UNIVERSITY
OF TECHNOLOGY
SYDNEY

# Essays in Applied Microeconomics with a Focus on Gender Heterogeneity

**by Liqing Chen**

Thesis submitted in fulfilment of the requirements for the degree of

**Doctor of Philosophy**

under the supervision of Prof. Adeline Delavande, Prof. Peter Siminski and Dr. Mario Fiorini

University of Technology Sydney
Faculty of Business

November 2025

# Certificate of Original Authorship

I, *Liqing Chen*, declare that this thesis is submitted in fulfilment of the requirements for the award of *Doctor of Philosophy*, in the *Business School* at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution. This research is supported by the Australian Government Research Training Program.

Production Note:
Signature: Signature removed prior to publication.

Date: 31st May

# Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor Adeline Delavande for her committed and invaluable support throughout my journey. I am also sincerely thankful to my co-supervisors Peter Siminski and Mario Fiorini.

The gambling paper presented in Chapter 3 was financially supported by the UTS Behavioural Lab. I would also like express my special thanks to Elif Incekara-Hafalir for her detailed feedback on this paper.

I am grateful to the University of Technology Sydney (UTS) and the UTS Business School, for giving me the opportunity to undertake and complete my doctoral thesis. I also thank many faculty members of the Economics Discipline Group (EDG) for their valuable comments on the three chapters of this thesis I acknowledge the use of generative AI tools, such as ChatGPT, for providing code templates and assisting in proofreading the final draft.

Lastly, to my family and friends who believed in me — particularly my mother whose quiet understanding and support meant everything. She respected my struggles and never once asked about my progress when I needed space the most.

# Abstract

This thesis comprises three empirical and experimental chapters, each examining different contexts but unified by a focus on understanding heterogenous effects arising from gender and other socio-economic factors.

**Chapter 1** investigates the impact of an information intervention on university students' perceptions of employer-valued soft skills, skill investment decisions, and job search behaviours, with a particular focus on gender differences. Non-cognitive skills are increasingly recognised for their strong predictive power in economic outcomes and their resistance to substitution by automation and technology. Using a randomised controlled experiment among third-year students in the UK, we examine whether providing targeted information on skill demand influences students' university investment and labour market strategies. We find that students generally hold accurate beliefs about employer-valued skills prior to the intervention. While the treatment has little effect on perceived skill demand, it significantly shifts how students believe employers evaluate their skills based on CVs, particularly reducing self-assessed employer evaluations. This effect is especially pronounced among female students, who revise their self-assessments downward; no such effect is observed among males. These differential perceptions translate into behavioural changes: treated females increase their participation in career events and improve academic performance, while treated males begin job searches earlier.

**Chapter 2** examines how exposure to different sibling gender (has a brother vs. a sister) influences non-cognitive outcomes for children before adolescence in Australian households. Using data from the Longitudinal Study of Australian Children (LSAC), I exploit the quasi-random assignment of sibling gender in two-child families to identify causal effects. Behavioural and emotional outcomes are measured using the Strengths and Difficulties Questionnaire (SDQ), a validated tool capturing five dimensions of non-cognitive skills: pro-social behaviour, hyperactivity/inattention, emotional symptoms, peer problems, and conduct problems. The results show that having a brother (as opposed to a sister) significantly reduces hyperactivity and inattention problems. This effect holds for both boys and girls and is most pronounced among younger children, suggesting gender-neutral mechanisms related to sibling interactions. Additionally, girls with brothers exhibit

increased pro-social behaviour, aligning with sibling differentiation theory. These findings offer insight into how family composition shapes non-cognitive skill formation and highlight implications for early childhood interventions and education policy. The beneficial effects of cross-gender sibling exposure on attention regulation can suggest that single-sex environments may constrain key developmental experiences.

**Chapter 3** explores how mental accounting and asymmetric frictions affect online gambling behaviour. Like many modern business and services, online gambling platforms in Australia make it easy to deposit and spend funds but introduce non-pecuniary friction when withdrawing funds. Using a two-stage online experiment mimicking this asymmetry, we find that when friction is introduced for withdrawals, participants tend to stick with the default setting and avoid withdrawing their money, even when this inaction increases their exposure to financial risk. This behaviour is consistent with, and possibly amplified by, mental accounting effects. The study also uncovers heterogenous effects: individuals with no prior experience in highly "gamblified" risky investment activities appear especially susceptible to mental accounting, while participants aged 50 and over are the most adversely affected by withdrawal friction.

# Table of contents

## Chapter 1 – Critical Women and Confident Men? Skills Perception and Investment at University

## Chapter 2 – The Effect of Sibling Gender on Non-Cognitive Outcomes: Evidence from Australian Children

## Chapter 3 — Trapped in the Cycle: How Mental Accounting and Asymmetric Friction Affect Online Gambling Behaviours

# Chapter 1 — Critical Women and Confident Men? Skills Perception and Investment at University

Joint work with Adeline Delavande, University of Technology Sydney; Emilia Del Bono, University of Essex; Angus Holford, University of Essex

## 1.1 Introduction

The widespread adoption of automation and digitalisation, driven by advancements in electronics and the Internet, has profoundly transformed the global economy. Computers and machines increasingly outperform humans in cognitively demanding tasks, leading to a decline in manual labour. Employment growth in cognitively intensive occupations, such as those in science, technology, engineering, and mathematics (STEM), peaked in the late 20th century but began to decline in the early 2000s (Beaudry et al., 2016). Similarly, Castex & Kogan Dechter (2014) document a reduction in the returns to cognitive test scores. Given the trajectory of technological advancements, this trend is expected to continue. However, occupations that primarily rely on social skills—such as teaching, caregiving, and management—are not as easily automated. Indeed, there is growing evidence of increasing demand for, and higher returns to, non-cognitive skills in the labour market (Autor, 2015; Deming, 2017; Edin et al., 2022). As the workforce shifts towards prioritising non-cognitive skills, employers are increasingly seeking graduates who demonstrate employability skills beyond academic performance[1].

Despite the rising importance of non-cognitive skills, higher education institutions have yet to fully integrate them into curricula. In developed countries such as the UK, the USA, and Australia, industry stakeholders frequently criticise graduates for deficiencies in key soft skills for employability (Casner-Lotto & Barrington, 2006; Jackson & Chapman, 2012). A persistent gap remains between the skills employers seek and those that graduates possess. However, it remains unclear whether students recognise this gap. The value of providing targeted information about non-cognitive skills to students is also unknown. Would it make a difference to their investment in skill development, signalling strategies, and labour market outcomes? Meanwhile, psychological research suggests that men and women exhibit distinct cognitive and behavioural traits (Croson & Gneezy, 2009; Weisberg et al., 2011), which may influence their reactions to performance-related information. For instance, Coffman et al. (2021) find that women revise their beliefs and choices more negatively than men following adverse feedback, even when controlling for prior performance and decisions. Nevertheless, gender differences in responses to information interventions remain underexplored.

This study investigates these questions through a stratified randomised information intervention that provides third-year university students in the UK with information about employability skills in demand and strategies to develop or signal them effectively. The intervention, conducted between January and February 2018, consisted of a short video outlining general recruitment requirements from a typical employer, followed by a list of five non-cognitive skills identified as "essential" or "in shortage" by the Association of Graduate Recruiters in 2016, along with their definitions. The intervention also presents statistics on the importance of these skills and information on relevant skill development events.

Using a rich dataset comprising longitudinal survey responses and administrative records, we first assess the intervention's effect on students' perceptions of five

[1] Employability skills are first defined by Goleman (1998) as "prime qualities that make and keep as employable", which refer to a set of non-cognitive soft skills that demonstrate vita personal attributes.

employability skills, including their awareness of in-demand skills, self-rating for these skills, and how they believe employers would evaluate their skills based on their CVs.

We find that while the intervention did not significantly affect students' awareness of in-demand skills, it negatively impacted their self-assessment and their perception of how employers would evaluate their skills. This effect was driven primarily by female students, who lowered their self-rating of skills despite receiving no feedback on their actual abilities or performance, whereas no significant change was observed among male students.

These changes in self-perception translated into different behavioural responses by gender. Among female students, the intervention led to greater investment in both academic and non-academic skill development. Treated females were more likely to enrol in the Big Employability Award, a university-verified program designed to enhance employability skills, and were also more likely to graduate with a First-Class degree. Additionally, they showed higher attendance at career workshops, greater participation in work experience aligned with their field of study, and improved final-year academic performance.

For male students, the intervention primarily influenced job search behaviours rather than skill investment. Treated males were significantly more likely to begin job searches earlier, as indicated by a higher likelihood of applying for graduate jobs and securing offers within two months of the intervention. However, this early job search had long-term negative consequences, as treated males were less likely to secure stable, long-term employment two years after graduation. A possible explanation is that male students interpreted the information as a signal of increased labour market competition. Since they did not adjust their self-assessment downward, they may have believed they already possessed the necessary skills and, as a result, prioritized immediate job entry over further skill investment. This premature focus on securing employment may have left them underprepared for long-term career stability, ultimately leading to less favourable job attributes over time.

Our findings contribute to the growing literature on gender differences in non-cognitive skills as explanations for labour market disparities. Labour economists increasingly recognise the role of psychological traits, preferences, and personality in shaping economic outcomes (Heckman & Kautz, 2012). Mueller and Plug (2004) first suggested that attributes such as self-confidence function as human capital variables, influencing productivity and wages. Several studies estimate that psychological traits explain between 2.5% and 27.6% of the gender wage gap (Machado, 2012; Manning & Swaffield, 2008; Reuben et al., 2015). More recent research attributes the wage gap to differences in negotiation and risk-taking behaviour (Biasi & Sarsons, 2021; Card et al., 2016; Dittrich et al., 2014; Exley et al., 2020). Our study differs by examining gender differences in information interpretation and subsequent investment decisions that shape labour market outcomes.

Additionally, we contribute to the literature on human capital investment by analysing non-cognitive skill development among undergraduates. While efforts to foster non-cognitive skills during adolescence and early adulthood have shown promising results (Bettinger et al., 2012; Carrell & Sacerdote, n.d.; Cook et al., 2014), few studies examine skill investment in adulthood. Delavande et al., (2022) explore university students' investment decisions in human capital, considering time allocation and expected returns.

Furthermore, we extend the literature on information interventions in education, which has gained traction in economics. Most studies focus on education decisions, such as Jensen (2010), who finds that providing information on higher returns to education increases schooling years, and (Bleemer & Zafar, 2018) who show that exposure to objective information on college returns influences attendance intentions. Our study investigates how providing employability information influences students' skill investment decisions and long-term labour market outcomes.

Finally, we contribute to research on extra-curricular engagement and labour market success. Previous studies highlight the role of extra-curricular activities in improving transitions from university to employment (Tchibozo, 2007; Milner et al., 2016; Roulin & Bangerter, 2013; Nghia, 2017) and document wage premiums for students engaged in sports or internships (Lechner & and Downward, 2017; Persico et al., 2004; Saniter & Siedler, 2014). Instead of evaluating wage or employment effects, we examine how employability skills are developed through extra-curricular participation.

This chapter is structured as follows. Section 2 describes our data and analytical sample. Section 3 outlines the intervention and hypotheses. Section 4 presents descriptive evidence and measurement strategies. Section 5 discusses empirical strategy and results, followed by robustness checks in Section 6. Section 7 concludes.

# 1.2 Data

## 1.2.1 BOOST2018

The BOOST2018 is a longitudinal survey comprising an entire cohort of undergraduate students who enrolled at a UK university between October 2015 (academic year 2015/16) and June 2018 (academic year 2017/18).[2] The institutional characteristics of this university are representative of other higher education institutions in the UK. Undergraduate degree programmes typically span three years, and students must successfully pass each academic year to progress to the next. Academic performance in the second and third years is used to calculate the degree mark, which determines the degree class awarded upon graduation.

