Hitotsubashi Journal of Economics 65 (2024), pp.177-198.
https://doi.org/10.15057/hje.2024009

ORDER BIAS IN SEQUENTIAL PRESENTATION? EVIDENCE FROM UNIVERSITY ENTRANCE INTERVIEWS*

Beomsoo Kim

Department of Economics, Korea University,
Seoul, 02841, South Korea
kimecon@korea.ac.kr

and

Sang Soo Park

Department of Economics, Korea University,
Seoul, 02841, South Korea
starpac@korea.ac.kr

Received April 2024; Accepted October 2024

Abstract

This study examines the influence of presentation order on evaluations using a unique dataset from university entrance interviews conducted by one of the major universities in South Korea. Our data has three main advantages over the previous literature: random assignment of interview order; better controlled situations for all interviewees; and large numbers of interviewers and interviewees. Most of the order effect parameters are statistically insignificant. More importantly, even the statistically significant ones are practically insignificant in influencing decision-making outcomes. We also found no practically significant path dependency.

Keywords: presentation order, random assignment, interview
JEL Classification Codes: D1, D7, H8

I. *Introduction: Order Effects in Existing Literature*

Choice of an alternative from multiple options that are sequentially presented is pervasive and holds tremendous importance in real life. Consider, for example, that you are a government

* Sang Soo Park acknowledges supports by the Institute of Information & Communications Technology Planning & Evaluation (IITP) Grant funded by the Korean Government (MSIT) (No.2022-0-00302, Development of global demand forecasting and analysis/prediction system of market/industry trends). Beomsoo Kim acknowledges supports by the Korea University Grant (K2312841)

officer responsible for allocating taxpayers' money to various venture capital companies. Your task is to evaluate their research project proposals and determine which projects should receive funding based on factors like success probability, profitability, and potential contribution to society. You must evaluate these proposals sequentially, and ideally, your decision should be solely based on each project's inherent features. However, if your decision is influenced by the order in which you examine the proposals, it may lead to inefficiencies in money allocation.

A significant body of literature has documented that the order in which alternatives are presented or reviewed can impact final decisions, known as the 'order effect' or 'order bias.' This phenomenon is observed in various contexts, including political voting (Miller and Krosnick, 1998), psychology (Mussweiler, 2003), economics (Neilson, 1998; Page and Page, 2010; Sarafidis, 2007), marketing (Novemsky and Dhar, 2005), music contests (Glejser and Heyndels on the Queen Elisabeth Music Contest, 2001), the legal system (Danziger et al. on judges' parole decisions, 2011; Bindler and Hjalmarsson, 2019), and even the healthcare system among physicians (Jin et al., 2024).

However, findings from the mentioned studies have faced challenges. Firstly, it is crucial to ensure that the order is determined randomly, independent of any observable or unobservable characteristics of the items. For instance, Danziger et al. (2011) found an order effect in judges' parole decisions in Israel, but subsequent research suggested that the ordering might not be random. Factors such as prisoners' decisions regarding legal representation and judges' intentional placement of cases may contribute to observed patterns (Weinshall-Margel and Shapard, 2011; Glockner, 2016). Glockner (2016) demonstrated that the perceived influence of order could be alternatively explained as a statistical artifact resulting from intentional case ordering for judges' time management.

Secondly, unobservable factors beyond presentation order can influence evaluators' behaviors. For example, the order of NBER working papers on webpages can impact downloads, possibly due to 'skimming' behavior. This is due to individuals focusing on the first few papers in the list due to time constraints. Miller and Krosnick's (1998) finding can also be explained by sort of 'skimming' behavior of voters.

Glejser and Heyndels (2001) found in their study on the Queen Elisabeth Contest, a six-day classical music competition, that performers on the closing night tended to be ranked approximately one place higher than those on the opening night. Despite random contestant orders and judges listening to all performances, the "curse of pioneering" effect might explain this trend. Later performers benefit from improved orchestra performance and increased familiarity of initially unpublished pieces as the contest continues. Achieving identical circumstances in each performance is nearly impossible.

Thirdly, the limited number of judges in many studies can undermine conclusions about 'order bias.' A single evaluator's decision may significantly impact results, as seen in studies with a small number of judges. For example, Danziger et al. (2011) had 8 judges and Glejser and Heyndels (2001) had 15 juries. Bruine de Bruin's (2006) study on figure skating had 7-14 judges for each competition.

We believe our data are largely free from these issues, as explained in detail in Section 2. They comprise interview data from a prominent university in Seoul, South Korea. The interviews were designed and conducted in a very well-controlled environment so that unobservable factors that can affects interviewers' evaluations are controlled for and the assignment of order is guaranteed to be random based on the university's regulation. The number of interviewers is large, too. Contrary to existing literature that supports existence of order effects, analysis of our data reveals that estimated order effect parameters are statistical insignificant in general and even statistically significant ones do not have practical importance. We can consider a possibility that what previous studies have identified as order bias might instead stem from statistical artifacts arising from the use of less pristine data.

The paper proceeds as follows. Section 2 discusses the university admission process in South Korea. Section 3 describes our data and econometric models. Section 4 presents the results and discuss. In section 5, we address another concern, so-called "path-dependency", related to the order effect. Section 6 concludes.

II. *University Admissions in South Korea*

University entrance in South Korea is facilitated through two main systems: the College Scholastic Ability Test (CSAT) system and the University Admission Officer (UAO) system. The CSAT is a nationwide examination conducted by the Korean government every year on the third Thursday of November every year. Every applicant for a college or university may take the test. Individual universities can admit applicant solely based on their CSAT scores but often such admissions are a limited fraction of total admissions.

The majority of admissions are granted through the other system. The UAO system bears similarities to the college admissions process in the United States. In the UAO system, students have the opportunity to apply to up to six different colleges or universities. A group of evaluation panels, comprised of professionals, thoroughly review the applicants' documents and selects candidates for interviews. Final admission will be given based on interviews, document reviews and, if applicable, requirements for CAST. Although this general procedure applies to all universities in South Korea, every university has its own policies and processes within the frame of CSAT and UAO systems. The details mentioned in Section 3 are specific to the university from which we obtained our data.

III. *Data and Econometric Models*

1. Interview Data

The data being utilized is confidential and pertains to individual-level interview data conducted in the year 2018 at a prominent university in Seoul, South Korea. This particular university offers several tracks within the University Admission Officer (UAO) system, with a notable focus on two tracks: the High School Principal Recommendation (HSPR) track and the High School Record (HSR) track. We will use these two tracks' interview data.[1] The university allocated sixty-seven percent of UAO system admissions to these two tracks.

Before the admission process commenced, the university allocates the number of admission that each department could offer for each track. For instance, the Department of

[1] The other tracks not included in this study have a small number of interviewees due to the limited allocation of the number admissions.

Economics was assigned 80 admissions for these two tracks out of a total of 120 available seats.

< HSPR track >

Under the HSPR track, high school principals recommend students for application, with a limit of up to 4% of the senior students in each school. The university's admission office evaluates the recommended students based on their high school records, statements of purpose, and recommendation letters. They then select interview candidates up to five times the number of admissions allotted for the track. Interviews take place over a single weekend, with morning and afternoon slots on both Saturday and Sunday. The interview slots are allocated based on departments, meaning that students applying for the same department are interviewed during the same slot. Multiple interviews are conducted simultaneously within each slot.

The interviewers consist of professors and professional admission officers who receive annual training or re-training. Most interviewers participate in both morning and afternoon slots. The track comprises two sequential sections: the "Interview with Article (IWA)" and the "High School Record Review (HSRR)". Each interviewee is interviewed by a group of two interviewers for IWA in one room and then by another group of two interviewers in a separate room for HSRR. In each section, two interviewers independently evaluate the interviewee's performance. They are not allowed to confer with each other.

