

Pre-Analysis Plan: Information frictions in job search and occupational segregation

Jung Hyuk Lee*

October 21, 2021

AEA RCT Registry: AEARCTR-0008243

AEA RCT Registry Title: Information frictions in job search and occupational segregation

Abstract

This document outlines a randomized controlled trial (RCT) for more than 1,000 Korean senior undergraduate students, aiming to evaluate the role of information frictions in fresh college graduates' occupational segregation. For the framed survey experiment on the occupational choices, seven sectors (three male-dominated, three female-dominated, one neutral) and four core job characteristics (wage, work hours, welfare institutions, job security) are chosen, based on the analysis of a nationally-representative survey of college graduates. At baseline, information on those sectors and characteristics are directly provided to the Treatment group. The Control group is not provided with any information. Students in both groups are followed for two more rounds of surveys throughout six months and examined on their labor market beliefs, preferences, job search and applications behaviors.

*University of Southern California, Ph.D. student; email: leejungh@usc.edu.

1 Introduction

Occupational segregation has been a key source of a gender wage gap in the labor market. To explain the persistence of the segregation, researchers have studied important factors from both labor demand and supply side, implicitly assuming complete information. However, can the job seekers' occupational choice be considered a solution of the optimization problem under complete information? The literature on college major choice reported on substantial gaps between students' beliefs and actual job characteristics. For example, [Conlon \(2019\)](#)'s survey of freshmen at the Ohio State University showed that the mean absolute errors about the average salary of graduates range from 22% to 42% of the true values. [Wiswall and Zafar \(2015\)](#)'s examination on New York University students and [Betts \(1996\)](#)'s study on University of California San Diego students showed no less bias. This substantial bias in college students' beliefs even in wages, about which it is not difficult to obtain information with a relatively little search effort on the internet, brings about the question on the contributions of the "complete information" component and the "incomplete information" component in fresh college graduates' occupational segregation. In other words, if college job seekers suffer from severe information frictions, the impact of gender gaps in skills and preferences on the segregation could be smaller than what would have occurred under complete information. Alternatively, if these information frictions could also vary by gender, they would result in segregated job search and excessive occupational segregation.

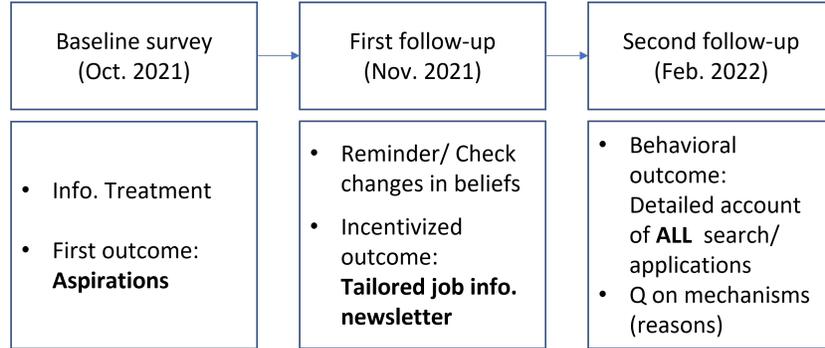
This project designs a survey experiment with more than one thousand Korean senior college students majoring in humanity and social sciences, which is designed (i) to detect the gaps in students' beliefs and the true parameters of both male-dominated sectors (MDS) and female-dominated sectors (FDS), (ii) to evaluate the role of biased beliefs in occupational segregation, and (iii) to assess the impact of provision of information on mitigating the segregation. In addition, we try to zoom in on college students' information gathering behaviors to suggest the heterogeneity of search costs as a main mechanism behind the information frictions. This pre-analysis plan focuses on outlining the experimental design and the main analyses.

2 Experimental Design

2.1 Overview

This study consists of three rounds of surveys – the baseline, the first follow-up, and the second follow-up survey. Figure 1 shows the planned timeline and the key features of three surveys. The baseline survey is administered in the beginning of the senior students' last semester (Fall 2021 semester), followed by the first follow-up in the middle of the semester, and then by the second follow-up in Spring 2022.