The BOOST2018 sampling frame includes all students who enrolled in the first year of an undergraduate degree in the academic year 2015/16. The sample consists of 2,621 students, including Home (UK resident), EU, and international students. The study was widely advertised across the university campus, and students who registered to participate were incentivised with a £5 sign-up reward. By the end of the Autumn term of the 2017/18 academic year, a total of 2,005 students had enrolled in BOOST2018, representing 76.5% of the entire cohort.

BOOST2018 comprises 14 waves of online and laboratory-based surveys, which are linked to university administrative records. In each academic year, participants were invited to complete three online surveys, conducted during the Autumn term (November), Spring term (March), and revision period (May), as well as one laboratory session at the Social Science Experimental Laboratory in January. The online surveys in November and March were long-format surveys (approximately 60 minutes), while the May survey was shorter (approximately 25 minutes). These surveys collected information on students' academic engagement, including attendance, study time, study habits, and participation in extracurricular activities. They also captured students' subjective beliefs about their abilities, expectations regarding academic outcomes, and career aspirations.

Additionally, university administrative records provide detailed demographic and academic data, including age, gender, ethnicity, parental education level, parental

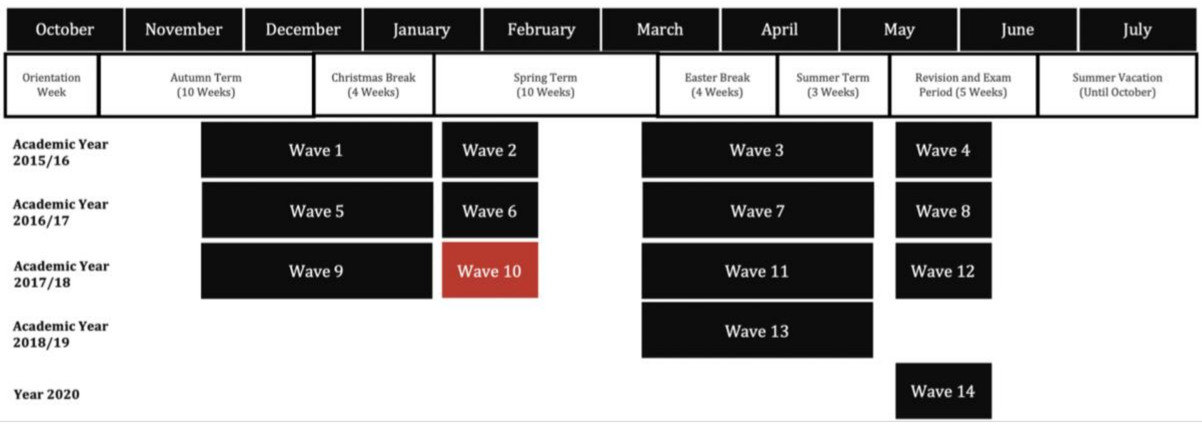
[2] About 20% of students went abroad/placement in the third year (academic year 2017/18) and completed their degree in June 2019.

occupation, and university participation rates within students' home regions. These records also contain students' academic performance data, their participation in career development events organised by the University Employability and Careers Service, and their attendance records, which were tracked using a swipe-card electronic system.

To encourage survey participation, monetary incentives were provided. Payments ranged from £8 to £20 for online surveys and averaged £30 for laboratory sessions. Over the three-year study period, BOOST2018 implemented three randomised interventions during the laboratory sessions. This paper focuses on the information intervention conducted in January 2018 (Wave 10), which coincided with the third-year laboratory session.

Figure 1.1 provides an overview of the data collection timeline. Waves 1 to 12 correspond to data collection between the 2015/16 and 2017/18 academic years. Wave 13 was collected in the 2018/19 academic year and included only students who remained enrolled as undergraduates due to study-abroad placements, industrial placements, or repeated years. The final survey, Wave 14, was conducted in May 2020, capturing realised labour market outcomes approximately two years post-graduation.

*Figure 1.1 Timeline of BOOST2018*




## 1.2.2 Analytical Samples

The target population consists of third-year university students who successfully progressed from Years 1 and 2 without delays. Column (1) of Table 1.1 presents the baseline characteristics of the entire cohort upon enrolment in the first year, showing a balanced gender distribution. Column (2) displays the characteristics of students who consented to participate in BOOST2018. Over 90% of participants were non-mature students (aged 21 or below at entry), and 85% were Home (UK resident) or EU students. BOOST participants are broadly similar to the full cohort, with the main observable difference being a higher proportion of females. This difference does not affect internal validity, as randomisation occurs within the respondent sample.

*Table 1.1 Summary statistics of participants*

| **Variable** | **(1)** Entire Cohort | **(2)** BOOST Survey | **(3)** BOOST in Year 3 | **(4)** Wave 9 | **(5)** Wave 910 | **(6)** Wave 911 | **(7)** Wave 1014 |
|---|---|---|---|---|---|---|---|
| *Basic Characteristics* | | | | | | | |
| Female | 0.542 | 0.624 | 0.635 | 0.549 | 0.516 | 0.491 | 0.464 |
| Non-Mature | 0.905 | 0.927 | 0.942 | 0.920 | 0.915 | 0.911 | 0.904 |
| British/EU | 0.912 | 0.927 | 0.934 | 0.925 | 0.940 | 0.940 | 0.942 |
| White | 0.826 | 0.830 | 0.836 | 0.837 | 0.848 | 0.855 | 0.859 |
| High-SES | 0.382 | 0.388 | 0.388 | 0.388 | 0.379 | 0.377 | 0.370 |
| Low-SES | 0.236 | 0.233 | 0.224 | 0.235 | 0.238 | 0.234 | 0.237 |
| Overall Mark in Year 1 | 59.19 | 59.12 | 60.22 | 61.45 | 61.00 | 61.07 | 61.17 |
| *Tariff Quintile* | | | | | | | |
| First (Lowest) | 0.146 | 0.149 | 0.148 | 0.149 | 0.154 | 0.148 | 0.150 |
| Second | 0.161 | 0.166 | 0.172 | 0.175 | 0.166 | 0.174 | 0.177 |
| Third | 0.166 | 0.188 | 0.178 | 0.174 | 0.181 | 0.173 | 0.173 |
| Fourth | 0.180 | 0.176 | 0.168 | 0.169 | 0.170 | 0.173 | 0.170 |
| Fifth (Highest) | 0.145 | 0.134 | 0.133 | 0.134 | 0.135 | 0.132 | 0.132 |
| *Department* | | | | | | | |
| Social Science | 0.401 | 0.401 | 0.408 | 0.381 | 0.386 | 0.384 | 0.357 |
| Science | 0.379 | 0.375 | 0.368 | 0.342 | 0.354 | 0.364 | 0.352 |
| Humanities | 0.220 | 0.224 | 0.224 | 0.277 | 0.260 | 0.252 | 0.291 |
| **Observations** | 2621 | 2005 | 1028 | 770 | 677 | 601 | 437 |

*Notes:* Column (1) shows the characteristics of students in the target population, i.e. those who enrolled in university. Column (2) includes students who signed up to participate in the survey, and Column (3) includes those who responded to the survey in Year 3. Columns (4)–(7) restrict the sample based on who attended lab sessions and responded to follow-up surveys. SES missing cases not shown.

Column (3) presents the characteristics of BOOST2018-registered students who successfully progressed to their third year at the time of the intervention. The main estimation sample consists of students who attended the Wave 10 laboratory session, representing 75% of the eligible third-year population, as shown in column (4). A comparison between column (3) and columns (4) to (7) indicates that the analytical sample is highly representative of the broader population.

The different analytical samples used in this study are presented in Columns (4) to (7). For instance, the Wave 910 showed in column (5) sample includes students who attended the Wave 10 lab session and responded to the Wave 9 survey. Similarly, sample of column (6) comprises students who participated in Waves 9, 10, and 11, which is the main analytical sample used for the analysis. It contains baseline data for key outcome variables, allowing for a more robust evaluation of treatment effects.

# 1.3 The Information Intervention

## 1.3.1 Overview

The information intervention took place in Year 3 during the laboratory session in Wave 10, which was also the beginning of the Spring term as shown in Figure 1.1 Timeline of BOOST2018. It was designed to assess whether providing students with knowledge about essential employability skills enhances their preparation for the job market. The intervention was implemented through a stratified randomised controlled trial (RCT), cross-randomised with two pre-existing interventions conducted in Years 1 and 2. After first stratifying by gender, mature status (age above 21), parental socio-economic status, department, and tariff quintile[3], students within each cell were then randomly assigned into a control group and a treatment group with equal probability. Among those 1,497 students, 750 of them were assigned to the control group and 747 of them were assigned to the treatment group.

A total of 1,496 email invitations were sent out, inviting students to register for a laboratory session.[4] Participants had the flexibility to choose from multiple sessions, held at different times and on different days over a three-week period. The response rate was approximately 50%, with 395 students in the control group and 375 students in the treatment group. A summary of the baseline characteristics and balance tests between the control and treatment groups are provided in Table 1.2 below.

The laboratory was equipped with individual partitioned booths, each containing a computer and noise-cancelling headphones to ensure a controlled and distraction-free environment. The lab session was divided into two parts. In the first part of the lab session, students received the information intervention with incentivised tasks or the alternative contents. The tasks differed based on the condition, but the financial payoff was comparable with an average of £31.34 for the control group and £29.72 for the treatment

[3] The tariff points are available through the linkage with the university administrative data and come from the University and Colleges Admission Service (UCAS). The UCAS Tariff points are a way of comparing the value of all post-16 qualifications in the UK, as students can access university by gaining academic qualifications, vocational qualifications or a mixture of the two. The total score is obtained by assigning a numerical value to each grade and qualification and summing these up. The higher the grade the student achieved per each qualification, the higher the number of points awarded.

group. In the second part, all students (both treatment and control groups) completed an identical survey, which assessed students' perceptions, investment decisions, and behavioural responses following the intervention.

*Table 1.2 Baseline balancing checks*

| Variable | (1) (N = 394) | (2) Treatment (N = 376) | (3) p-value |
|---|---|---|---|
| *Basic Characteristics* | | | |
| Female | 0.563 | 0.576 | 0.734 |
| Non-Mature | 0.944 | 0.939 | 0.753 |
| British/EU | 0.840 | 0.870 | 0.244 |
| High-SES | 0.391 | 0.402 | 0.761 |
| Low-SES | 0.239 | 0.229 | 0.747 |
| *Tariff Quintile* | | | |
| First (Lowest) | 0.154 | 0.143 | 0.663 |
| Second | 0.192 | 0.183 | 0.739 |
| Third | 0.129 | 0.135 | 0.801 |
| Fourth | 0.170 | 0.175 | 0.842 |
| Fifth (Highest) | 0.165 | 0.138 | 0.302 |
| *Department* | | | |
| Social Science | 0.380 | 0.382 | 0.950 |
| Science | 0.344 | 0.347 | 0.926 |
| Humanities | 0.276 | 0.271 | 0.867 |
| *Self-rating on skills* | | | |
| Negotiate and Influencing | 68.234 | 69.288 | 0.436 |
| Commercial Awareness | 63.883 | 64.028 | 0.920 |
| Problem Solving | 75.060 | 75.500 | 0.727 |
| Interpersonal Skill | 77.057 | 76.975 | 0.950 |
| Teamwork | 79.721 | 80.801 | 0.415 |
| Average | 72.791 | 73.156 | 0.727 |
| *Proportion of employers demand skills* | | | |
| Negotiate and Influencing | 73.467 | 73.632 | 0.901 |
| Commercial Awareness | 73.735 | 72.994 | 0.587 |
| Problem Solving | 80.128 | 80.923 | 0.525 |
| Interpersonal Skill | 78.595 | 79.905 | 0.333 |
| Teamwork | 82.821 | 83.172 | 0.781 |
| Average | 77.749 | 78.125 | 0.728 |
| *Others* | | | |
| Had experiences related to study/desired career | 0.293 | 0.313 | 0.583 |
| Had other employment/non-academic experiences | 0.746 | 0.736 | 0.762 |
| The BigE Award enrollment[a] | 0.368 | 0.368 | 0.988 |
| Hours in experiences related to study/desired career | 3.198 | 3.376 | 0.836 |
| Hours in other employment/non-academic experiences | 11.073 | 8.511 | 0.040* |
| Hours in the BigE Award[a] | 3.777 | 2.355 | 0.104 |
| Number of Career Events attended[b] | 0.883 | 0.888 | 0.963 |
| Study time (Hours per week) | 15.042 | 14.290 | 0.368 |
| Attendance (%)[b] | 0.655 | 0.640 | 0.325 |
| Attendance (Hours per week)[b] | 5.129 | 4.986 | 0.368 |
| Mark (Year 1) | 61.638 | 61.195 | 0.533 |

*Notes:* Baseline statistics reported here reflect values measured in the last wave prior to treatment, except for those marked with symbol [a]. Symbol [b] indicates that the statistics were computed using administrative data from the third year, prior to the lab session (treatment).