In the IWA, where each interview lasts exactly for 6 minutes, interviewees are given with the same article in advance and interviewers ask questions related to the article. The articles are slot-specific. For example, interviewees for the morning slot are given one article, and those for the afternoon slot another.

In order to ensure fair and objective evaluation as much as possible, the university has imposed two additional restrictions. Firstly, interview questions are pre-set. Four predetermined interview questions are given to interviewees, and only these questions must be asked during the interview. Secondly, each interviewee will have exactly 24 minutes of preparation time after receiving the article and questions before their interview begins in a preparation room.

Discussions are held before the interview begins among interviewers. During the discussions, some suggested answers to the questions are provided and explained, which serve as a standard for interviewees' potential responses. Other answers that interviewees may answer are proposed and discussed among interviewers as well. The purpose of having discussion time is for interviewers to reach a common understanding on how to grade interviewees' answers.

Once the IWA section is completed, students proceed to the other room for the HSRR section for 6 minutes. Here, another group of two interviewers asks questions about the students' high school experiences, including curricular and extracurricular activities. Unlike the IWA section, the questions in the HSRR section are not pre-set, and each interviewee might face individualized questions that do not have specific suggested answers. The evaluation of the HSRR section can be regarded as subjective compared to the IWA section.

< HSR track >

Applicants for the HSR track should prepare similar documents to those for the HSRR track, except that a recommendation from their high school principal is not needed. After their

Table 1. Descriptive Statistics

	High School Principal Recommendation (HSPR)		High School Record (HSR)
	Interview with Article (IWA)	High School Record Review (HSRR)	
No. of interviewees (A)	4,081	4,080	4,022
No. of interview groups (B)	155	155	235
No. of interviewers (=2*B)	310	310	470
No. of records per feature	8,162	8,160	8,044
No. of interviewees per group	16~31	16~31	8~23

Notes: The number of records per feature is A * the number of interviewers per group(=2).
We use individual level interview data from a university in 2018.

documents are assessed, selected candidates are invited for an interview. Unlike the HSRR track, the interview for the HSR track is a single 10-minute section, which is a combination of the IWA and HSRR sections in the HSRR track. An article is provided with pre-set questions, similar to the IWA section. The interviewee is given 20 minutes to read the article and prepare their answers to the pre-set questions immediately before the interview in a preparation room. During the interview, a group of two interviewers asks not only the pre-set questions but also questions related to the interviewee's high school record.

Table 1 presents descriptive statistics for the interview data from the two different tracks. The IWA section of HSRR track had a total of 4,081 interviewees, while the HSRR section had 4,080 interviewees, as one student only took part in the IWA and not in the HSRR. There were 155 interview groups, resulting in 310 interviewers. The last column of Table 1 displays the data for the HSR track, with 4,022 interviewees divided into 235 interview groups. Since each interviewee was assessed by two interviewers, the total number of records for each feature is twice the number of interviewees. The IWA section contains 8,162 observations for each feature, except for two fewer observations in the HSRR section due to one student dropping out. The HSR track comprises 8,044 observations.

2. Randomization of Interview Order

One significant advantage of this data is that the order of the interviews is randomized. All interview candidates are required to arrive within specific time windows: 7:30 am to 8:20 am for morning interviews and 12:50 pm to 1:30 pm for afternoon interviews. Late arrival is not permitted to participate in the interview process. After the time window closes, the interviewees are assigned to interview groups randomly. Each interview group consists of candidates from the same department. In cases where a department has a large number of interviewees, multiple interview groups are formed within that department. The assignment of interview order within each interview group is also randomized, ensuring homogeneity in the distribution of observable and unobservable characteristics across the interview sequence.

Although the randomization of interview groups and interview order within a group is guaranteed by the university's institutional regulation, we aim to verify that the interview order is indeed random by conducting an econometric test. The concept is to assess whether observable characteristics are not correlated with interview order by regressing interviewees'

Table 2. Random Assignment Test

	High School Principal Recommendation (HSPR)		High School Record (HSR)
	Interview with Article (IWA)	High School Record Review (HSRR)	
No. of interviewees	4081	4080	4022
R-squared	0.008	0.008	0.007
Wald statistic	0.953	0.961	1.281
p-value	0.547	0.533	0.170

Notes: Dependent variable is birth month and independent variable is binary variable indicating the interview order of the interviewee. The null hypothesis is coefficient for binary variable indicating the interview order of the interviewee are jointly zero.

observable characteristics on their interview orders.

Unfortunately, the university did not provide individual information other than interviewees' birth months due to privacy issues. Therefore, we only have birth month as an observable characteristic. We regress birth month (with 1 representing January and 12 representing December, denoted as $Birth\ month_i$) on interview order ($I\{Order_i=k\}$)) as follows:

$$Birth\ month_i=\beta_1+\sum_{k=2}^{K}\beta_k * I\{Order_i=k\}+\varepsilon_i,\ \ E[\varepsilon_i\,|\,Order_i]=0.$$

Here, β_1 represents the intercept, β_k (where k ranges from 2 to K) represents the coefficients for each interview order, I{A} is an indicator function that takes the value 1 if A is true and 0 otherwise, and ε_i is the error term. The value of K is 31 for the IWA and HSRR data and 19 for the HSR data. The null hypothesis, H_0, is $\beta_2=\beta_3=\cdots=\beta_K=0$, implying no statistically significant birth month differences across the interview orders.

The Wald statistics for the H_0 for each track are reported in Table 2. Under the H_0, the Wald statistics asymptotically follow a chi-squared distribution with degrees of freedom equal to $K-1$, from which the p-values in the table are computed. As seen there, the obtained p-values for all three tracks were sufficiently large, indicating a failure to reject the null hypothesis. This implies that there were no statistically significant birth month differences observed across the interview orders.

3. Outcomes

In the IWA interview section, the two interviewers independently evaluate the interviewees based on four distinct features: Analytic Ability, Practical Ability, Critical Thinking, and Interview Attitude. Each feature, except for Interview Attitude that we excluded from our analysis, is graded on a scale of A+ to F. As F grades are rare, they were merged with the D grades.[2] We assigned a numeric value 5 to an A+, 4 to A, 3 to B, 2 to C, and 1 to D or F for our analysis.

For the HSRR interview section, there are two features that are evaluated: Major Preparedness and Motivation. In the HSR interview section, the interviewers independently

[2] Merging D and F does not change our results qualitatively.

evaluate the interviewees on four features: Analytical Ability, Practical Ability, Critical Thinking, and Motivation. The first three features are based on preset questions, while Motivation is related to high school record-related questions.[3] The same 5-point scale numeric transformation was applied to both HSRR and HSR sections.

4. Econometric Model

We adopted an ordered probit model as a natural choice for our econometric analysis due to the nature of our outcome variable, which is a 5-point Likert scale representing ordered categories. However, there are some issues with using a typical ordered probit model in our case. Let $Y_{i,j} \in \{1, 2, \cdots, 5\}$ be the evaluation score for the interviewee i that the interviewer j gave. Since there are two interviewers for the same interviewee, evaluation scores for the same interviewee i are given always in pairs, namely $(Y_{i,j1}, Y_{i,j2})$, which means we need to use a bivariate ordered probit model. A more difficult issue to handle than the necessity of a bivariate ordered probit model is that one interviewer conducted interviews for multiple interviewees, which necessitates controlling for interviewer-specific heterogeneity over the interviewees that the same interviewer interviewed. In addition, since we only have interview-related variables such as interviewee's interview order for the right-hand side variable, we include unobservable interviewee-specific heterogeneity in the outcome pair $(Y_{i,j1}, Y_{i,j2})$. Furthermore, putting aside all the technical issues in econometric modelling with an ordered probit model, the interpretation of the bivariate ordered probit model can be complex for our purpose. Therefore, we decided to use a linear regression approach for our main analysis, treating the outcome variable as a usual cardinal variable[4] . The bivariate ordered probit model results are comparable to the linear regression results.