Figure 1: Timeline of the project



The sample is divided into three groups: Group 1 is male students *without* information treatment, Group 2 is female students *without* information treatment, and Group 3 is female students *with* information treatment.

At the baseline, Group 1 and 2 are compared to identify the existence of information frictions and their contributions to occupational segregation in college graduates’ job market. The survey includes rich questions about diverse job search methods and the search process, enabling the in-depth analysis that links heterogeneous search costs to gender gaps in the information set.

Group 2 and 3 (both females) are compared for the experimental evaluation of the impact of provision of information on the segregation. Group 3 (Treatment group) is to be provided with the accurate information on four key characteristics – wages, work hours, welfare level, and job seekers – of representative job sectors (which include both male-dominated sectors and female-dominated sectors) that are calculated from a nationally-representative survey of the first jobs of college graduates.

Three sets of outcome variables are examined that correspond to each round of surveys. At the baseline survey, female students’ stated aspirations to each experimental sector are asked. At the first follow-up survey, they are asked to participate in the demand survey of the “tailored job information newsletter” and choose the firms and characteristics that they want to receive information about. At the second follow-up survey administered after approximately 6 months after the baseline, they report the detailed history of their actual job search/applications behaviors.

2.2 Treatment

The key reason why there is sparse literature adopting experimental methods to occupational choice is that there are a vast diversity of jobs with innumerable characteristics (Kofoed et al. (2019)). To overcome this multi-dimensionality, this study implements a framed field experiment with carefully chosen job sectors and characteristics.

Based on the analysis on Korea’s Graduate Occupational Mobility Survey (GOMS), we choose seven key sectors that account for about 30% of the total first jobs of college graduates with humanity and social sciences majors. Three sectors are male-dominated: wholesale trade, financial service activities, manufacture of electric components. Three sectors are female-dominated: education, business support activities, social work activities. One sector is neutral in gender composition: retail trade.

As a next step, also based on the GOMS, we choose four core characteristics that Korean college students responded to be their most important considerations when searching for jobs – wage, work hours, welfare institutions, and job security. To construct objective measures of those characteristics, we define them as in Table 1 and explain these to students throughout the survey process.

Table 1: Construction of sector parameters

| | Population parameters (GOMS) |
|------------------------------|--|
| Wage | The (weighted) average of annual salary (including bonuses, if any) over sectors, deflated by Consumer Price Index (2020=100) |
| Work hours | The (weighted) average of reported weekly work hours (regular and overtime) over sectors |
| Welfare institutions measure | (1) Count the number of applicable welfare institutions among the following ten: - severance pay, payed leave, overtime pay, bonus, weekly allowance, pension, medical insurance, employment insurance, industrial accident insurance (2) Take a (weighted) average within sectors and rescale to 0-5. |
| Job security measure | (1) Count the applicable job security status between the two: - whether the worker is a regular employee - whether the worker’s employment contract is not based on a fixed-term (2) Take a (weighted) average within sectors and rescale to 0-5. |

Notes: The table shows how each population parameter is constructed. Most of the 48,227 respondents of the GOMS survey in the past five years responded non-missing values for all questions used for the variable construction.

Without further information except for the examples of jobs belonging to each sector, students in the Control group (Group 2) are asked to specify (i) their subjective probabilities of getting a job offer from each sector before or right after graduation if they start making efforts from now, (ii) their personal weights across four characteristics as job choice criteria, (iii) their beliefs on the average of the four population characteristics of seven pre-specified sectors, and (iv) their aspirations to each sector.

What is different for the Treatment group (Group 3) among the survey sequence is (iii). For

Group 3, instead of asking about students’ beliefs, I provide a table displaying the population averages of the wages, work hours, welfare, and job security for top 20 sectors that are calculated from the responses by female graduates in the GOMS, with a short instruction explaining how to read the numbers correctly. All numbers are provided both in the absolute term and the relative term to the financial services sector for easy comparison. Also, the best and worst three sectors for each of four characteristics are colored differently for instant legibility. Appendix A shows the treatment material that is accompanied by related questions. To secure the salience of the intervention, I adopt various strategies. To make sure that the participants actually read the numbers, I borrow Conlon (2019)’s strategy of letting them write down the numbers about the seven pre-specified sectors that they see in the blank. After that, I ask them to evaluate whether those numbers are larger or smaller than they expected. Then the questionnaire moves on to (iv) and ask their aspirations to seven sectors. Notably, the questions are carefully ordered to affect only the beliefs on the four characteristics, avoiding their influence on the subjective probabilities.