### 1.3.2 The Treatment

The treatment intervention was conducted at the beginning of the laboratory session in Wave 10 and comprised five distinct components. Students in the control group were instead given an incentivised non-verbal reasoning test to assess their problem-solving ability, followed by an incentivised writing task that required them to describe a learning experience.

#### *1.3.2.1 A Video of a Typical Recruiter*

The information intervention began with a 90-second video featuring a graduate recruiter discussing general recruitment requirements. The recruiter outlined:

- The minimum degree classification required (a 2:1 degree, i.e. above 60% in final GPA),
- Skills valued beyond academic performance, and
- Examples of skill development opportunities.

The recruiter shared her own experiences working in factories and bars to illustrate commercial awareness, one of the employability skills. The full transcript of the treatment video is provided in Appendix 1. Students were required to watch the video for at least 90 seconds before proceeding, with the option to rewatch the content if they wished.

#### *1.3.2.2 Quiz on Skills Definition*

Following the video, students completed a short quiz in which they were asked to match five employability skills to their correct definitions.[5] The five listed skills were:

- Teamwork
- Interpersonal skills
- Problem-solving
- Commercial awareness
- Negotiating and influencing

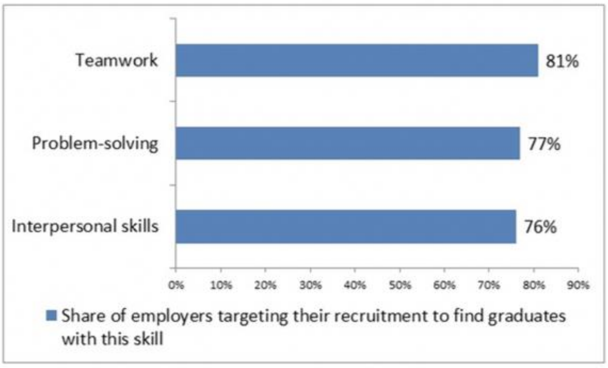
[5] See A 1.1 from Appendix 1 for detailed contents of the information treatment.

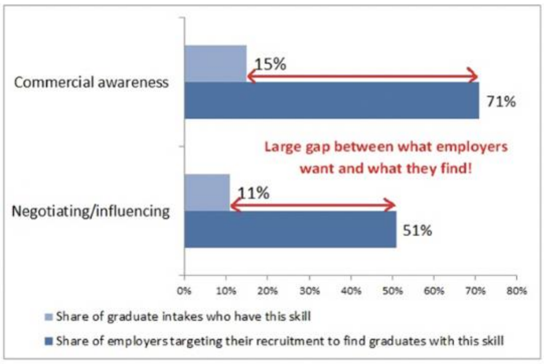
Students who provided incorrect answers were notified of their mistake and shown the correct responses.[6]

### *1.3.2.3 Information about Essential Skills and Skills in Shortage*

After completing the quiz, students were presented with data on essential skills and skills that are in shortage, based on a survey conducted by the Association of Graduate Recruiters (2016). Essential skills were defined as the top three attributes that employers prioritise when hiring graduates, while skills in shortage referred to attributes that employers struggle to find among graduates.

*Figure 1.2 Screenshots of information treatment for "Essential Skills" and "Skills in Shortage"*




These information were accompanied by graphs illustrating the proportion of employers who tailor recruitment efforts to identify graduates with these skills. Below these graphs, explanatory text was provided to ensure students correctly interpreted the data. For example, one of the statements read:

*"81% of employers prioritise hiring graduates who demonstrate strong teamwork skills."*

For skills in shortage, a second graph was provided to highlight the discrepancy between employer demand and graduate supply for these skills. See Figure 1.2 as example of the illustration. The importance of not only possessing these skills but also demonstrating them effectively was emphasised.

---

[6] 70% of students in the treatment group had got at least four correct answers out of five while the most common mistake happened with interpersonal skills.

#### *1.3.2.4 Information about Career Events and Big Employability Award*

The next component of the intervention provided information about upcoming courses and career events organised by the university's Employability Service. Students were given the option to opt-in for an email reminder to register for one of these events. These career-focused sessions were designed to help students develop and effectively demonstrate their employability skills. Additionally, the intervention introduced the Big Employability Award (BigE Award), a university-endorsed certification that formally recognises students' participation in a range of extracurricular activities. The BigE Award documents students' involvement in Student Union roles, university-run employment, volunteering, leadership positions among other activities, which would be included alongside students' degree transcripts. This certification serves two key functions: (i) Identify and validate the skills students have developed throughout their university experience, and (ii) Help students effectively signal their skills to employers by providing concrete examples that can be referenced in job applications and interviews.

#### *1.3.2.5 Incentivised mock interview essay*

The treatment finished with an incentivised task designed to help students practice signalling skills in a mock job interview setting. Participants were given eight minutes to compose a written response to two interview-style questions. It required them to demonstrate the understanding of the challenges faced by organisations within the industry they wish to work in as well as to describe a situation where they successfully persuaded someone to adopt their point of view. Students received £2 for every 200 characters (approximately two lines of meaningful text), with earnings capped at a maximum of £10. Following the writing task, participants were explicitly informed of the two key employability skills that employers assess through such interview questions.

### 1.3.3 Hypotheses of the Treatment Effect

The provision of information on employability is expected to improve students' understanding of labour market demands and encourage them to invest in the development or signalling of these skills. Specifically, we hypothesise the following four directions of effects of the intervention.

***Hypothesis 1***: *Updating of Skills Perceptions – Students will revise their beliefs about the importance of specific employability skills based on the information provided, adjusting their self-assessment accordingly.*

***Hypothesis 2***: *Increased Investment in Employability Skills – Students will engage more in extracurricular activities such as university clubs and societies, volunteering, internships, or employment, as these activities were explicitly mentioned in the intervention as key avenues for developing employability skills.*

***Hypothesis 3***: *Greater Academic Investment and Achievement – Since the intervention emphasised the importance of obtaining a good degree classification, we expect an increase in academic effort and improved academic outcomes.*

***Hypothesis 4***: *Enhanced Job Market Readiness – The intervention is expected to improve students' transition from university to the workforce by reducing job search duration, increasing post-graduation earnings, and improving non-pecuniary job attributes.*

To evaluate these hypotheses, we employ a combination of self-reported survey data and administrative records in our empirical analysis.

## 1.4 Measuring Skills Perceptions and Employability Skills

This section outlines how students' skills perception, investment in employability skills, and academic investments are measured. It also details labour market outcomes collected two years after graduation. Table 1.3 presents summary statistics for all key variables described in the following subsections.

### 1.4.1 Skills Perception

We assess three dimensions of students' skill perceptions: (i) perceived employer demand for skills, (ii) self-assessment of skills, and (iii) self-assessment on skills signalling to employers. Each dimension includes nine key non-cognitive skills commonly cited by graduate employers in recruitment decisions, which are 'Managing up', 'Dealing with conflict', 'Negotiating and Influencing', 'Commercial awareness', 'Business communication', 'Self-awareness', 'Problem-solving', 'Interpersonal skills' and 'Teamwork'. Students were asked to rate these skills on a 0–100 scale in response to the following questions:

*"Rate their importance to employers. What proportion of graduate employers do you think tailor their recruitment process specifically to find graduates who already have each of these skills?*

*Rate yourself. For each of the 9 skills listed below, please rate yourself on the scale from 0 (you have no skill at all in this field) to 100 (your skill is perfect).*
*Rate how well you signal these skills. For each of the 9 skills listed below, how well do you think an employer looking at your CV would rate you on the scale from 0 (you have no skill at all in this field) to 100 (your skill is perfect)?"*

To assist students unfamiliar with these terms, clickable definitions were provided. These questions were asked in Waves 9, 10, 11, and 12, with self-rating of signalling skills introduced in Wave 10. Our analysis focuses on five skills explicitly mentioned in the information intervention.

#### *1.4.1.1 Baseline Analysis of Skills Perceptions*

To assess whether students had accurate perceptions of employer demand for these skills at baseline, we compare their ratings with actual employer survey data. Figure 1.3 and Figure 1.4 display density plots of students' skill ratings by gender and treatment status, with a vertical line indicating the true employer demand. The distributions for males and females, as well as control and treatment groups, mostly overlap for all five skills, indicating no baseline differences in perceived employer demand. Despite substantial variance, students' average ratings were close to actual employer demand, except for 'Negotiating and Influencing', where students overestimated the demand by approximately 20 percentage points.

Panel A of Table 1.3 reports baseline self-assessment scores, suggesting pre-existing gender differences. Female students rated themselves significantly lower than males in 'Negotiating and Influencing' and 'Commercial Awareness' by 3 and 4 points, respectively. Conversely, females rated themselves higher in 'Interpersonal Skills' and 'Teamwork' by 4 and 4.5 points, respectively (significant at 1% level). No significant gender difference was found in self-rating of 'Problem-Solving'. Female students also rated themselves 5 points lower in signalling 'Commercial Awareness' (significant at 5% level).

*Table 1.3 Descriptive statistics of outcome variables*

| | (1) All | (2) Male | (3) Female | (4) p-value |
|---|---|---|---|---|
| *Panel A: Elicited Subjective Beliefs* | | | | |
| *Proportion of employers demand skills* | | | | |
| Negotiate and Influencing | 73.255 | 72.492 | 73.793 | 0.370 |
| | (17.20) | (18.11) | (16.54) | |
| Commercial Awareness | 73.368 | 72.774 | 73.787 | 0.499 |
| | (17.75) | (18.73) | (17.05) | |
| Problem Solving | 80.865 | 79.919 | 81.531 | 0.251 |
| | (16.22) | (18.48) | (14.40) | |
| Interpersonal Skill | 79.573 | 78.504 | 80.327 | 0.213 |
| | (17.08) | (18.88) | (15.68) | |
| Teamwork | 83.195 | 82.194 | 83.901 | 0.227 |
| | (16.39) | (18.41) | (14.79) | |
| Average | 78.051 | 77.177 | 78.668 | 0.216 |
| | (13.98) | (15.74) | (12.58) | |
| *Self-rating on skills* | | | | |
| Negotiate and Influencing | 68.660 | 70.569 | 67.315 | 0.022** |
| | (17.35) | (16.64) | (17.73) | |
| Commercial Awareness | 63.773 | 66.267 | 62.023 | 0.005*** |
| | (18.70) | (17.43) | (19.38) | |
| Problem Solving | 75.288 | 75.597 | 75.071 | 0.705 |
| | (16.60) | (17.10) | (16.25) | |
| Interpersonal Skill | 77.007 | 74.718 | 78.619 | 0.007*** |
| | (17.08) | (18.57) | (15.78) | |
| Teamwork | 80.123 | 77.470 | 82.000 | 0.002*** |
| | (17.20) | (19.46) | (15.15) | |
| Average | 72.882 | 72.707 | 73.006 | 0.794 |
| | (13.46) | (14.54) | (12.66) | |
| *Self-rating on signalling skills*[a] | | | | |
| Negotiate and Influencing | 66.685 | 67.964 | 65.778 | 0.196 |
| | (20.54) | (19.98) | (20.91) | |
| Commercial Awareness | 62.530 | 65.373 | 60.513 | 0.014** |
| | (24.05) | (23.09) | (24.54) | |
| Problem Solving | 76.398 | 77.458 | 75.647 | 0.201 |
| | (17.33) | (16.50) | (17.88) | |
| Interpersonal Skill | 77.567 | 76.594 | 78.256 | 0.247 |

