007** | 0.205 | 0.647 |
| Raven, gender | 0.840 | **0.001** | 0.532 | 0.342 | 0.057 | 0.363 |
| **Raven, age, gender** | 0.415 | **0.009** | 0.547 | 0.231 | 0.152 | 0.560 |

Table 112: p-values for z-test: Hiroshima CU (excl. sci-eng) vs. Doshisha (excl. econ), Full-fb

| Covariates | 1 − 10 | 11 − 20 | 21 − 30 | 31 − 40 | 41 − 50 | 51 − 60 |
|---|---|---|---|---|---|---|
| age | 0.775 | 0.385 | 0.941 | 0.954 | 0.975 | 0.706 |
| gender | 0.767 | 0.553 | 0.678 | 0.955 | 0.683 | 0.781 |
| **age, gender** | 0.971 | 0.240 | 0.797 | 0.380 | 0.999 | 0.854 |
| Raven | 0.928 | 0.620 | 0.519 | 0.950 | 0.532 | 0.847 |
| Raven, age | 0.263 | 0.954 | 0.258 | 0.262 | 0.229 | 0.576 |
| Raven, gender | 0.826 | 0.606 | 0.669 | 0.912 | 0.917 | 0.509 |
| **Raven, age, gender** | 0.138 | 0.518 | 0.263 | 0.551 | 0.680 | 0.677 |