For our main analysis, we considered the following model as our main model:

Model (1):

$$Y_{i,l} = \beta_0 + \beta_1 * Order_i + \beta_2 * X_i + u_i + v_l + \epsilon_{i,l},\ l = j_1 \text{ or } j_2, \text{ where } j_1 < j_2 \in \{1, \cdots, J\},$$

where X represents observable characteristics related to the interview, such as department that each interviewee is applying for and interview slots (Saturday morning and afternoon, Sunday morning and afternoon). The department dummy variables are used to control for potential differences in various majors. The inclusion of interview slot dummies (Saturday morning and afternoon, Sunday morning and afternoon) aims to account for disparities in pre-set interview questions.

The key independent variable, $Order_i$, represents the interview order and is expected to reflect any changes in the score values.

This model allows us to examine whether the scores change monotonically according to the interview order. The error terms need a special attention. It consists of three components:

1. The unobservable interviewee heterogeneity, u_i, X_i cannot capture;

[3] There is another feature that we do not include in the analysis: Interview Attitude in all three interview sections. It's grading scale is A, B, C, or F, which is different from that of features we analyzed, hence not directly comparable.

[4] Due to limited space we can only provide the details and the estimation results of the bivariate ordered probit model upon the request of the reader.

2. The unobservable interviewer heterogeneity v_l, where $l \in \{1, \cdots, J\}$, which affects the scores of all interviewees interviewed by the same interviewer l;
3. The remaining part ($\epsilon_{i,l}$), which we assume has a zero mean but is likely to exhibit heteroscedasticity since $Y_{i,l}$ is a discrete variable.

Let us denote the composite error term as $e_{i,l}$ i.e. $e_{i,l} = u_i + v_l + \epsilon_{i,l}$. We can reasonably assume that the three error components are independent of each other and have zero conditional means, thus satisfying $E[e_{i,l} | Order_i, X_i] = 0$. However, due to the shared u_i component between Y_{i,j_1} and Y_{i,j_2}, we have $E[e_{i,j_1} e_{i,j_2} | Order_i, X_i] \neq 0$. Additionally, since v_l is shared with other interviewees who were interviewed by the same interviewer l, we have $E[e_{i,l} e_{j,l} | Order_i, X_i, Order_j, X_j] \neq 0$. Moreover, as mentioned earlier, the $\epsilon_{i,l}$ term is heteroskedastic. Hence, the variance structure of the composite error term is complex.

We reported two different estimation results:

1. The least square dummy variable model (LSDV) estimation, where interviewer fixed-effect dummy variables are included, with the ordinary least square (OLS) standard errors in parentheses in Tables 3A-4B;[5]
2. Fixed effect estimation in the linear panel data analysis (FE), where the variables are de-meaned by interviewers to account for interviewer heterogeneity, with a sandwich form robust standard errors in squared brackets in Tables 3A-4B.

The parameter estimates obtained from both estimation methods are identical (except for β_0), but the standard errors differ. Although the OLS standard errors are inconsistent and known to be biased downward, we decided to report them for interested readers to compare the two results.

The literature suggests that the first interviewees may have a higher probability of getting a better score due to interviewer's time or effort constraints, which is known as the primacy effect. Furthermore, there is evidence of a favorable outcome for the last interviewees, known as the recency effect, which cannot be solely explained by limited memory and is observed in both end-of-sequence and step-by-step evaluations. End-of-sequence evaluation refers to evaluating all presentations before assigning scores, while step-by-step evaluation involves scoring after each presentation.

We would examine whether including being the first or last interviewee within the group leads to any differences in our estimates for *Order*. The new regression model is as follows:

Model (2):

$Y_{i,l} = \beta_0 + \beta_1 * Order_i + \beta_2 * First_i + \beta_3 * Last_i + \beta_4 * X_i + u_i + v_l + \epsilon_{i,l}$, $l = j_1$ or j_2, where $j_1 < j_2 \in \{1, \cdots, J\}$.

[5] We have 310 and 470 interviewers for HSPR and HSR respectively. There might be interviewer specific heterogeneity. For example, some interviewers are generous, and others might be more stringent in their evaluations. Interviewer fixed effects will capture these time invariant interviewer specific variations. In our data, university organize several hours of training as well as hour long open discussion about the interview questions and suggested answers. We provide results without including interviewer fixed effects in the Appendix A. The results do not change almost at all even if we do not include interviewer fixed effects meaning that there is not much variation across interviewers.

Table 3A. High School Principal Recommendation (HSPR) Track: Model (1)

VARIABLES	Interview with Article (IWA)			High School Record Review (HSRR)	
	Application Ability	Analytical Ability	Critical Thinking	Preparedness	Motivation
	(1) score	(2) score	(3) score	(4) score	(5) score
Order	-0.00241	-0.00231	-0.00138	0.00101	0.00214
	(0.00145)*	(0.00147)	(0.00144)	(0.00130)	(0.00122)*
	[0.00174]	[0.00183]	[0.00178]	[0.00159]	[0.00133]
Constant	3.248***	3.431***	3.459***	2.895***	2.830***
	(0.232)	(0.235)	(0.229)	(0.354)	(0.332)
Dependent Mean	3.176	3.365	3.044	3.481	3.566
Dependent S.D.	1.108	1.128	1.107	1.027	0.968
Observations	8,162	8,162	8,162	8,160	8,160
R-squared	0.145	0.151	0.165	0.196	0.208

Notes: Departments, Interview slots (Saturday morning and afternoon, Sunday morning and afternoon), and interviewer dummy variables are included as control variables. ( ) are the OLS standard errors and [ ] the robust standard errors from the FE estimation.
* is significant at 10%, ** at 5%, *** at 1%.

Table 3B. High School Record (HSR) Track: Model (1)

VARIABLES	Application Ability	Analytical Ability	Critical Thinking	Motivation
	(1) score	(2) Score	(3) score	(4) score
Order	0.00247	0.00687	0.00425	0.00217
	(0.00232)	(0.00240)***	(0.00237)*	(0.00192)
	[0.0029]	[0.0030]**	[0.00299]	[0.00224]
Constant	1.995***	1.935***	2.650***	2.283***
	(0.278)	(0.287)	(0.283)	(0.230)
Dependent Mean	3.191	3.383	3.245	3.420
Dependent S.D.	1.112	1.154	1.145	0.972
Observations	8,044	8,044	8,044	8,044
R-squared	0.142	0.149	0.157	0.232

Notes: Departments, Interview slots (Saturday morning and afternoon, Sunday morning and afternoon), and interviewer dummy variables are included as control variables. ( ) are the OLS standard errors and [ ] the robust standard errors from the FE estimation.
* is significant at 10%, ** at 5%, *** at 1%.

The $First_i$ is a dummy variable indicating i is the first interviewee of the interview group i.e. $First_i = I\{Order_i = 1\}$. The $Last_i$ is a dummy variable indicating i is the last interviewee of the interview group.