| | | | | |
|---|---|---|---|---|
| | (17.61) | (16.62) | (18.27) | |
| Teamwork | 85.348 | 84.410 | 86.014 | 0.177 |
| | (14.48) | (13.95) | (14.83) | |
| Average | 73.706 | 74.360 | 73.242 | 0.321 |
| | (13.76) | (13.12) | (14.20) | |
| *Panel B: Investment in Employability Skills* | | | | |
| Had experiences related to study/desired career | 0.307 | 0.274 | 0.330 | 0.145 |
| | (0.461) | (0.447) | (0.471) | |
| Had other employment/non-academic experiences | 0.748 | 0.673 | 0.801 | 0.001*** |
| | (0.434) | (0.470) | (0.400) | |
| The BigE Award enrollment | 0.377 | 0.230 | 0.480 | 0.000*** |
| | (0.485) | (0.422) | (0.500) | |
| Hours in study/desired career related exp. | 3.399 | 3.754 | 3.149 | 0.575 |
| | (11.70) | (15.66) | (7.796) | |
| Hours in other employment exp. | 10.062 | 9.363 | 10.554 | 0.450 |
| | (17.29) | (22.59) | (12.26) | |
| Hours in the BigE Award | 3.292 | 1.960 | 4.234 | 0.016** |
| | (12.18) | (9.393) | (13.75) | |
| Career events attended[bc] | 0.917 | 0.839 | 0.972 | 0.283 |
| | (1.535) | (1.364) | (1.645) | |
| *Panel C: Academic Inputs and Outcomes* | | | | |
| Mark (Year 3)[c] | 64.727 | 64.168 | 65.156 | 0.269 |
| | (10.57) | (11.44) | (9.850) | |
| Degree mark[c] | 63.790 | 63.655 | 63.892 | 0.711 |
| | (8.425) | (9.246) | (7.754) | |
| First class[c] | 0.258 | 0.280 | 0.241 | 0.236 |
| | (0.438) | (0.450) | (0.428) | |
| Upper second class or above[c] | 0.735 | 0.687 | 0.776 | 0.000*** |
| | (0.441) | (0.464) | (0.417) | |
| Study time (hrs/week) | 15.115 | 13.589 | 16.195 | 0.004*** |
| | (11.05) | (10.21) | (11.50) | |
| Attendance (%)[bc] | 0.647 | 0.601 | 0.683 | 0.000*** |
| | (0.211) | (0.218) | (0.199) | |
| Attendance (hrs/week)[bc] | 5.058 | 4.942 | 5.145 | 0.209 |
| | (2.196) | (2.282) | (2.127) | |
| *Panel D: Job Search and Labour Market Outcomes[d]* | | | | |

| | | | | |
|---|---|---|---|---|
| Job application submitted | 0.327 | 0.321 | 0.330 | 0.813 |
| | (0.469) | (0.468) | (0.471) | |
| Job offer secured | 0.092 | 0.092 | 0.091 | 0.960 |
| | (0.289) | (0.290) | (0.288) | |
| Employed within 3 months | 0.668 | 0.651 | 0.680 | 0.561 |
| | (0.471) | (0.478) | (0.468) | |
| Employed within 6 months | 0.763 | 0.747 | 0.773 | 0.559 |
| | (0.426) | (0.436) | (0.420) | |
| Postgraduate education | 0.402 | 0.349 | 0.436 | 0.096* |
| | (0.491) | (0.478) | (0.497) | |
| Fixed/Long-term contract | 0.844 | 0.868 | 0.828 | 0.311 |
| | (0.364) | (0.340) | (0.379) | |
| Short-term/Temp. contract | 0.133 | 0.110 | 0.148 | 0.308 |
| | (0.340) | (0.314) | (0.356) | |
| Annual gross earnings | 21796.94 | 23689.53 | 20491.05 | 0.037** |
| | (13288.4) | (14904.8) | (11912.0) | |

*Notes:* Means and standard deviations (in parentheses) are from Wave 9 for the for the analytical sample Wave91011 unless stated otherwise. [a]Data collected from Wave 10 after the information intervention. [b]Administrative data in Year 3 prior to the lab session (treatment). [c]Wave 10 participants only. Derived from admin records; degree mark = 40% Year 2 + 60% Year 3. First class = mark ≥ 70; upper second class = mark ≥ 60. The maximum value of study hours per week is 80, capped at 99% percentile. [d]Panel E reports post-treatment data from Wave 11 (first two variables) and Wave 14. Outliers in annual gross earnings are replaced to 1% and 99% percentiles respectively.

*Figure 1.3 Density of skills perception on employers' demand at baseline - by gender*

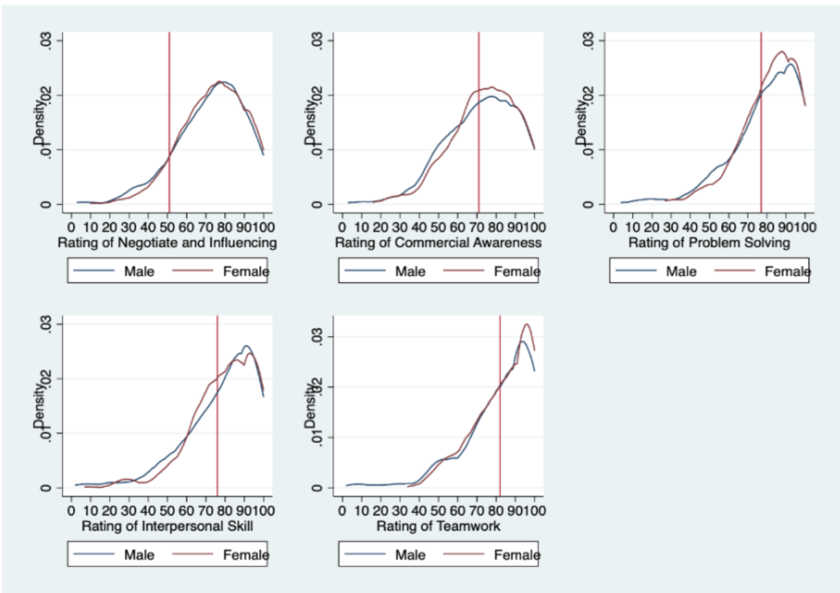
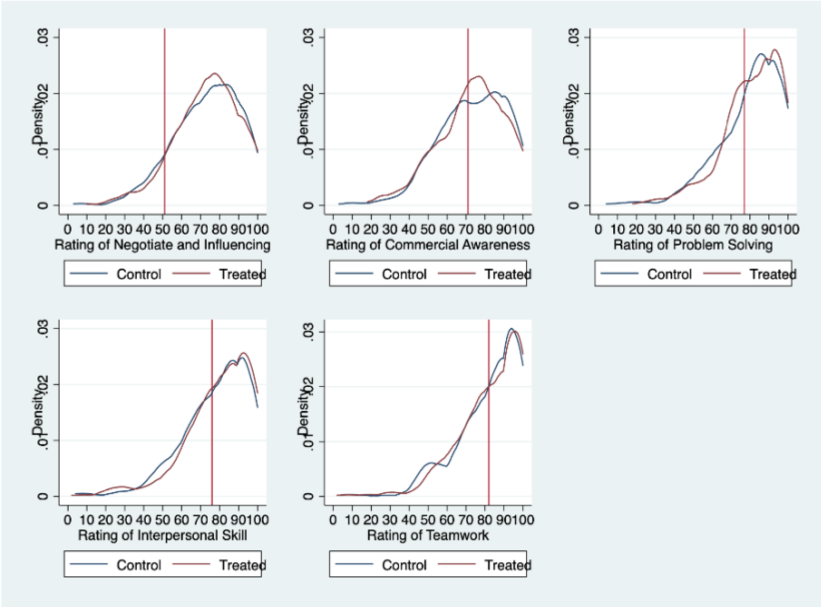


*Figure 1.4 Density of skills perception on employers' demand at baseline - by treatment*




## 1.4.2 Investment in Employability Skills

Investment in employability skills was assessed through students' participation in extracurricular activities, which were explicitly emphasised in the intervention as key avenues for skill acquisition and professional development. Data on these activities were derived from two primary sources: self-reported survey responses and administrative records of university-facilitated events.

The survey instruments captured students' engagement in a range of extracurricular activities, including participation in university-affiliated sports clubs and societies, involvement in the Student Union, voluntary work within or outside the university, internships, and paid employment. They were administered in Waves 9 and 11, referencing students' participation during the corresponding academic term. Additionally, students were asked about their participation in the Big Employability Award (BigE Award), a university-led certification programme introduced during the intervention. This initiative formally recognises students' co-curricular engagement—such as employment and volunteering—by issuing a verified certificate that can be appended to their academic transcript. This investment was also measured through survey responses in Waves 9 and 11 regarding their enrolment in the programme.

For activities such as volunteering, internships, and paid employment, students were asked to indicate whether their experience was relevant to their field of study or intended career path. Instead of treating each activity as an independent measure of employability skill investment, we adopted the classification approach used by Delavande et al. (2020), which groups activities into two broad categories. The first category includes experiences explicitly related to students' fields of study or desired careers, while the second encompasses employment and non-academic experiences unrelated to their academic specialisation.

As a result, we constructed an indicator variable, 'Study/desired career-related experience', which takes a value of 1 if a student engaged in paid employment, an internship, or volunteering that was reported as relevant to their field of study or career aspirations, and 0 otherwise. A similar indicator variable, 'Other employment/non-academic experience', was defined for students who participated in sporting competitions, university-affiliated extracurricular activities, or employment, internships, or volunteering unrelated to their field of study. Additionally, we measured the time commitment to these activities by recording the number of hours spent per week in each category.

On the other hand, the university's administrative records document all career-related events available to students, including details on event descriptions, student registrations, attendance, and event dates. These events and workshops organised by the University Employability and Careers Service serve as structured opportunities for students to develop and enhance their employability skills similar to the BigE Award.

Our primary variable of interest is the number of career events attended, irrespective of event type. To quantify students' engagement, we aggregated the total number of events attended before and after the treatment, aligning the timing of event participation with the first week of the Wave 10 lab sessions. To ensure consistency with other baseline variables, only career events attended during students' third year were included in the baseline measure.

#### *1.4.2.1 Baseline Analysis of Participation in Extracurricular Activities*

Panel B of Table 1.3 indicates that, prior to the intervention, approximately 30% of the analytical sample[7] had engaged in experiences directly related to their field of study or

[8] The analytical sample includes 677 students who participated in both Waves 9 and 10.

desired career. As expected, participation in non-academic employment or extracurricular activities was more prevalent, with 75% of students reporting involvement.

Gender differences were evident, with 80% of female students engaging in non-academic employment or extracurricular activities compared to 67% of males, a statistically significant difference at the 5% level. On average, students dedicated approximately 3.4 hours per week to ‘Study/desired career-related experiences’ and 10 hours per week to ‘Other employment/non-academic experiences’. However, there was no statistically significant gender difference in the mean number of hours worked.

Regarding the BigE Award, nearly half of the female students (48%) had already enrolled, whereas less than a quarter of male students (23%) had done so. Moreover, among students who had registered, females devoted significantly more time to the programme, averaging four hours per week—more than double the time commitment of their male counterparts. These gender differences were statistically significant.

Panel B of Table 1.3 also indicates that, students had, on average, booked just under one career event in their third year before the intervention, with no statistically significant gender difference found.

### 1.4.3 Academic Investment

Our intervention also emphasised the importance of degree classification, particularly the final degree mark and degree class, as key determinants of students’ academic success and employability prospects. This emphasis may have implicitly influenced students' academic investments. To evaluate these effects, we examine students' academic outcomes, including their third-year marks, final degree mark, and degree classification, as indicators of academic engagement and performance.