2.3 Outcome variables

As outlined in the overview, there are three sets of outcome variables that correspond to each round of surveys.

Figure 2: Outcome variables

| Baseline survey | First follow-up | Second follow-up |
|--|--|---|
| <ul style="list-style-type: none"> • Aspirations | <ul style="list-style-type: none"> • Change in Beliefs • Aspirations | <ul style="list-style-type: none"> • Change in Beliefs • Aspirations |
| | <ul style="list-style-type: none"> • Sectors/firms searched • Firm choices for the newsletter | <ul style="list-style-type: none"> • Sectors/firms searched • Sectors/firms applied for |

Notes: Primary outcome variables for each round of surveys are in bold text.

At the baseline survey, students are asked to respond their aspirations to each of seven experimental sectors in 1-10 Likert scale. The question used is: “Based on the information that you currently have, please report your aspirations to each sector as your first job after graduation in a Likert scale, with 1 (least likely) to 10 (most likely).”

At the first follow-up survey, changes in students’ beliefs and aspirations are measured. In addition, they are provided with an incentivized choice problem. We ask them to participate in constructing a pilot version of a “tailored job information newsletter.” The newsletter is explained to be what provides information on firms and characteristics about which students

self-select to receive information about. For example, if a student chooses “ABC electronics” as a firm name and “wage and welfare institutions” as firm characteristics, the newsletter provides that information. We list the names of about 50 real firms in the seven sectors and their current job openings and make students choose as many firms as they want (maximum 10 choices). The outcome variable is the total number and sector composition of firms chosen. We also ask students’ willingness to pay for the newsletter, if the service is to launch formally.

At the second follow-up survey, students report all the jobs that they have searched/applied for in the previous six months in detail, regardless of the seven sectors. The primary outcome variable is also the list of firms searched/applied for and their sector composition.

3 Analysis

3.1 Baseline decomposition (Group 1 vs. Group 2)

The first part of the analysis is focusing on detecting the information frictions and estimate their role in gender segregation reported as aspirations to seven sectors cross-sectionally in the baseline survey. We compare the beliefs of the both gender students on the four characteristics of seven sectors with the population parameters calculated in the GOMS. Based on the reported beliefs, we decompose the aspirational gaps into gender gaps in four components and assess their contributions: i) subjective probability of getting a job offer, ii) preferences over job characteristics, iii) (actual) gaps in population parameters of MDS and FDS, and iv) gender-differential gaps in beliefs and population parameters.

3.2 Experimental evaluation (Group 2 vs. Group 3)

3.2.1 Key specification

The second part of the analysis is to estimate the impact of information treatment (direct provision of information) on the three sets of outcome variables. Since the randomization provides conditional independence, the form of the main specification is simple:

$$Y_i = \alpha_0 + \alpha_1 Treat_i + X_i' \beta + \epsilon_i$$

where Y_i is outcome variables for respondent i , $Treat_i$ is an indicator equal to 1 if respondent i belongs to the Treatment group (Group 3), and X_i' is a vector of control variables. The survey includes a rich set of potential control variables, including demographics (age, location, etc.), educational attainment/human capital accumulation (quality of school, specific major, GPA, certificates, etc.), and background (parents’ income and education). Since the randomization is implemented at the individual level, we will report both estimated results (i) without controls

and (ii) with controls systematically selected, for example, by double LASSO procedure. In addition, for the analysis on follow-up survey results, we will also control for baseline characteristics, such as job search/applications history.