Table 4A. High School Principal Recommendation (HSPR) Track: Model (2)

VARIABLES	Interview with Article (IWA)			High School Record Review (HSRR)	
	Application Ability	Analytical Ability	Critical Thinking	Preparedness	Motivation
	(1) score	(2) score	(3) score	(4) score	(5) score
Order	-0.00386	-0.0045	-0.00283	0.000419	0.000778
	(0.00162)**	(0.00164)***	(0.00160)*	(0.00145)	(0.00136)
	[0.0021]*	[0.0022]**	[0.0022]	[0.00179]	[0.0015]
First	0.0337	-0.0616	-0.0246	0.0154	-0.0493
	(0.0766)	(0.0644)	(0.0627)	(0.0570)	(0.0534)
	[0.0784]	[0.0778]	[0.0774]	[0.0661]	[0.0624]
Last	0.222	0.224	0.163	0.0919	0.128
	(0.0635)***	(0.0644)***	(0.0627)***	(0.0570)	(0.0534)**
	[0.0819]***	[0.0816]***	[0.0818]**	[0.0788]	[0.0676]*
Constant	3.256***	3.454***	3.472***	3.068***	2.866***
	(0.232)	(0.235)	(0.229)	(0.354)	(0.332)
Dependent Mean	3.176	3.365	3.044	3.481	3.566
Dependent S.D.	1.108	1.128	1.107	1.027	0.968
Observations	8,162	8,162	8,162	8,160	8,160
R-squared	0.147	0.153	0.165	0.197	0.208

Notes: Departments, Interview slots (Saturday morning and afternoon, Sunday morning and afternoon), and interviewer dummy variables are included as control variables in addition to First and Last. ( ) are the OLS standard errors and [ ] the robust standard errors from the FE estimation.
* is significant at 10%, ** at 5%, *** at 1%.

Table 4B. High School Record (HSR) Track: Model (2)

VARIABLES	Application Ability	Analytical Ability	Critical Thinking	Motivation
	(1) score	(2) Score	(3) score	(4) score
Order	0.00263	0.00630	0.00301	0.00100
	(0.00277)	(0.00287)**	(0.00283)	(0.00229)
	[0.00357]	[0.00357]*	[0.00371]	[0.0027]
First	-0.0164	-0.0499	-0.0299	0.00154
	(0.0546)	(0.0565)	(0.0557)	(0.0452)
	[0.0651]	[0.0725]	[0.0692]	[0.0546]
Last	-0.0247	-0.0196	0.0365	0.0641
	(0.0546)	(0.0565)	(0.0557)	(0.0452)
	[0.0638]	[0.0654]	[0.0682]	[0.0517]
Constant	1.997***	1.944***	2.660***	2.290***
	(0.278)	(0.287)	(0.284)	(0.230)
Dependent Mean	3.191	3.383	3.245	3.420
Dependent S.D.	1.112	1.154	1.145	0.972
Observations	8,044	8,044	8,044	8,044
R-squared	0.142	0.149	0.157	0.232

Notes: Departments, Interview slots (Saturday morning and afternoon, Sunday morning and afternoon), and interviewer dummy variables are included as control variables in addition to First and Last. ( ) are the OLS standard errors and [ ] the robust standard errors from the FE estimation.
* is significant at 10%, ** at 5%, *** at 1%.

IV. *Results*

Model (1)'s results are presented in Tables 3A and 3B and Model (2)'s in Tables 4A and 4B. In Tables 3A, we present regression results for the HSPR track within the IWA section, focusing on three attributes (Application Ability, Analytical Ability, and Critical Thinking), as well as the HSRR section, centering on two attributes (Preparedness and Motivation).

For Application Ability, in Column (1), we can see the average grade score decreases as one gets interviewed later because the coefficient for *Order* is negative. Putting aside its statistical insignificance by the robust standard error,[6] the magnitude of the effect is very small. To put the magnitude in context, considering there were 18 to 31 interviewees in a group, we can estimate the 31st interviewee would receive a 0.07 (= -0.00241* (31-1)) lower grade compared to the 1st interviewee. The sample average of dependent variable is 3.176 and a 0.07 is only 2% of the sample average and 0.063 times the dependent variable's sample standard deviation (1.108).

For Analytical Ability, Column (2) indicates that the analytical ability score decreases by 0.00231 as the interview order increases by one. This effect is neither statistically nor, as we saw in the Application Ability, practically significant.

Regarding Critical Thinking, Column (3) shows that as the interview order increases, the Critical Thinking score decreases by 0.00138, and this effect is not statistically and practically significant.

For Preparedness, Column (4) indicates that the *Order* variable does not have a statistically significant effect on the preparedness score and the magnitude is even smaller.

For Motivation, Column (5) shows that the *Order* variable is statistically insignificant[7] and the magnitude of the effect is small. A 31st interviewee would receive a 0.06(=0.00214*(31-1)) higher grade compared to a 1st interviewee, which corresponds to only 1.8% (=0.06/3.566) of the sample average and 0.062 times sample standard deviation of the dependent variable.

Overall, none of five outcomes are statistically significant at the conventional level when we use robust standard errors. In addition, the magnitude of being the last interviewee, which is 31st in our data, ranges from -0.07 to 0.06, which is a maximum of 2.3% of the sample mean.

In Table 3B, we present the regression results for the HSR track on four features (Application Ability, Analytical Ability, Critical Thinking, and Motivation) using Model (1). The table format is the same as Table 3A. For Analytical Ability the *Order* coefficient is statistically significant at 5% significance level but the magnitude is very small. Considering that the number of interviewees varies between 8 and 23, we can see the 23rd interviewee would receive 0.15 higher grade on average for Analytical Ability than the first interviewee, which corresponds to 4.4% of the sample mean (3.383). In the other three cases, the *Order* is statistically insignificant and smaller in magnitude as well.

Table 4A presents HSPR track results using Model (2). In column (1), the results show that the *Order* statistically significantly impacts the Application Ability score negatively. A 31st

[6] It is significant at 90% confidence level by the OLS standard error, but the OLS standard error is inconsistent as mentioned earlier.

[7] It is significant at 90% significant level by the OLS standard error, but the OLS standard error is inconsistent as mentioned earlier.

interviewee would receive a 0.15 (=-0.00386*30-0.0337) lower grade point when she is not the last interviewee or a 0.07 (=-0.00386*30+0.222-0.0337) higher grade score when she is the last interviewee on average than the first interviewee of the group. It is just 4.7% or 2.3% of the sample mean value. In Column (2), the coefficient for *Order* is -0.0045, indicating that the 31st interviewee would receive a 0.15 (=-0.0045*30+0.224+0.0616) higher or a 0.07 (=-0.0045*30 +0.0616) lower grade score than the first interviewee depending on whether or not she is the last interviewee, similarly to Application Ability case. In Column 3, the *Order* variable is statistically insignificant and, the same calculation shows the 31st interviewee would receive a 0.10 (=-0.00283*30+0.163+0.0246) higher if she is the last or a 0.06 (=-0.00283*30+0.0246) lower grade point if she is not the last than the first interviewee. In Columns (4) and (5), neither Preparedness nor Motivation is statistically significant. The same calculations show the 31st and last interviewee would receive 0.09 (= 0.000419*30 + 0.0919-0.0154) and 0.2 (= 0.000778*30+0.128+0.0493) higher grade points than the first interview in Preparedness and Motivation, respectively. When she is not the last interviewee, she will get 0.003 (= 0.000419*30-0.0154) lower and 0.07 (=0.000778*30+0.0493) higher than the first interviewee.

Overall, the differences between an earlier and a later interviewee, regardless of her being the last, are not at all substantial. The conclusion is not affected by whether we use Model (1)'s results or Model (2)'s.

Table 4B presents HSR track results using Model (2). The *Order* variable is statistically significant only for Analytical Ability in column (2), with a magnitude similar to Model (1) presented in Table 3B column (2). Overall, we found qualitatively similar results between Model (1) and Model (2).