The university follows a modular structure, a system commonly adopted by UK higher education institutions. Students enrol in four to eight modules per academic year, with module assessments comprising a combination of coursework and final examinations. Marks are awarded on a 0 to 100 scale, and students must achieve a minimum overall mark of 40 to pass each module. The final degree mark is calculated using a weighted average, with 40% derived from second-year marks and 60% from third-year marks. Degree classifications follow standard UK grading criteria: students obtaining a final mark of 70 or above are awarded First-Class Honours, while those scoring between 60 and 69 receive an Upper Second-Class Honours.

Administrative records contain detailed module-level grade data for all BOOST2018 participants throughout their three-year undergraduate studies. Using these records, we compute the post-intervention mark, referred to as ‘Mark (Year 3)’ in Panel C of Table 1.3, by calculating the weighted average of students' overall third-year module marks.

Additionally, we define First-Class as an indicator variable equal to 1 if a student achieved a final degree mark of 70 or above, while Upper Second-Class or Above is set to 1 if the final mark exceeds 60. The first-year overall mark is used as a baseline measure, as it contributes to the final degree calculation. As anticipated, female students exhibited significantly higher academic outcomes than their male counterparts.

### 1.4.4 Job Search Behaviours

To assess students' job search behaviours, we collected multiple measures through the Wave 11 survey, which was administered towards the end of third-year studies, approximately two to three months prior to graduation. Our analysis focuses on two key indicators derived from survey questions that provide insights into potential variations in the timing of job-seeking activities.

We construct the following two binary outcome variables: ‘Job application submitted’ and ‘Job offer secured’. ‘Job application submitted’ indicates 1 if a student reported having submitted at least one job application, while ‘Job offer secured’ indicates 1 if a student reported having received at least one confirmed job offer. As baseline data on job search behaviours were not collected in prior waves, Panel D of Table 1.3 presents the descriptive statistics for these job search variables based on Wave 11 responses.

### 1.4.5 Labour Market Outcomes

To evaluate the long-term impact of the intervention on actual labour market outcomes, we conducted follow-up interviews with BOOST participants in May 2020 (Wave 14), approximately 22 months post-graduation for those who completed their degrees on time.

[8] Due to panel attrition, the Wave 14 sample was reduced to N = 437. The survey collected retrospective monthly data on employment and education activities throughout the post-graduation period. Based on this information, we constructed three binary variables regarding the immediate employment status post-graduation. For instance, ‘Employed within 3 months of graduation’ was recorded as 1 if the student reported being employed at least once in the first three months post-graduation. The same logic applied to ‘Employed within 6 months of graduation’. ‘Postgraduate education’ equals to 1 if the graduate reported engaging in postgraduate education or training for at least one month after completing their undergraduate degree.

Additionally, we examined earnings and non-pecuniary job attributes, such as employment contract type and job stability. Participants were asked to classify their employment basis using the following multiple-choice question:

*“Which of these best describes your employment basis?*
*(1) On a permanent or open-ended contract*
*(2) On a fixed-term contract lasting 12 months or longer*
*(3) On a fixed-term contract lasting less than 12 months*
*(4) Temporarily, through an agency*
*(5) Temporarily, other than through an agency*
*(6) Other”*

Respondents who were unemployed at the time of the survey were also asked about the nature of their most recent job. To categorise employment stability, we constructed two key binary variables based on responses to the employment basis question. The variable ‘Fixed/Long-term contract’ equals 1 if the respondent selected (1) or (2). ‘Temporal/Short-term contract’ equals 1 if the respondent selected (3), (4), or (5).

Given variations in pay frequency, we also standardised annual gross earnings across different reporting periods. For self-reported annual salaries, we retained the original reported value. For monthly and weekly wages, we multiplied by 12 and 52, respectively.

[8]Around 15% of third year students who did not complete their degrees on time due to study abroad/placement or fail/restart/dropout during year 3. Those who graduated a year later were still included in the survey and their responses had been harmonised with students graduated under the standard track.

For hourly wage earners, we estimated annual earnings by applying the median weekly hours worked to the reported hourly rate, using data from the 2019 Annual Population Survey, conditional on industry sector and contract type.[9]

Analysis of Panel E in

Table 1.3 indicates that the majority (76%) of graduates secured employment within six months of graduation. Additionally, 44% of female graduates engaged in postgraduate education or training for at least one month, compared to 35% of male graduates. This gender difference is statistically significant at the 10% level. The average annual gross earnings were estimated at £21,800, with a statistically significant gender wage gap of £4,190 per year (5% level), reflecting disparities in post-graduation earnings.

# 1.5. Empirical results

## 1.5.1 Empirical Strategy

We investigate the treatment effect of the information intervention on skills perception, investment in employability skills and study, as well as job search behaviours and labour market outcomes following the below regression:

$$y_{i,t} = \alpha + \gamma Treat_i + \beta y_{i,t-1} + \delta X_i + \epsilon_{i,t}$$

where $y_{i,t}$ is the post-treatment outcomes of interest, $Treat_i$ indicates the treatment status. The use of baseline $y_{i,t-1}$and a series of control variables $X_i$ including all the stratifying variables and whether receive intervention in the first and second year improve the precision of results (Bruhn & McKenzie, 2009; McKenzie, 2012).

For most variables collected from the survey (e.g., skill perceptions, experience, and study hours), baseline measurements were obtained prior to the intervention at Wave 9 during the Autumn term and are included as additional control variables in the regression analysis. However, certain variables such as job search behaviours and labour market outcomes, do not have baseline measures. For these variables, we rely solely on their post-intervention values. Additionally, for outcomes derived from administrative data, baseline

[9] Outliers in annual gross earnings are replaced to 1% and 99% percentiles respectively.

values were determined based on either the academic term or the first week of the laboratory session.

Given the notable gender differences observed in the descriptive statistics, we focus on assessing the heterogeneity of treatment effects by gender. To formally test whether these gender differences in treatment effects are statistically significant, we employ a fully interacted regression model.

$$y_{i,t} = \alpha + \gamma Treat_i + \beta y_{i,t-1} + \delta X_i + \theta(fe_i \times Treat_i) + \zeta(fe_i \times y_{i,t-1}) + \eta(fe_i \times X_i) + \epsilon_{i,t}$$

where $fe_i = 1$ indicates being female, and it is fully interacted with other variables. The outcome of interest here is $\theta$, which represents the difference of the treatment effect between males and females. The outputs of this regression are not shown, and we only report the p-value of $\theta$.

## 1.5.2 Treatment Effect on Skills Perception

Table 1.4 reports the estimated treatment effects on the three dimensions of students' awareness of employability skills.

### *1.5.2.1 Rating on Importance of Skills*

We begin by assessing the treatment effect on students' perceptions of the importance of skills demanded by employers. Panel A presents the results for the average rating of the importance of five employability skills, while Figure 1.5 provides a more detailed breakdown of the treatment effects for each skill.

Across columns (1)–(3) of Table 1.4 Panel A, we observe no significant treatment effects overall. This outcome variable reflects students' subjective beliefs regarding the average proportion of employers that prioritise graduates with these five key skills.

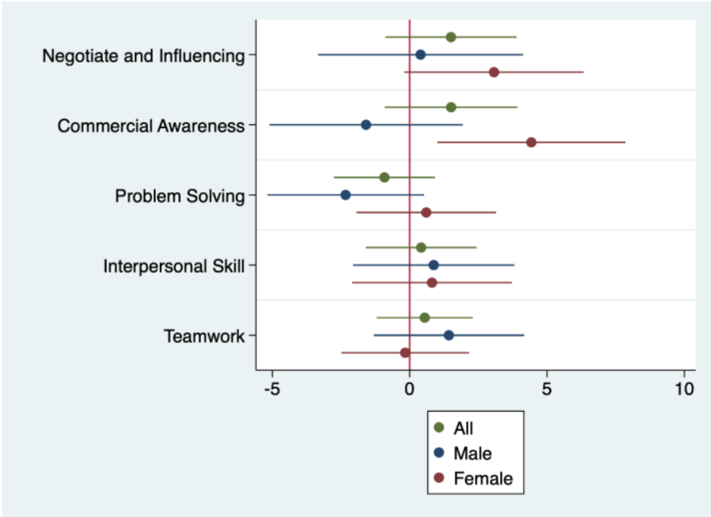
One possible explanation for the absence of treatment effects is the minimal discrepancy between the objective information provided and students' pre-existing beliefs at baseline. For instance, students initially rated the importance of 'Teamwork' at 83, indicating their belief that 83% of employers prioritise teamwork skills in their recruitment process. However, the intervention revealed that the true proportion was 81%. Similarly, the baseline ratings for 'Commercial Awareness' (73), 'Problem-Solving' (81), and 'Interpersonal Skills' (80) were closely aligned with the actual figures presented in the intervention (71, 77, and 76, respectively).

*Table 1.4 Treatment effects on subjective beliefs*

| | (1) All | (2) Male | (3) Female |
|---|---|---|---|
| | *Panel A: Average rating on the importance of skills to employers* | | |
| Treatment | 0.441 | -1.088 | 1.500 |
| | (0.933) | (1.415) | (1.251) |
| Baseline | 0.211*** | 0.152** | 0.272*** |
| | (0.043) | (0.063) | (0.053) |
| p-value of the difference | | | 0.330 |
| $R^2$ | 0.058 | 0.039 | 0.082 |
| Observations | 675 | 292 | 383 |
| | *Panel B: Average self-rating on skills* | | |
| Treatment | -1.987*** | -0.891 | -3.012*** |
| | (0.758) | (1.132) | (1.026) |
| Baseline | 0.465*** | 0.391*** | 0.540*** |
| | (0.036) | (0.055) | (0.042) |
| p-value of the difference | | | 0.064* |
| $R^2$ | 0.299 | 0.263 | 0.331 |
| Observations | 676 | 293 | 383 |
| | *Panel C: Average self-rating on signalling skills* | | |
| Treatment | -3.230*** | -2.581* | -3.722*** |
| | (0.982) | (1.389) | (1.368) |
| p-value of the difference | | | 0.422 |
| $R^2$ | 0.018 | 0.010 | 0.017 |
| Observations | 770 | 332 | 438 |

*Notes:* All regressions include control for gender, ethnicity, mature student status, social-economic status, tuition fee status, department, tariff quintiles and whether receive intervention in the first and second year. Estimation sample for Panel A and B consists of students who attended both Wave 10 and 9 as information collected in Wave 9 are used for baseline. Estimation sample for Panel C only consists of students who attended Wave 10 as no baseline answers were collected. The subject beliefs were elicited after the intervention in Wave 10. Robust standard errors in parentheses. * p<0.1, ** p<0.05, * p<0.01

*Figure 1.5 Treatment effect on rating the importance of skills to employers*




Nevertheless, given the large variance in students' baseline perceptions, it is still possible that the treatment narrowed the bounds of students' ratings without changing the average ratings. However, this does not appear to be the case, as the distribution of the ratings for five skills remained mostly unchanged post-intervention.[10]

Notably, we do not observe any significant treatment effect on 'Negotiating and Influencing', despite this skill exhibiting the largest perception gap at baseline. Students initially overestimated its importance by more than 20 percentage points; however, there is no evidence of downward belief updating. Even when the sample is segmented based on whether students initially overestimated or underestimated the proportion of employers prioritising this skill, no significant treatment effect is detected.

A potential explanation for this finding is that 'Negotiating and Influencing' was explicitly identified as a skill in shortage during the intervention. As a result, students may have internalised this message and, rather than revising their estimates downward, maintained their perception of strong employer demand. This interpretation is further supported by the treatment effect observed for 'Commercial Awareness,' the other skill classified as being in shortage. Although female students had already slightly overestimated the proportion of employers prioritising 'Commercial Awareness' in

---

[10] We compare the distribution of self-reported ratings on the perceived employers' demand for skills before and after intervention by treatment status, suggesting no major shift. Figures are provided in A 1.2 from Appendix 1.

recruitment (73.79%—2.79 percentage points above the actual figure), the intervention led to a significant positive treatment effect of 5 percentage points.