Since the data points for the subjective probability, beliefs on characteristics, and most outcome variables are individual-sector level (e.g. student i 's aspiration to sector s), aggregated variables in individual level will be used for the estimation. Student i 's *relative* aspirations to male-dominated sectors can be defined as:

$$Y_i = \frac{\sum Y_{im}}{\sum Y_{if}}$$

where Y_{im} is i 's aspiration to each of three male-dominated sectors and Y_{if} is her aspiration to each of three female-dominated sectors. Other outcome variables are also standardized to be relative terms in the similar fashion.

3.2.2 Analysis on heterogeneity

Since we posit that the treatment effect comes from the correction of beliefs by provision of accurate information, we evaluate the differential impact of the treatment by the initial information gap, measured as differences between students' beliefs and the population parameters. The identifying assumption is that female students with larger information gap would show more substantial changes in the outcome variables.

3.3 Additional analysis on mechanism

If provision of information demonstrates statistically significant treatment effects, we plan to go the extra mile to shed light on the mechanism behind the information frictions, even in the environment where job information is searched with relatively low costs. Our hypothesis is that there still remains substantial heterogeneity in search costs across search methods and information contents that cause gender gap in information. To achieve the goal, the surveys include rich questions about students' job search methods and the quality/costs of information they obtain from various sources. We try to construct gender-specific matrices of job search costs of which rows are search methods and columns are information contents, respectively and develop a comprehensive framework that incorporates the heterogeneity in search costs, gender gaps in information, and occupational segregation.

References

- Betts, Julian R**, “What do students know about wages? Evidence from a survey of undergraduates,” *Journal of human resources*, 1996, pp. 27–56.
- Conlon, John J**, “Major malfunction: A field experiment correcting undergraduates’ beliefs about salaries,” *Journal of Human Resources*, 2019, pp. 0317–8599R2.
- Kofoed, Michael S et al.**, “The effect of same-gender or same-race role models on occupation choice evidence from randomly assigned mentors at west point,” *Journal of Human Resources*, 2019, 54 (2), 430–467.
- Wiswall, Matthew and Basit Zafar**, “Determinants of college major choice: Identification using an information experiment,” *The Review of Economic Studies*, 2015, 82 (2), 791–824.