In summary, our estimates for the coefficient of *Order* range from -0.00241 to 0.00687 based on Model (1) and from -0.0045 to 0.0063 for Model (2). Readers may wish to compare our results with those in existing literature, such as Bruine de Bruin's (2006) study. In her Table 1, the coefficient for the serial position is 0.02, which appears to be vastly different from our findings. It was statistically significant but she did not interpret the magnitude. It is worth noting that a direct comparison with Bruine de Bruin's (2006) is not feasible due to differences in the dependent variable (numeric grade vs. percentile rank), and interpretation of the coefficient estimate is not easy. However, we found that she would have obtained a coefficient estimate around 0.0017 if she used numeric grades like ours instead of percentile ranks through our simulations that mimicked her data generation process.[8] The order effect in Bruine de Bruin (2006) is very similar to ours.

< Another possibility of no order effect >

Many of our estimates yield statistically insignificant results despite the ample sample sizes. Moreover, the magnitudes of the coefficients are notably small, irrespective of their statistical significance. In essence, our findings strongly suggest the absence of an order effect.

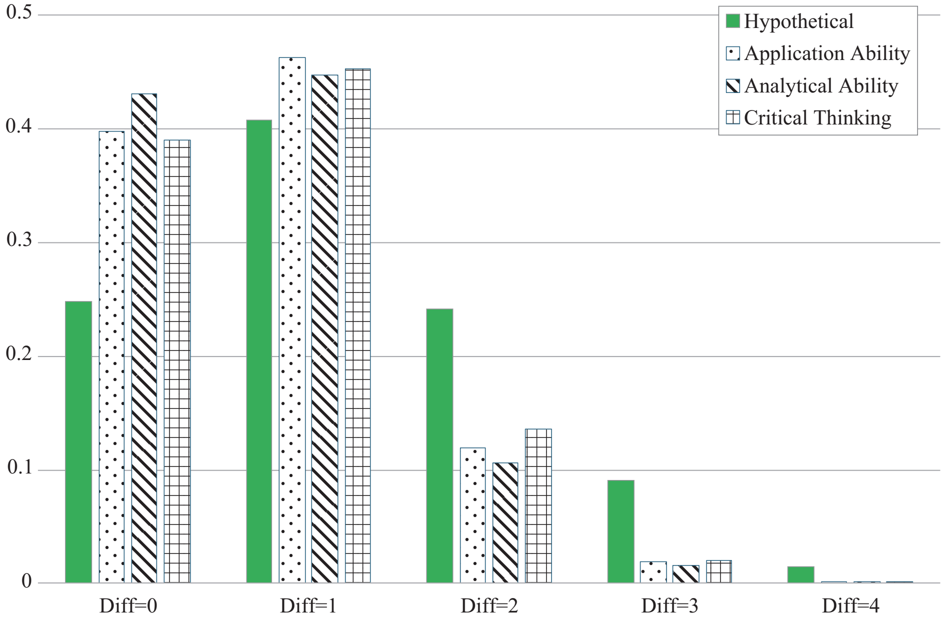
However, it's worth considering whether our failure to detect an order effect stems from interviewers assigning random scores. If interviewers indeed allocated grades arbitrarily, regardless of the substance of interviewees' responses, we would still observe a lack of an order effect. While this scenario is highly improbable, we seek to explore this possibility here.

Essentially, we aim to demonstrate that the observed grading patterns deviate from the

[8] We can provide simulation details for the interested reader upon reader's request.

FIGURE 1. SCORE DIFFERENCES BETWEEN TWO INTERVIEWERS, INTERVIEW WITH ARTICLE

IWA



expected ones when interviewers assign scores randomly. Let $y_{i,1}$ and $y_{i,2}$, ranging from $\{1, 2, \cdots, 5\}$, be the grades given to interviewee i by interviewers 1 and 2, respectively. If the two interviewers independently and randomly assign grades from $\{1, 2, \cdots, 5\}$ with probabilities $Pr[y_{i,1}=k]=Pr[y_{i,2}=k]=p_k$, where $\sum_{k=1}^{5} p_k=1$, the probability of both grades coinciding is calculated as:

$$Pr[y_{i,1}=y_{i,2}]=\sum_{k=1}^{5} Pr[y_{i,1}=k, y_{i,2}=k]=\sum_{k=1}^{5} Pr[y_{i,1}=k]Pr[y_{i,2}=k]=\sum_{k=1}^{5} p_k^2.$$

We can compute the probability of the two grades being d grade apart as follows.

$$\begin{aligned} Pr[|y_{i,1}-y_{i,2}|=d] &= \sum_{k=1}^{5-d}(Pr[y_{i,1}=k, y_{i,2}=k+d]+Pr[y_{i,1}=k+d, y_{i,2}=k]) \\ &= 2\sum_{k=1}^{5-d} Pr[y_{i,1}=k, y_{i,2}=k+d]=2\sum_{k=1}^{5-d} p_k p_{k+d} \end{aligned}$$

p_k can be 1/5 for all $k=1, \cdots, 5$ but it is more natural to think that interviewers give mid-range grades more frequently than they do extreme grades such as 1 or 5. So, we took the actual probabilities from all two tracks' results. For example, $p_5=\frac{No.\ of\ A+}{2\times No.\ of\ interviewees}$. The estimated probabilities are: $p_1=0.05$; $p_2=0.22$; $p_3=0.33$; $p_4=0.25$; $p_5=0.15$. Plugging-in these numbers, we obtain probabilities of grades to coincide is 0.248, to be one grade apart 0.4072, two grades apart 0.242, three grades apart 0.091, and lastly four grade apart 0.015.

In Figures 1 to 3, the horizontal axis represents the grade difference between two