#### *1.5.2.2 Self-rating on Skills and Signalling Skills*

Panels B and C of Table 1.4 examine the other two dimensions of skills perception: self-assessment of skills and self-assessment of skills signalling. The results indicate significant negative treatment effects on both measures. Column (1) of Panel B shows that students in the treatment group lowered their average self-rating across the five skills by approximately 2 points, while column (1) of Panel C reveals a reduction of 3.23 points in their average rating of how well they signal these skills to employers. Both estimates are statistically significant at the 1% level.

Next, we examine the effects separately for male and female students. Column (3) of Panel B shows a significant negative effect of 3.012 points on the average self-ratings of female students, significant at the 1% level. In contrast, column (2) indicates no significant effect for male students. The difference in treatment effects between males and females is 2.121 points, with a p-value of 0.064, suggesting that treated female students tend to be more self-critical in evaluating their skills compared to their male counterparts. Figure 1.6 provides further insights into these gender differences, demonstrating that treated female students significantly lowered their self-assessments across four of the five skills, with 'Teamwork' being the only exception. Conversely, treated male students exhibited a significant decline in self-rating for only one skill.

*Figure 1.6 Treatment effects on self-rating of skills*

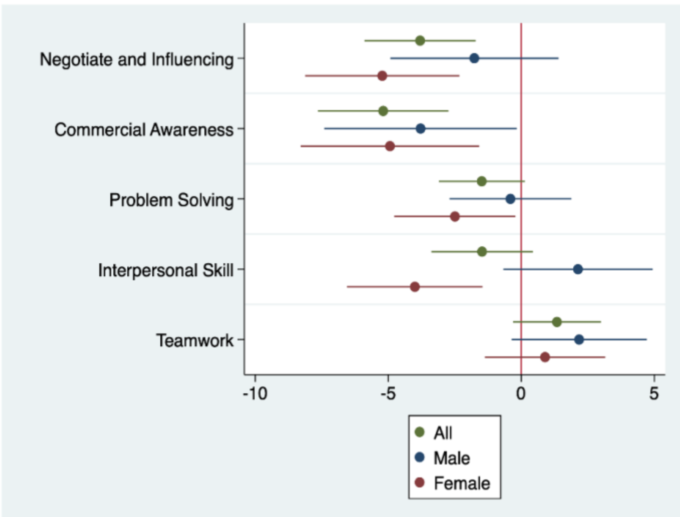
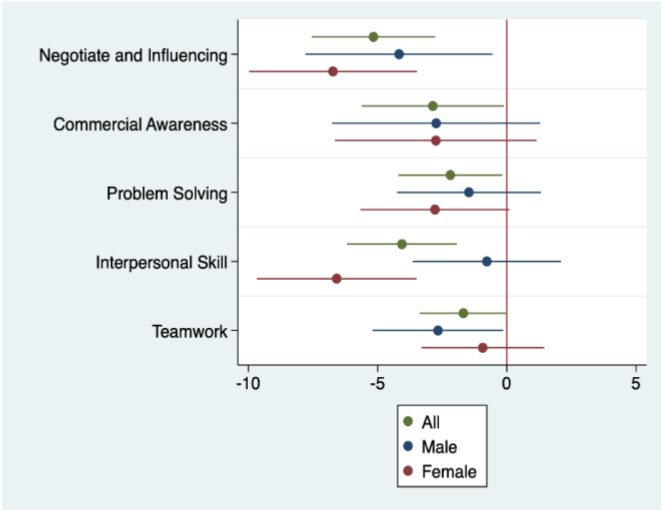


*Figure 1.7 Treatment effects on self-rating of signalling skills*




Columns (2) and (3) of Table 1.4 Panel C suggest that both treated male and female students perceived a need for improvement in signalling their skills to employers. Figure 1.7 indicates no substantial gender differences in treatment effects, except for 'Interpersonal Skills,' where females showed a more pronounced decline.

Notably, female students significantly revised their self-assessments downward immediately after the intervention. However, as these outcomes were also measured in subsequent Wave 11, we assess whether these effects persisted over time. Two months after the intervention, we find no significant treatment effects on self-perception for either males or females. One possible explanation is that the initial negative shock induced by the intervention may have motivated students to invest in skill development during the interim period, ultimately leading to an adjustment in their self-perceptions back to their pre-treatment levels. This interpretation aligns with subsequent findings.

## 1.5.3 Treatment Effect on Investment in Employability Skills

Table 1.5, Table 1.6, and Table 1.7 examine the treatment effects on investments in employability skills.

Panel A of Table 1.5 indicates no significant treatment effect on the likelihood of students securing study- or career-related experience during university term time. This finding may be attributed to the inherent difficulty of obtaining such opportunities, particularly within a short timeframe. Notably, the data were collected only one month after the intervention. However, there is some evidence to suggest that students in the treatment group were actively seeking relevant work, as indicated by a significant 6.1 percentage point reduction (at the 10% level) in participation in non-career-related or non-academic activities, as shown in Panel B.

*Table 1.5 Treatment effects on employability investment*

| | (1) All | (2) Male | (3) Female |
|---|---|---|---|
| | *Panel A: Study/desired career related experience* | | |
| Treatment | 0.001 | -0.060 | 0.030 |
| | (0.025) | (0.037) | (0.034) |
| Baseline | 0.215*** | 0.258*** | 0.187*** |
| | (0.035) | (0.058) | (0.043) |
| p-value of the difference | | | 0.451 |
| $R^2$ | 0.144 | 0.255 | 0.164 |
| | *Panel B: Other employment/ non-academic experience* | | |
| Treatment | -0.061* | -0.072 | -0.044 |
| | (0.036) | (0.060) | (0.047) |
| Baseline | 0.532*** | 0.440*** | 0.611*** |
| | (0.039) | (0.060) | (0.051) |
| p-value of the difference | | | 0.714 |
| $R^2$ | 0.285 | 0.299 | 0.327 |
| | *Panel C: The BigE Award Enrollment* | | |
| Treatment | 0.056** | 0.040 | 0.067* |
| | (0.026) | (0.036) | (0.039) |
| Baseline | 0.731*** | 0.705*** | 0.750*** |
| | (0.031) | (0.060) | (0.038) |
| p-value of the difference | | | 0.605 |
| $R^2$ | 0.593 | 0.594 | 0.573 |
| Observations | 599 | 248 | 351 |

*Notes:* All regressions include control for gender, ethnicity, mature student status, social-economic status, tuition fee status, department, tariff quintiles and whether receive intervention in the first and second year. Estimation sample consists of those who who attended Wave 10 and responded in both Wave 9 and 11. Information collected in Wave 9 are used for baseline while the outcomes are collected in Wave 11. Robust standard errors in parentheses. * p<0.1, ** p<0.05, * p<0.01

*Table 1.6 Treatment effects on employability (in hours)*

| | (1) All | (2) Male | (3) Female |
|---|---|---|---|
| | *Panel A: Study/desired career related experience (hours)* | | |
| Treatment | 0.188 | -1.452 | 0.849* |
| | (0.572) | (1.275) | (0.490) |
| Baseline | 0.127*** | 0.103** | 0.293*** |
| | (0.049) | (0.047) | (0.073) |
| p-value of the difference | | | 0.090* |
| $R^2$ | 0.094 | 0.333 | 0.270 |
| | *Panel B: Other employment/non-academic experience (hours)* | | |
| Treatment | 0.236 | -2.035 | 1.788 |
| | (1.334) | (1.581) | (2.055) |
| Baseline | 0.434*** | 0.409*** | 0.501*** |
| | (0.046) | (0.042) | (0.126) |
| p-value of the difference | | | 0.141 |
| $R^2$ | 0.220 | 0.490 | 0.169 |
| | *Panel C: The BigE Award (hours)* | | |
| Treatment | -0.251 | -0.126 | -0.381 |
| | (0.747) | (0.986) | (0.984) |
| Baseline | 0.266*** | 0.472** | 0.192** |
| | (0.094) | (0.217) | (0.077) |
| p-value of the difference | | | 0.854 |
| $R^2$ | 0.165 | 0.283 | 0.157 |
| Observations | 599 | 248 | 351 |

*Notes:* All regressions include control for gender, ethnicity, mature student status, social-economic status, tuition fee status, department, tariff quintiles and whether receive intervention in the first and second year. Estimation sample consists of those who who attended Wave 10 and responded in both Wave 9 and 11. Information collected in Wave 9 are used for baseline while the outcomes are collected in Wave 11. Robust standard errors in parentheses. * p<0.1, ** p<0.05, * p<0.01

This interpretation is further supported by column (3) of Panel A in Table 1.6, which documents a significant 0.85-hour increase (at the 10% level) in weekly working hours among treated female students. Given that no significant effect was observed on the extensive margin, this increase is likely driven by students who were already engaged in study- or career-related employment. Conversely, treated male students exhibited a decline of 1.5 hours in weekly working hours within study- or career-related employment (column 2), though this coefficient was not statistically significant. Nonetheless, the difference in treatment effects between males and females is statistically significant at the 10% level, suggesting that treated female students extended their working hours in career-related roles to a greater extent than their male counterparts.

Panel C of Table 1.5 also highlights a positive impact on participation in the university's BigE Award. Students in the treatment group were 5.6 percentage points more likely to enrol in the programme, a result that is significant at the 5% level. When examining female students separately, the effect size increased by 1.1 percentage points, though the significance level dropped to 10%. In contrast, the effect size decreased for male students and was no longer significant. Changes in participation intensity, as measured in Table 1.6, Panel C, appear minimal.

Lastly, Table 1.7 presents the treatment effect on attendance at career events. While no significant overall treatment effect is detected, treated female students (column 3) attended 0.273 more career events than those in the control group. Given that the average number of bookings among female students prior to the intervention was 0.980, this effect represents an increase of nearly 30%. Additionally, the gender difference in treatment effects is statistically significant, with a p-value of 0.019.

*Table 1.7 Treatment effects on career events*

| | *Number of Career events attended* | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| | All | Male | Female |
| Treatment | 0.082 | -0.120 | 0.273** |
| | (0.084) | (0.115) | (0.122) |
| Baseline | 0.388*** | 0.381*** | 0.385*** |
| | (0.062) | (0.068) | (0.085) |
| p-value of the difference | | | 0.019** |
| $R^2$ | 0.236 | 0.309 | 0.252 |
| Observations | 770 | 332 | 438 |

*Notes:* All regressions include control for gender, ethnicity, mature student status, social-economic status, tuition fee status, department, tariff quintiles and whether receive intervention in the first and second year. Estimation sample consists of students who received treatment in Wave 10. Outcome variable is obtained from administrative data. The baseline is specified as the number of career events attended in third year (2017/18) prior to the treatment. As individuals chose to attend the lab session on different days of a given period, a universal cut-off point is applied to define baseline and post-treatment for all participants. Columns (1)-(3) select the first day of the lab session (22/January/2018). Robust standard errors in parentheses. * p<0.1, ** p<0.05, * p<0.01

In summary, no significant behavioural changes were observed among treated male students. However, treated female students exhibited greater investment in employability skills through three key activities: (i) increased likelihood of enrolling in the BigE Award, (ii) increased working hours in study- or career-related employment, and (iii) increased attendance at career events.

### 1.5.4 Treatment Effect on Academic Investment

Table 1.8 presents the effects of the intervention on academic outcomes in Year 3 and at graduation.