A Treatment material

Figure A1: The first page

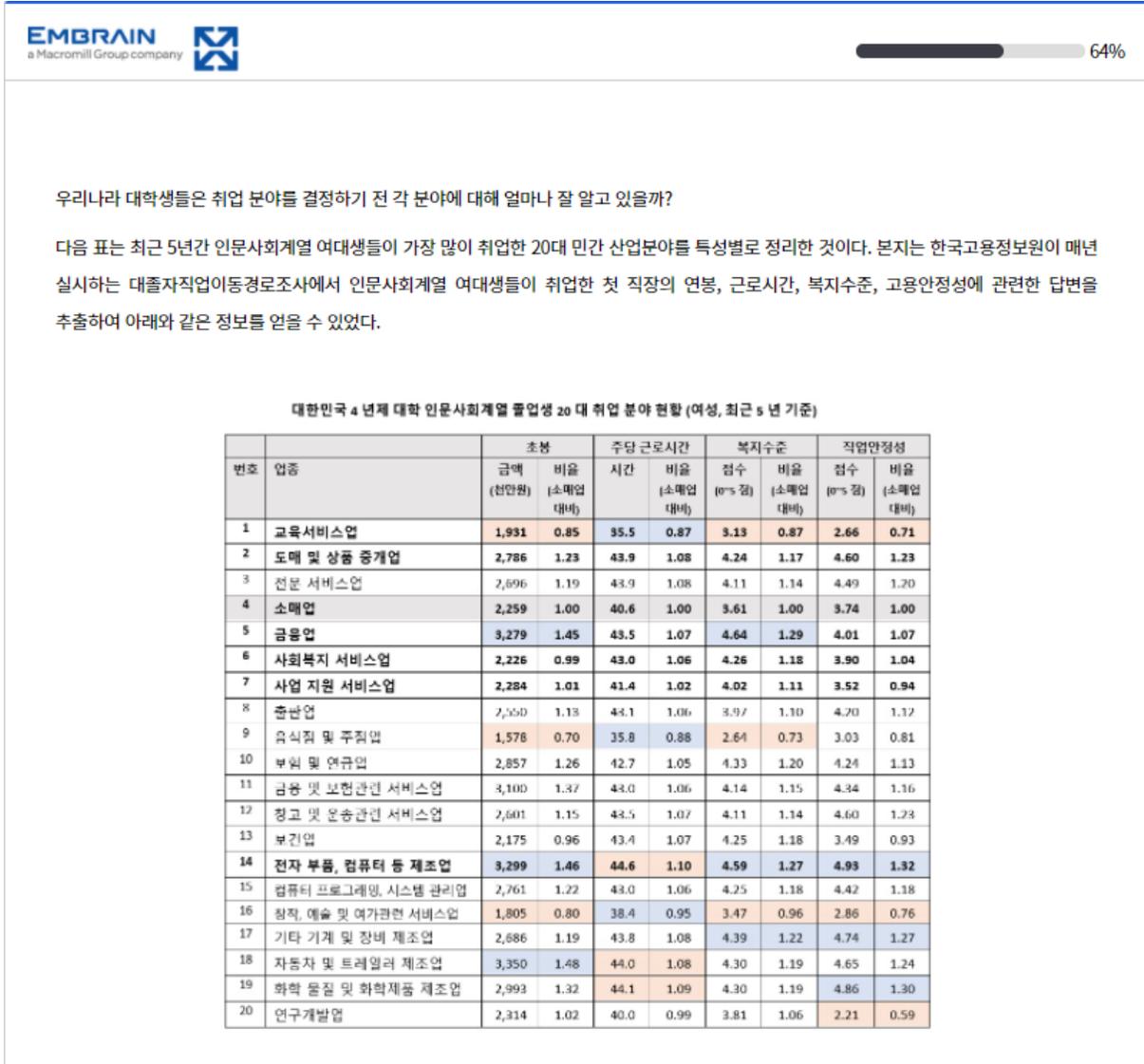


Figure A2: The second page

각 항목별로 “비율” 열의 숫자는 본지가 기준 분야로 삼은 “소매업” 대비 비율을 나타낸다. (예를 들어, 교육서비스업의 초봉 비율란에 쓰여 있는 0.85는 교육 서비스업의 평균 연봉이 소매업의 85% 수준임을 의미함.) 복지수준과 직업안정성 지표는 대졸자직업이동경로조사에 활용된 각종 지표들을 종합하여 산출되었다. 먼저 복지수준 지표는 각 업종별로 9가지 복지제도 (법정퇴직금, 유급휴가, 시간 외 수당, 상여금, 유급주휴, 국민연금/특수지역연금, 건강보험, 고용보험, 산재보험)의 해당 갯수의 평균값을 계산한 후 0~5점의 척도로 변환한 것으로, 점수가 높을수록 복지수준이 높음을 의미한다. 다음으로 직업안정성 지표는 각 업종별로 상용근로자 및 기간제 근로자의 비율을 계산한 후 역시 0~5점의 척도로 변환한 것으로, 역시 점수가 높을수록 직업 안정성이 높음을 의미한다.

표 안의 파란색 표시는 해당 20개 업종 중 가장 높은 순위를 보인 세 업종을 의미하고, 붉은색 표시는 가장 낮은 순위를 보인 세 업종을 의미한다. 예를 들어, 평균 초봉이 가장 높았던 업종은 순서대로 자동차 및 트레일러 제조업, 전자 부품, 컴퓨터 등 제조업, 금융업이었고, 평균 복지 수준이 가장 낮았던 업종은 순서대로 음식점 및 주점업, 교육 서비스업, 창작, 예술 및 여가관련 서비스업이었다.

참고로, 위의 업종 분류는 각 기업이 속한 산업에 의해 결정되며, 기업 내의 직무에 의해 결정되는 것은 아니다. 예를 들어, 삼성전자는 위 표에서 전자 부품, 컴퓨터 등 제조업에 포함되며, 삼성전자에 속한 생산직, 사무직, 비서직 등 모든 직무가 이에 포함된다.