FIGURE 2. SCORE DIFFERENCES BETWEEN TWO INTERVIEWERS, HIGH SCHOOL RECORD REVIEW

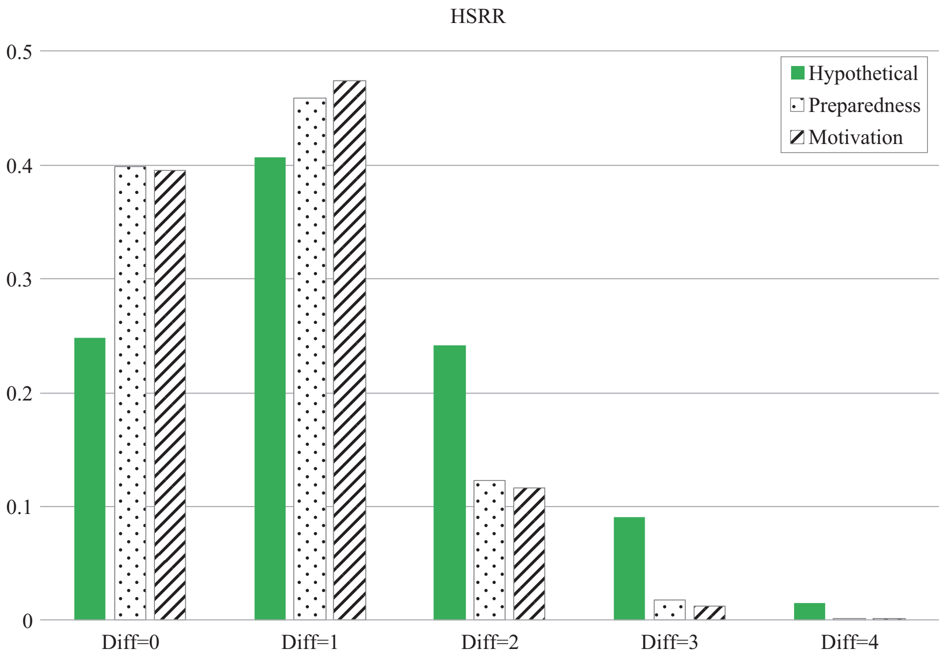
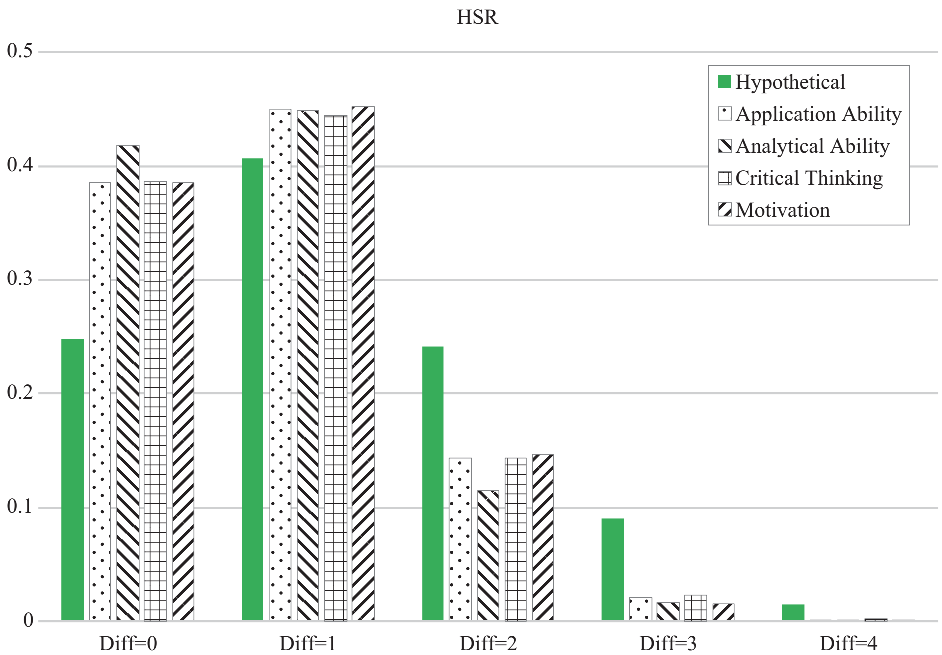


FIGURE 3. SCORE DIFFERENCES BETWEEN TWO INTERVIEWERS, HIGH SCHOOL RECORD



interviewers for an interviewee, ranging from 0 to 4, and the vertical axis indicates the percentage out of the total number of interviewees. Let us utilize Analytical Ability of IWA section presented in Figure 1 to elucidate the results; the other results adhere to the same format. Forty three percent of interviewees (1,757 out of 4,081 groups) received the same grade from both interviewers, a notable deviation from the hypothetical value of 24.8% shown as a green solid bar. The observed probability of being one grade apart is 44.69% (1,824 out of 4,081 groups), differing from the derived value of 40.72%. Moreover, the observed probability of being two grades apart is 10.56% (431 out of 4,081 groups), significantly lower than the previously calculated 24.2%. The probabilities of being three and four grades apart are 1.5% and 0.1%, respectively, in the data, contrasting with the hypothetical values of 9.1% and 1.5%. Overall, comparisons in Figure 1 illustrate substantial disparities between observed and hypothetical probabilities for all features, suggesting that interviewers do not assign grades randomly. The same comparisons are applicable to Figures 2 (HSRR section) and 3 (HSR track). These findings further reinforce the argument that interviewers' grade assignments exhibit non-random patterns.

V. *Path Dependency: A Different View*

Some literature has shown presence of path dependency in sequential decision making. Path dependency refers to a phenomenon wherein earlier evaluations influence later ones in a sequential presentation, roughly speaking. Jin et al. (2024) used administrative data of an emergency department (ED) in a Southeast Asian country and examined physicians' decisions on whether or not to admit a patient. They found physician's admission probability is higher when they admitted immediate previous patient compared to when they didn't even after controlling for patient's demographics and medical conditions, and physician fixed effects. Bhargava and Fisman (2014) used speed dating experiment data conducted by Fisman et al. (2006, 2008). They also found a male participant's willingness to see his partner again (a "yes" response) decreases when his previous partner's attractiveness is greater compared to when his previous partner's attractiveness is less, after controlling for the participant fixed effect. Interestingly, they could not obtain similar results from female participants' data.[9] Radbruch and

[9] We believe their findings are subject to critique. In Jin et al. (2024), the presentation order of patients to physicians depends on the ED's triage protocol, so the presentation order may not be random (see, for example, Nakao et al., 2017). We can reasonably assume that more severe patients will be attended to sooner than less severe patients, even if the more severe ones arrive later, based on the triage protocol. Another issue with their results is the use of a usual fixed effect (FE) estimation while the model is a dynamic panel data model, for which Arellano-Bond's system GMM estimation is recommended. The randomization of presentation order in Bhargava and Fisman (2014)'s data seemed sufficient, but we suspect their results might suffer from the so-called errors-in-variable problem. They employed the following model.

$$Y_{i,l}=\beta_0+\beta_1 * X_l+\beta_2 * X_{l-1}+e_{i,l}.$$

Here, $Y_{i,l}$ is the "yes-or-no" decision of evaluator i for l-th partner and X_l and X_{l-1} the attractiveness of l-th and $l-1$-st partners respectively. The attractiveness of partners is measured by two research assistants (RAs) independently, not by the participants themselves. Admittedly, the RAs' evaluations will contain random deviations from the participants' subjective evaluations. Since participants' decisions would be based on self-evaluations of partners' attractiveness rather than the RAs' 'objective' attractiveness measure, their X represents mismeasured participant evaluations, hence an errors-

Schiprowski (2024) found that quality of the previous interviewees negatively influences the evaluation on the current interviewee, based on data from high-stake grant admissions.

The relationship between path dependency and order bias is not yet thoroughly studied and we don't attempt to investigate it either. Our aim here is to examine if our data show path dependency additionally by estimating the following model.

Model (3):

$$Y_{i,l}=\beta_0+\beta_1 * Y_{i,l-1}+v_i+e_{i,l},$$

where $Y_{i,l}$ is interviewer i's evaluation of the l-th order interviewee in the given interview slot. The v_i is interviewer-specific (or any type of order-invariant) unobservable heterogeneity. We used Arellano-Bond's system GMM to estimate β_1 because the model is a dynamic panel data model. Due to the nature of the GMM estimation that differenced variables are used, effects of any type of order-invariant variables such as interview slot are eliminated so we don't need to include those variables in the regression.

The results are reported in Table 5A and 5B. The $L.y$ rows in the tables present $\hat{\beta}_1$s. Figures in parentheses underneath are robust standard errors. The rows named *Lagged y for IVs* present lag orders of $Y_{i,l}$ that were used for IVs. For example, *L6~L9* in Application Ability in IWA means $Y_{i,l-6}\sim Y_{i,l-9}$ were used for IVs. Two rows under the row name *Arellano-Bond test (p-val) for* show tests for autocorrelation in residuals of differenced regression that the system GMM uses. It is required that AR(1) test rejects the null of no serial correlation and AR(2) fails to reject the no serial correlation null hypothesis. Two rows under the row name *OverID Test (p-val)* shows Sargan's and Hansen's overidentification tests respectively. It is required that the null hypothesis of no excess endogenous variables among the IVs should not be rejected by the overidentification tests. *P*-values were presented for the four tests. All the regressions passed required signification tests. The numbers of observation decreased since we use one period lagged variable.