*Table 1.8 Treatment effects on academic investment*

| | *Year 3 mark* | | | *Degree mark* | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | All | Male | Female | All | Male | Female |
| Treatment | 0.656 | -1.062 | 1.837** | 0.507 | -0.084 | 1.093* |
| | (0.729) | (1.235) | (0.931) | (0.455) | (0.799) | (0.560) |
| Baseline | 0.578*** | 0.629*** | 0.542*** | 0.590*** | 0.636*** | 0.557*** |
| | (0.043) | (0.075) | (0.056) | (0.027) | (0.043) | (0.036) |
| p-value of the difference | | | 0.061* | | | 0.227 |
| $R^2$ | 0.353 | 0.395 | 0.359 | 0.537 | 0.562 | 0.552 |
| Observations | 583 | 254 | 329 | 725 | 314 | 411 |

| | *First class* | | | *Upper second class or above* | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | All | Male | Female | All | Male | Female |
| Treatment | 0.052* | 0.016 | 0.084** | -0.013 | -0.040 | 0.015 |
| | (0.027) | (0.042) | (0.035) | (0.028) | (0.044) | (0.038) |
| Baseline | 0.025*** | 0.025*** | 0.024*** | 0.022*** | 0.022*** | 0.022*** |
| | (0.001) | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) |
| p-value of the difference | | | 0.218 | | | 0.873 |
| $R^2$ | 0.367 | 0.403 | 0.398 | 0.317 | 0.356 | 0.353 |
| Observations | 725 | 314 | 411 | 725 | 314 | 411 |

*Notes:* All regressions include control for gender, ethnicity, mature student status, social-economic status, tuition fee status, department, tariff quintiles and whether receive intervention in the first and second year. Estimation sample consists of students who received treatment in Wave 10. Outcome variable is obtained from administrative data. Degree mark is computed as 40% of the second year mark and 60% of the third year mark. First class is defined as those who achieved 70 or above while upper second class or above is defined as those who achieved above 60. The baseline is specified as the average overall mark of first year (2015/16). Robust standard errors in parentheses. * p<0.1, ** p<0.05, * p<0.01

We first examine the overall mark in Year 3, as it represents the final academic outcome where any investments made after the intervention should be reflected. While no significant effects are observed in columns (1) and (2), we find that female students who received the intervention achieved, on average, 1.84 points higher in their final Year 3 mark than those in the control group. This effect is statistically significant and substantively meaningful. Given that the average baseline mark for female students was 61.36, this result suggests an approximate 3% increase in marks during the final term of university. Furthermore, this improvement appears to have contributed to an increase in the final degree mark. As the final degree mark is calculated using a weighted combination of 60% of the Year 3 mark and 40% of the second-year mark, the significant positive effect of 1.09 points observed in column (6) under 'Degree mark' can be attributed entirely to the improvement in Year 3 performance. Since the intervention took place in the middle of the third year, it could not have influenced the second-year mark.

Turning to final degree classifications, we find that students in the treatment group were 5.2 percentage points more likely to graduate with first-class honours. The effect is stronger and more statistically significant for female students but diminishes and becomes insignificant for male students. However, there is no statistically significant difference between the treatment effects for males and females. Notably, the intervention had no impact on the likelihood of obtaining an upper second-class or higher degree, suggesting that it primarily shifted students from achieving upper second-class honours to first-class honours rather than improving overall pass rates.[11]

### 1.5.5 Treatment effect on job search behaviours and labour market outcomes

Table 1.9 presents the estimated treatment effects on job search behaviours. Notably, significant positive effects are observed exclusively among treated male students. By three months prior to graduation, males in the treatment group were 11 percentage points more likely to have initiated their job search (column (2)) and 6.7 percentage points more likely

[11] We also examined potential mechanisms underlying the information treatment effect on academic outcomes using administrative attendance data and self-reported study hours. Results presented in A 1.3 from Appendix 1 show no significant effects on these intermediate outcomes, suggesting the mechanism remains inconclusive.

to have secured at least one job offer (column (5)). Both estimates are statistically significant at the 10% level. These findings, combined with the mean values reported in Table 1.3 Panel E, suggest that treated males were more inclined to begin job searching earlier and secure offers more quickly. This is particularly notable given that only approximately 32% of males in our sample had started submitting job applications at that stage, and fewer than 10% had secured a job offer. Conversely, no significant changes in job search behaviours were observed among female students.

*Table 1.9 Treatment effects on job search behaviours*

| | *Job Application Sent* | | | *Job Offer Secured* | | |
|---|---|---|---|---|---|---|
| | (1) All | (2) Male | (3) Female | (4) All | (5) Male | (6) Female |
| Treatment | -0.003 | 0.110* | -0.080 | 0.022 | 0.067* | -0.005 |
| | (0.038) | (0.060) | (0.050) | (0.024) | (0.035) | (0.034) |
| p-value of the difference | | | 0.015** | | | 0.139 |
| $R^2$ | 0.066 | 0.113 | 0.095 | 0.053 | 0.114 | 0.063 |
| Observations | 653 | 267 | 386 | 653 | 267 | 386 |

*Notes:* All regressions include control for gender, ethnicity, mature student status, social-economic status, tuition fee status, department and tariff quintiles. Estimation sample consists of students who received treatment in Wave 10 as well as responded in Wave 11. Job search questions were asked roughly three months after the treatment and there was no baseline. Robust standard errors in parentheses. * p<0.1, ** p<0.05, * p<0.01

Table 1.10 presents the estimated effects on realised labour market outcomes collected two years after graduation. We find that treated female students were 6.9 and 6.7 percentage points more likely to be employed within three- and six-months post-graduation, respectively (columns (3) and (6)), whereas treated male students were 8 and 5.5 percentage points less likely to be employed within the same periods relative to their control group counterparts (columns (2) and (5)). However, none of these estimates are statistically significant. Additionally, treated female students were 11.3 percentage points less likely to pursue postgraduate education or training, a result that is both statistically significant and substantial.

With respect to employment contract type, treated male graduates were 14 percentage points less likely to be employed under stable long-term contracts and 12 percentage points more likely to hold casual or short-term contracts. Both estimates are statistically significant at the 5% level. While the corresponding coefficients for female graduates suggest the opposite pattern, they do not reach statistical significance. However, the differences in treatment effects by gender are statistically significant, with p-values of 0.018 and 0.008, respectively.

*Table 1.10 Treatment effects on realised labour market outcomes*

| | *Employed within 3 months after Graduation* | | | *Employed within 6 months after Graduation* | | |
|---|---|---|---|---|---|---|
| | (1) All | (2) Male | (3) Female | (4) All | (5) Male | (6) Female |
| Treatment | -0.004 | -0.080 | 0.069 | 0.013 | -0.055 | 0.067 |
| | (0.045) | (0.077) | (0.060) | (0.040) | (0.068) | (0.054) |
| p-value of the difference | | | 0.124 | | | 0.156 |
| $R^2$ | 0.107 | 0.184 | 0.180 | 0.125 | 0.195 | 0.196 |
| Observations | 435 | 175 | 260 | 435 | 175 | 260 |
| | *Fixed/Long-term Employment* | | | *Casual/Short-term Employment* | | |
| | (1) All | (2) Male | (3) Female | (4) All | (5) Male | (6) Female |
| Treatment | -0.042 | -0.140** | 0.030 | 0.017 | 0.120** | -0.058 |
| | (0.035) | (0.054) | (0.047) | (0.033) | (0.049) | (0.046) |
| p-value of the difference | | | 0.018** | | | 0.008*** |
| $R^2$ | 0.117 | 0.350 | 0.165 | 0.104 | 0.354 | 0.145 |
| Observations | 398 | 161 | 237 | 398 | 161 | 237 |
| | *Post Education Training* | | | *Gross Annual Earnings (log)* | | |
| | (1) All | (2) Male | (3) Female | (4) All | (5) Male | (6) Female |
| Treatment | -0.062 | 0.002 | -0.113* | 0.002 | -0.059 | 0.071 |
| | (0.046) | (0.076) | (0.063) | (0.060) | (0.088) | (0.082) |
| p-value of the difference | | | 0.240 | | | 0.624 |
| $R^2$ | 0.159 | 0.235 | 0.193 | 0.130 | 0.272 | 0.167 |
| Observations | 435 | 175 | 260 | 396 | 163 | 233 |

*Notes:* All regressions include control for gender, ethnicity, mature student status, social-economic status, tuition fee status, department and tariff quintiles. Estimation sample consists of students who attended Wave 10 and responded in Wave 14 (2 years after graduation). Robust standard errors in parentheses. * p<0.1, ** p<0.05, * p<0.01

Finally, columns (4)–(6) report the treatment effects on gross annual log-transformed earnings. While none of the coefficients are precisely estimated, the estimated treatment effect for male graduates suggests a negative direction, contrary to expectations.

Although preferences for employment contract types vary among individuals, stable long-term contracts are generally considered more desirable or at least a more favourable non-pecuniary job attribute than casual or short-term contracts. It is therefore surprising that male graduates who received the intervention were more likely to experience less favourable employment conditions. One potential explanation for this outcome emerges when considering the significant changes observed in job search behaviours, coupled with the lack of significant effects on employability skill investment or academic performance among male students. It appears that treated males may not have interpreted the information intervention as intended. Instead, they may have perceived it as signalling an increasingly competitive labour market, prompting them to prioritise early job search efforts and accept any available position, even if it was not ideal.

## 1.6 Robustness Check

We conduct a series of sensitivity analyses to examine whether our estimates are influenced by the exclusion of baseline outcomes and/or the choice of control variables. The main conclusions remain robust.

A potential concern in our analysis is the presence of nonresponse and panel attrition in the longitudinal survey data. Although treatment was randomly assigned in Wave 10, participation in subsequent survey waves was voluntary. Of the Wave 10 participants, 654 students (85%) responded two months post-intervention in Wave 11, while only 435 students (56.5%) remained in Wave 14, approximately 28 months after the intervention. This attrition could impact the credibility of our results in two ways.

First, if attrition is systematically related to treatment status, then estimates of labour market outcomes could be subject to selection bias. For instance, treated students may be more or less likely to participate in later waves based on their perceived benefits or losses from the intervention, potentially leading to over- or underestimation of treatment effects. To assess this, we regress the conditional probability of participating in Wave 14 on treatment assignment and other stratifying variables. The regression result suggests no evidence of systematic selection bias.[12]

*Table 1.11 Treatment effect with inverse probability weighting*

| | (1) All | (2) Male | (3) Female |
|---|---|---|---|
| | *Panel A: Outcomes Collected in Wave 11* | | |
| Study/desired career related experience | 0.002 | -0.056 | 0.025 |
| | (0.024) | (0.036) | (0.031) |
| Other employment/non-academic experience | -0.046 | -0.038 | -0.039 |
| | (0.038) | (0.059) | (0.048) |
| The BigE Award Enrollment | 0.061* | 0.066*** | 0.048*** |
| | (0.036) | (0.050) | (0.051) |
| Study/desired career related experience (hours) | -0.013 | -1.614 | 0.672*** |
| | (0.569) | (1.055) | (0.511) |
| Other employment/non-academic experience (hours) | -1.151 | -3.586** | -0.053 |
| | (1.295) | (1.674) | (1.783) |
| The BigE Award (hours) | -0.525 | 0.109 | -0.927 |
| | (0.687) | (0.941) | (0.901) |
| Study Hours per week | -0.509 | 0.140 | -0.703 |
| | (0.906) | (1.217) | (1.249) |
| Job Application Sent | 0.000 | 0.105* | -0.087* |
| | (0.037) | (0.055) | (0.047) |
| Job Offer Secured | 0.020 | 0.062* | -0.003 |
| | (0.023) | (0.032) | (0.316) |
| Observations | 654 | 267 | 387 |
| | *Panel B: Outcomes Collected in Wave 14* | | |
| Employed within 3 months after Graduation | -0.010 | -0.087 | 0.054 |
| | (0.043) | (0.066) | (0.056) |
| Employed within 6 months after Graduation | 0.008 | -0.058 | 0.043 |
| | (0.038) | (0.058) | (0.052) |
| Fixed/Long-term Employment | -0.034 | -0.136*** | 0.044 |
| | (0.032) | (0.044) | (0.041) |
| Casual/Short-term Employment | 0.013 | 0.115*** | -0.066* |
| | (0.030) | (0.039) | (0.040) |
| Post Education Training | -0.059 | -0.013 | -0.103* |
| | (0.044) | (0.067) | (0.059) |
| Gross Annual Earnings (log) | 0.003 | -0.064 | 0.064 |
| | (0.057) | (0.074) | (0.073) |
| Observations | 435 | 175 | 260 |

*Notes:* All estimates represent average treatment effect(ATE) obtained from inverse probability weighting(IPW) estimator, which contains two steps. First, estimate the parameters of the treatment using a probit model and compute the predicted inverse probability weights based on stratifying variables. Then, apply the weights to compute weighted averages of the outcomes for treated and control groups. The estimated ATE is the difference of these weighted averages. Stata built-in function implements both steps at once so that the standard errors are correct. Robust standard errors in parentheses. * p<0.1, ** p<0.05, * p<0.01

[12] Regression results are presented in A 1.4 in Appendix 1.