< 후 략 >

대한민국 4 년제 대학 인문사회계열 졸업생 20 대 취업 분야 현황 (여성, 최근 5 년 기준)

| 번호 | 업종 | 초봉 | | 주당 근로시간 | | 복지수준 | | 직업안정성 | |
|----|--------------------|-------------|-------------------|---------|-------------------|---------------|-------------------|---------------|-------------------|
| | | 금액 (천만원) | 비율 (소매업 대비) | 시간 | 비율 (소매업 대비) | 점수 (0~5 점) | 비율 (소매업 대비) | 점수 (0~5 점) | 비율 (소매업 대비) |
| 1 | 교육서비스업 | 1,931 | 0.85 | 35.5 | 0.87 | 3.13 | 0.87 | 2.66 | 0.71 |
| 2 | 도매 및 상품 중개업 | 2,786 | 1.23 | 43.9 | 1.08 | 4.24 | 1.17 | 4.60 | 1.23 |
| 3 | 전문 서비스업 | 2,696 | 1.19 | 43.9 | 1.08 | 4.11 | 1.14 | 4.49 | 1.20 |
| 4 | 소매업 | 2,259 | 1.00 | 40.6 | 1.00 | 3.61 | 1.00 | 3.74 | 1.00 |
| 5 | 금융업 | 3,279 | 1.45 | 43.5 | 1.07 | 4.64 | 1.29 | 4.01 | 1.07 |
| 6 | 사회복지 서비스업 | 2,226 | 0.99 | 43.0 | 1.06 | 4.26 | 1.18 | 3.90 | 1.04 |
| 7 | 사업 지원 서비스업 | 2,284 | 1.01 | 41.4 | 1.02 | 4.02 | 1.11 | 3.52 | 0.94 |
| 8 | 출판업 | 2,550 | 1.13 | 44.1 | 1.06 | 3.97 | 1.10 | 4.70 | 1.17 |
| 9 | 음식점 및 주점업 | 1,578 | 0.70 | 35.8 | 0.88 | 2.64 | 0.73 | 3.03 | 0.81 |
| 10 | 보원 및 연금업 | 2,857 | 1.26 | 42.7 | 1.05 | 4.33 | 1.20 | 4.24 | 1.13 |
| 11 | 금융 및 보험관련 서비스업 | 3,100 | 1.37 | 43.0 | 1.06 | 4.14 | 1.15 | 4.34 | 1.16 |
| 12 | 항공 및 운송관련 서비스업 | 2,601 | 1.15 | 43.5 | 1.07 | 4.11 | 1.14 | 4.60 | 1.23 |
| 13 | 보건업 | 2,175 | 0.96 | 43.4 | 1.07 | 4.25 | 1.18 | 3.49 | 0.93 |
| 14 | 전자 부품, 컴퓨터 등 제조업 | 3,299 | 1.46 | 44.6 | 1.10 | 4.59 | 1.27 | 4.93 | 1.32 |
| 15 | 컴퓨터 프로그래밍, 시스템 관리업 | 2,761 | 1.22 | 43.0 | 1.06 | 4.25 | 1.18 | 4.42 | 1.18 |
| 16 | 창작, 예술 및 여가관련 서비스업 | 1,805 | 0.80 | 38.4 | 0.95 | 3.47 | 0.96 | 2.86 | 0.76 |
| 17 | 기타 기계 및 장비 제조업 | 2,686 | 1.19 | 43.8 | 1.08 | 4.39 | 1.22 | 4.74 | 1.27 |
| 18 | 자동차 및 트레일러 제조업 | 3,350 | 1.48 | 44.0 | 1.08 | 4.30 | 1.19 | 4.65 | 1.24 |
| 19 | 화학 물질 및 화학제품 제조업 | 2,993 | 1.32 | 44.1 | 1.09 | 4.30 | 1.19 | 4.86 | 1.30 |
| 20 | 연구개발업 | 2,314 | 1.02 | 40.0 | 0.99 | 3.81 | 1.06 | 2.21 | 0.59 |