All $\hat{\beta}_1$s but that of *Application Ability* in *IWA* section (column (1) in Table 5A) were not significant at conventional levels. $\hat{\beta}_1$ of *Application Ability* in *IWA* section is 0.0533 with a *p*-value smaller than 0.001. Let us understand how to interpret the estimate. Since $E[Y_l|Y_{l-1}, v]=(\beta_0+v)+\beta_1 * Y_l$, one can see $E[Y_2|Y_1, v]=(\beta_0+v)+\beta_1 * Y_1$, $E[Y_3|Y_1, v]=(\beta_0+v)+\beta_1 * E[Y_2|Y_1, v]=(\beta_0+v)(1+\beta_1)+\beta_1^2 * Y_1$, etc. So, we can recursively compute the expected value of Y at L-th order (Y_L) given Y_1 as follows.

$$E[Y_L|Y_1, v]=(\beta_0+v)(1+\beta_1+\cdots+\beta_1^{L-2})+\beta_1^{L-1} * Y_1.$$

If the interviewer gives one unit higher score to the first interviewee, then one can expect the L-th order interviewee will receive $E[Y_L|Y_1+1, v]-E[Y_L|Y_1, v]=\beta_1^{L-1}$point higher score on average. Given the coefficient estimate (0.0533), $\hat{\beta}_1^{L-1}$ is about 0.00015 ($=0.0533^3$), small enough already at $L=4$. So the magnitude of effect of path dependency, although statistically

in-variable situation. It is known that errors-in-variable cause attenuation bias when there is only one regressor in a linear regression model, but the directions or magnitudes of bias for a multiple linear regression model, such as the one in their analysis, are not clear (see, for example, Wooldridge, 2020, pp.310-313).

Table 5A. High School Principal Recommendation (HSPR) Track: Model (3)

VARIABLES		Interview with Article (IWA)			High School Record Review (HSRR)	
		Application Ability	Analytical Ability	Critical Thinking	Preparedness	Motivation
		(1) score	(2) score	(3) score	(4) score	(5) score
L.y		0.0533***	0.0238	0.0215	0.0603	0.3766
		(0.0160)	(0.0166)	(0.0141)	(0.2519)	(0.3985)
Constant		3.004***	3.285***	2.978***	3.272***	2.227
		(0.0580)	(0.0635)	(0.0494)	(0.8764)	(1.4189)
Lagged y for IVs		L6~L9	L2~L8	L2~L8	L8~L17	L7~L10
Arellano-Bond test (p-val) for	AR(1)	0.027	0.000	0.000	0.032	0.045
	AR(2)	0.250	0.275	0.883	0.623	0.299
OverID test (p-val)	Sargan	0.731	0.114	0.149	0.173	0.316
	Hansen	0.810	0.207	0.266	0.100	0.314
Dependent Mean		3.173	3.365	3.043	3.481	3.569
Dependent S.D.		1.113	1.132	1.112	1.030	0.972
Observations		7,542	7,542	7,542	7,540	7,540
R-squared		.	.	.	.	.

Notes: We estimated the models with STATA's *xtabond2* command with *gmmstyle* and *robust* options. So, ( ) are the robust standard errors. * is significant at 10%, ** at 5%, *** at 1%. Figures in Arellano-Bond, Sangan, and Hansen tests are p-values.

Table 5B. High School Record (HSR) Track: Model (3)

VARIABLES		Application Ability	Analytical Ability	Critical Thinking	Motivation
		(1) score	(2) Score	(3) score	(4) score
L.y		0.1544	-0.0938	0.2809	0.1750
		(0.2957)	(0.2139)	(0.2633)	(0.2212)
Constant		2.701***	3.706***	2.338***	2.823***
		(0.9432)	(0.7245)	(0.8547)	(0.7541)
Lagged y for IVs		L4~L11	L4~L11	L5~L11	L4~L11
Arellano-Bond test (p-val) for	AR(1)	0.036	0.035	0.006	0.003
	AR(2)	0.571	0.962	0.240	0.267
OverID test (p-val)	Sargan	0.713	0.657	0.371	0.315
	Hansen	0.707	0.649	0.402	0.329
Dependent Mean		3.194	3.389	3.248	3.421
Dependent S.D.		1.120	1.160	1.153	0.977
Observations		7,104	7,104	7,104	7,104
R-squared		.	.	.	.

Notes: We estimated the models with STATA's *xtabond2* command with *gmmstyle* and *robust* options. So, ( ) are the robust standard errors. * is significant at 10%, ** at 5%, *** at 1%. Figures in Arellano-Bond, Sangan, and Hansen tests are p-values.

significant, is negligible. For other $\widehat{\beta}_1$s, putting aside that they are not statistically significant at even 10% level, similar conclusions can be made. Basically, $\widehat{\beta}_1^{L-1}$ drops to negligible levels at $L=3$ or $L=4$ already for all $\widehat{\beta}_1$s. Overall, our analysis shows no evidence for substantial path dependency.

VI. *Conclusion*

Numerous studies across different fields have established that the order of presentation can have a significant impact on evaluation results. However, as addressed by other researchers or even by the very authors, the data used in existing papers have issues that potentially cause biases: non-random order of contestants; unobservable heterogeneities in the situations that potentially affected the outcomes; limited sample sizes, especially small number of judges.

We had a luxury that we could use a unique data set that we strongly believe is free from those issues. Randomization of order was carefully done at the institutional level. The Interview process was devised, regulated, and monitored to a great detail so that uncontrollable factors that could affect interviewers or interviewees are eliminated as much as possible. The number of interviewers was hundreds.

To summarize, we found no significant order effects. Most of the order effect parameter estimates are small in magnitude and statistically insignificant. One statistically significant case, focusing in Model (1), was Analytical Ability in HSR track but its magnitude is still small that it would not result in substantial changes in the evaluation outcomes. This finding aligns with two existing studies: Bruine de Bruin's (2006) study is in the same ballpark with ours; Miller and Krosnick (1998) concluded that the magnitude of name-order effects suggests unsubstantial effects to undermine the democratic process. The conclusion was qualitatively the same regarding the so-called "path dependency".

To be humble, as we should be, we may not conclude from our results, along with those of Bruine de Bruin (2006) and Miller and Krosnick (1998), that presentation order has no effect on decision making. However, a concern may arise regarding the findings of order effects, questioning whether they stem from unclean data used in those studies. Perhaps we can go as far as to say that carefully designed institutional regulations may eliminate or reduce order effects, if indeed existing findings for the presence of order effects are due to use of unclean data. At least, we think our finding implies that carefully designed institutional details are needed and ought to be implemented if society wants to mitigate unwanted effects from presentation order in human decision-makings.

Appendix

Appendix A: Regression results of models (1) and (2) without interviewer fixed effects

Below, we present regression results of models (1) and (2) without interviewer fixed effects. Each of Appendix Tables 3A～4B corresponds to Tables 3A～4B in the main text, respectively.

Appendix Table 3A. High School Principal Recommendation (HSPR) Track: Model (1)

VARIABLES	Interview with Article (IWA)			High School Record Review (HSRR)	
	Application Ability	Analytical Ability	Critical Thinking	Preparedness	Motivation
	(1) score	(2) score	(3) score	(4) score	(5) score
Order	-0.00242	-0.00238	-0.00137	0.00102	0.00218
	(0.00151)	(0.00152)	(0.00150)	(0.00141)	(0.00133)
Constant	3.353***	3.321***	3.434***	3.550***	3.567***
	(0.111)	(0.112)	(0.111)	(0.104)	(0.0983)
Dependent Mean	3.176	3.365	3.044	3.481	3.566
Dependent S.D.	1.108	1.128	1.107	1.027	0.968
Observations	8,162	8,162	8,162	8,160	8,160
R-squared	0.061	0.080	0.068	0.052	0.045

Notes: Departments, Interview slots (Saturday morning and afternoon, Sunday morning and afternoon), and interviewer dummy variables are included as control variables. ( ) are the OLS standard errors and [ ] the robust standard errors from the FE estimation.
* is significant at 10%, ** at 5%, *** at 1%.

Appendix Table 3B. High School Record (HSR) Track: Model (1)

VARIABLES	Application Ability	Analytical Ability	Critical Thinking	Motivation
	(1) score	(2) Score	(3) score	(4) score
Order	0.00238	0.00677***	0.00412*	0.00209
	(0.00242)	(0.00248)	(0.00247)	(0.00210)
Constant	2.792***	2.817***	3.262***	3.545***
	(0.138)	(0.142)	(0.141)	(0.120)
Dependent Mean	3.191	3.383	3.245	3.420
Dependent S.D.	1.112	1.154	1.145	0.972
Observations	8,044	8,044	8,044	8,044
R-squared	0.041	0.064	0.059	0.054

Notes: Departments, Interview slots (Saturday morning and afternoon, Sunday morning and afternoon), and interviewer dummy variables are included as control variables. ( ) are the OLS standard errors and [ ] the robust standard errors from the FE estimation.
* is significant at 10%, ** at 5%, *** at 1%.

Appendix Table 4A. High School Principal Recommendation (HSPR) Track: Model (2)

VARIABLES	Interview with Article (IWA)			High School Record Review (HSRR)	
	Application Ability	Analytical Ability	Critical Thinking	Preparedness	Motivation
	(1) score	(2) score	(3) score	(4) score	(5) score
Order	-0.00387**	-0.00459***	-0.00282*	0.000431	0.000818
	(0.00168)	(0.00169)	(0.00167)	(0.00157)	(0.00148)
First	0.0339	-0.0621	-0.0243	0.0153	-0.0491
	(0.0659)	(0.0665)	(0.0657)	(0.0614)	(0.0581)
Last	0.222***	0.225***	0.163**	0.0914	0.127**
	(0.0659)	(0.0665)	(0.0657)	(0.0614)	(0.0581)
Constant	3.361***	3.342***	3.446***	3.553***	3.580***
	(0.112)	(0.113)	(0.111)	(0.104)	(0.0987)
Dependent Mean	3.176	3.365	3.044	3.481	3.566
Dependent S.D.	1.108	1.128	1.107	1.027	0.968
Observations	8,162	8,162	8,162	8,160	8,160
R-squared	0.063	0.081	0.068	0.053	0.046

Notes: Departments, Interview slots (Saturday morning and afternoon, Sunday morning and afternoon), and interviewer dummy variables are included as control variables in addition to First and Last. ( ) are the OLS standard errors and [ ] the robust standard errors from the FE estimation.
* is significant at 10%, ** at 5%, *** at 1%.

Appendix Table 4B. High School Record (HSR) Track: Model (2)

VARIABLES	Application Ability	Analytical Ability	Critical Thinking	Motivation
	(1) score	(2) Score	(3) score	(4) score
Order	0.00249	0.00617**	0.00282	0.000892
	(0.00290)	(0.00297)	(0.00295)	(0.00252)
First	-0.0174	-0.0508	-0.0313	0.000756
	(0.0571)	(0.0585)	(0.0582)	(0.0496)
Last	-0.0234	-0.0183	0.0381	0.0651
	(0.0571)	(0.0585)	(0.0582)	(0.0495)
Constant	2.794***	2.827***	3.272***	3.551***
	(0.139)	(0.142)	(0.142)	(0.121)
Dependent Mean	3.191	3.383	3.245	3.420
Dependent S.D.	1.112	1.154	1.145	0.972
Observations	8,044	8,044	8,044	8,044
R-squared	0.041	0.064	0.059	0.054

Notes: Departments, Interview slots (Saturday morning and afternoon, Sunday morning and afternoon), and interviewer dummy variables are included as control variables in addition to First and Last. ( ) are the OLS standard errors and [ ] the robust standard errors from the FE estimation.
* is significant at 10%, ** at 5%, *** at 1%.

References

Bindler, A. and R. Hjalmarsson (2019), "Path Dependency in Jury Decision Making," *Journal of the European Economic Association* 17(6), pp.1971-2017.

Danziger, S., J. Levav and L. Avnaim-Pesso (2011), "Extraneous Factors in Judicial Decisions." *Proceedings of the National Academy of Sciences* 108(17), pp.6889-6892.

De Bruin, W.B. (2006), "Save the Last Dance II: Unwanted Serial Position Effects in Figure Skating Judgments," *Acta Psychologica* 123(3), pp.299-311.

Feenberg, D., I. Ganguli, P. Gaule and J. Gruber (2017), "It's Good to Be First: Order Bias in Reading and Citing NBER Working Papers," *Review of Economics and Statistics* 99(1), pp.32-39.

Fisman, R., S. Iyengar, E. Kamenica and I. Simonson (2006), "Gender Differences in Mate Selection: Evidence from a Speed Dating Experiment," *Quarterly Journal of Economics* 121(2), pp.673-697.

—— (2008), "Racial Preferences in Dating," *The Review of Economic Studies* 75(1), pp.117-132.

Glejser, H. and R. Heyndels (2001), "Efficiency and Inefficiency in the Ranking in Competitions: The Case of the Queen Elisabeth Music Contest," *Journal of Cultural Economics* 25, pp.109-129.

Glöckner, A. (2016), "The Irrational Hungry Judge Effect Revisited: Simulations Reveal That the Magnitude of the Effect is Overestimated," *Judgment and Decision Making* 11(6), pp.601-610.

Jin, L., R. Tang, H. Ye, J. Yi and S. Zhong (2024), "Path Dependency in Physician Decisions," *The Review of Economic Studies* 91(5), pp.2916-2953.

Miller, J.M. and J.A. Krosnick (1998), "The Impact of Candidate Name Order on Election Outcomes," *Public Opinion Quarterly*, pp.291-330.

Mussweiler, T. (2003), "Comparison Processes in Social Judgment: Mechanisms and Consequences," *Psychological Review* 110(3), pp.472-489.

Nakao, H., I. Ukai, and J. Kotani (2017), "A Review of the History of the Origin of Triage from a Disaster Medicine Perspective," *Acute Medicine & Surgery* 4(4), pp.379-384.

Neilson, W.S. (1998), "Reference Wealth Effects in Sequential Choice," *Journal of Risk and Uncertainty* 17, pp.27-48.

Novemsky, N. and R. Dhar (2005), "Goal Fulfillment and Goal Targets in Sequential Choice," *Journal of Consumer Research* 32(3), pp.396-404.

Page, L. and K. Page (2010), "Last Shall Be First: A Field Study of Biases in Sequential Performance Evaluation on the Idol Series," *Journal of Economic Behavior & Organization* 73(2), pp.186-198.

Radbruch, J. and A. Schiprowski (2024), "Interview Sequences and the Formation of Subjective Assessments," *The Review of Economic Studies*, rdae039, https://doi.org/10.1093/restud/rdae039.

Sarafidis, Y. (2007), "What Have You Done for Me Lately? Release of Information and Strategic Manipulation of Memories," *The Economic Journal* 117(518), pp.307-326.

Weinshall-Margel, K. and J. Shapard (2011), "Overlooked Factors in the Analysis of Parole Decisions," *Proceedings of the National Academy of Sciences of the United States of*

America 108(42), E833-E833.
Wooldridge, J.M. (2020), *Introductory Econometrics: A Modern Approach*, 7th ed., Cengage.