Second, nonresponse and attrition may introduce unequal sampling probabilities across the target population. While this is not an issue if survey participation is independent of factors affecting the outcome—as it would be in a compulsory survey—it becomes problematic if attrition is correlated with characteristics that influence the outcome. For example, graduates in demanding jobs may be both more likely to earn higher incomes and less likely to respond to Wave 14, thereby biasing the estimated effect of treatment on earnings.

To mitigate potential attrition bias, we apply inverse probability weighting (IPW) to adjust for under- and over-representation due to differential response rates. Instead of treating all observations equally, we assign weights based on the predicted probability of inclusion in our analytical sample, using stratifying variables. Table 1.11 presents the estimates obtained from the IPW estimator. While most conclusions remain unchanged, some estimates are moderately sensitive to attrition correction. For instance, the previously observed reduction in participation in non-related or non-academic experiences (Panel B, Table 1.5), which was significant at the 10% level, is no longer statistically significant. Conversely, several treatment effects that were initially insignificant become larger and statistically significant, reinforcing the observed gender-based differences in response to the intervention.

Following the information treatment, both male and female treated students exhibited an immediate increase in enrolment in the Big Employability Award, with probabilities rising by 6.6 and 4.8 percentage points, respectively. Treated females increased their weekly hours spent on study- or career-related activities by approximately 0.7 hours, while treated males reduced their time spent in other employment or non-academic activities by 3.6 hours per week. Furthermore, treated female students were less likely to submit job applications two months post-intervention and less likely to be employed in casual or short-term contracts 22 months after graduation. In contrast, treated male students exhibited the opposite effects, suggesting divergent employment trajectories as a result of the intervention.

## 1.7 Conclusion

Universities have long faced criticism from industry stakeholders for producing graduates with strong academic credentials but insufficient employability skills. As employers increasingly prioritise candidates who demonstrate such competencies, a crucial question arises: are university students aware of the skills in demand, and do they know how to acquire them? This study seeks to address this gap in skills perception between employers and graduates through an information intervention.

Using a comprehensive dataset combining longitudinal survey responses and administrative records, we examine the direct effects of the intervention on students' perceptions of employability skills. Our findings indicate that the intervention led to lower self-assessments of skills, with significant effects observed among female students but no corresponding impact on male students. Additionally, both male and female students in the treatment group reported a decline in their perceived ability to signal these skills to employers. For female students, the intervention had a positive effect on career event participation and academic outcomes. In contrast, treated male students exhibited a higher propensity to initiate their job search earlier but were also less likely to secure stable, long-term contracts upon graduation.

A key finding of this study is the substantial heterogeneity in treatment effects by gender. One possible explanation is that male and female students interpreted the information provided during the intervention differently, potentially reflecting a gender confidence gap. Female students appeared to take the information at face value, engaging in skill development and signalling activities as recommended. Conversely, male students may have perceived the emphasis on employability skills as an indication of increased labour market competition, prompting them to expedite their job search rather than invest further in skill acquisition.

Our findings align with a growing body of literature that links the gender pay gap to differences in self-confidence, suggesting that women may be less assured of their abilities than men. This confidence gap is not limited to university students but has also been observed among economists (Sarsons & Xu, 2021). The underlying causes of this phenomenon remain elusive, and the existing economic literature on this topic is relatively limited. While our study highlights notable gender differences in skills perception and investment decisions, it does not establish whether these differences are directly driven by the gender confidence gap. Future research should further investigate these dynamics, particularly the role of confidence in shaping labour market outcomes and career trajectories.

# Appendix 1

## *A 1.1 Information Treatment Materials*

"So the degree of graduate studies isn't really important at all. We take you from any degree discipline, and the reason for that is we are looking for a diverse organisation. We are looking for people to bring lots of different perspectives to that work and that means we want people that are studying a wide variety of different topics. The key is to get a good degree grade, and so to be getting a 2:1 degree in any subject is absolutely what we are looking for. The skills that students need to be developing beyond their academics whist at university as things like getting some work experience perhaps, doing some work with some societies or clubs while at university and that could be sports societies, cultural societies and it doesn't have to be related to the job they want to go into. It's more about developing their abilities and their skills through doing other things and that what we are looking for. So work experience is becoming more and more important to student these days. A lot of competition for jobs and having a bit of insight into business before they join an organisation is really really helpful, especially develop that commercial awareness that we are looking for. And that work experience doesn't have to be related to the workplace they are joining and certainly in my own experience I had worked in factories, as a waitress in a bar. Nothing related to the office environment at all. What it does teach you is a lot of skills that you can transfer to the workplace. Things like dealing with customers, dealing with difficult situations, thinking on your feet and working in teams, training and developing other people and all of those skills are really useful in any work environment so it can be really related to the new job that they apply to."

Here are 5 skills that employers are looking for when hiring graduates:

1. Teamwork
2. Interpersonal skills
3. Problem-solving
4. Commercial awareness
5. Negotiating and influencing

Can you match each skill to its correct definition?

<< >>

What is the correct definition for **Teamwork?**

Knowledge of the sector you want to work in, the challenges of the organizations operating in that sector, and what makes your prospective employer distinctive.

Ability to discuss an issue with others to find ways to reach an agreement, and ability to change others' minds about a topic by taking their perspectives into account.

Being able to operate smoothly and efficiently within a group.

Being able to using logic, as well as imagination, to make sense of your situation and come up with an intelligent solution.

Ability to understand what motivates others and how they use their knowledge to get the best results.

>>

Correct!

**Teamwork is being able to operate smoothly and efficiently within a group.**

>>

These are the **top 3 skills employers** are looking for when hiring graduates:

**1. Teamwork**: Ability to operate smoothly and efficiently within a group and the ability to build positive working relationships that help everyone to achieve their goals.

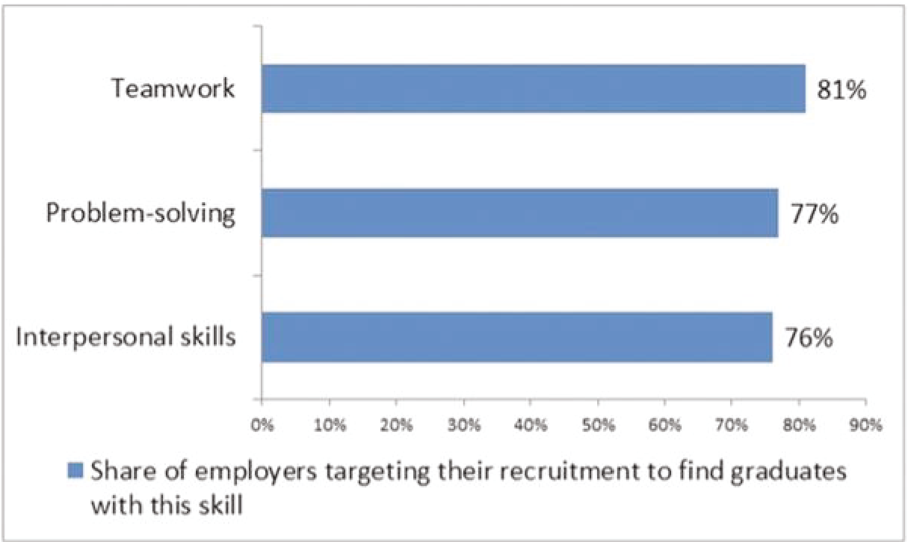
**2. Interpersonal skills:** Ability to understand what motivates others and how they use their knowledge to get the best results.

**3. Problem-solving:** Ability to identify and define a problem, generating solutions, and implementing the best alternative by taking a logical and analytical approach.

Source: Association of Graduate Recruiters 2016

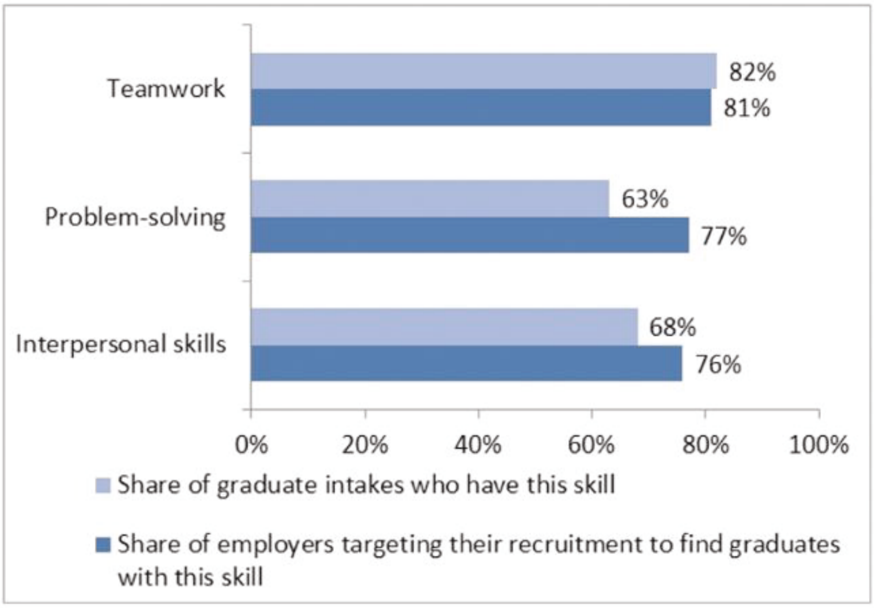
>>

This is the share of employers tailoring their recruitment to find graduates with each of these skills:




This means that:

- 81% of employers target their hiring to find graduates able to work in a team.
- 77% of employers of employers target their hiring to find graduates who are good at solving problems.
- 76% of employers target their hiring to find graduates who have good interpersonal skills.




Because most graduates have these 3 skills, you need to make sure **you show you have them when you look for a job!** **These are essential skills!**

<<

>>

But there are also skills that employers say are hard to find among graduates. So you might be able to **distinguish yourself from the crowd** if you show you have either of these:

**1. Commercial awareness:** Knowledge of the sector you want to work in, the challenges of the organizations operating in that sector, and what makes your prospective employer distinctive.

**2. Negotiating and Influencing:** The ability to discuss an issue with others to find ways to reach an agreement, and ability to change others' minds about a topic by taking their perspectives into account.

<<

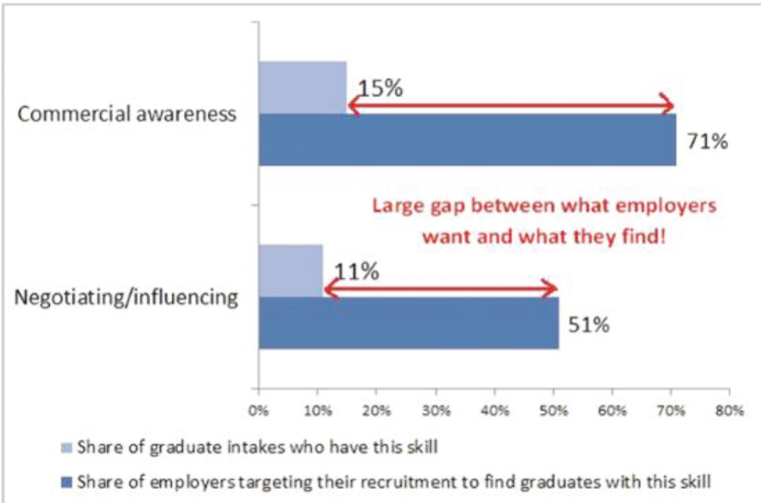
>>

Take a look at the gap between what share of employers want these skills and the share of graduates intakes that have them.


This means that:

- 71% of employers are targeting their hiring to find graduates with commercial awareness, but they say only 15% of graduates have this skill!
- 51% of employers are targeting their hiring to find graduates with negotiating and influencing skills, but only 11% of graduates have this skill!

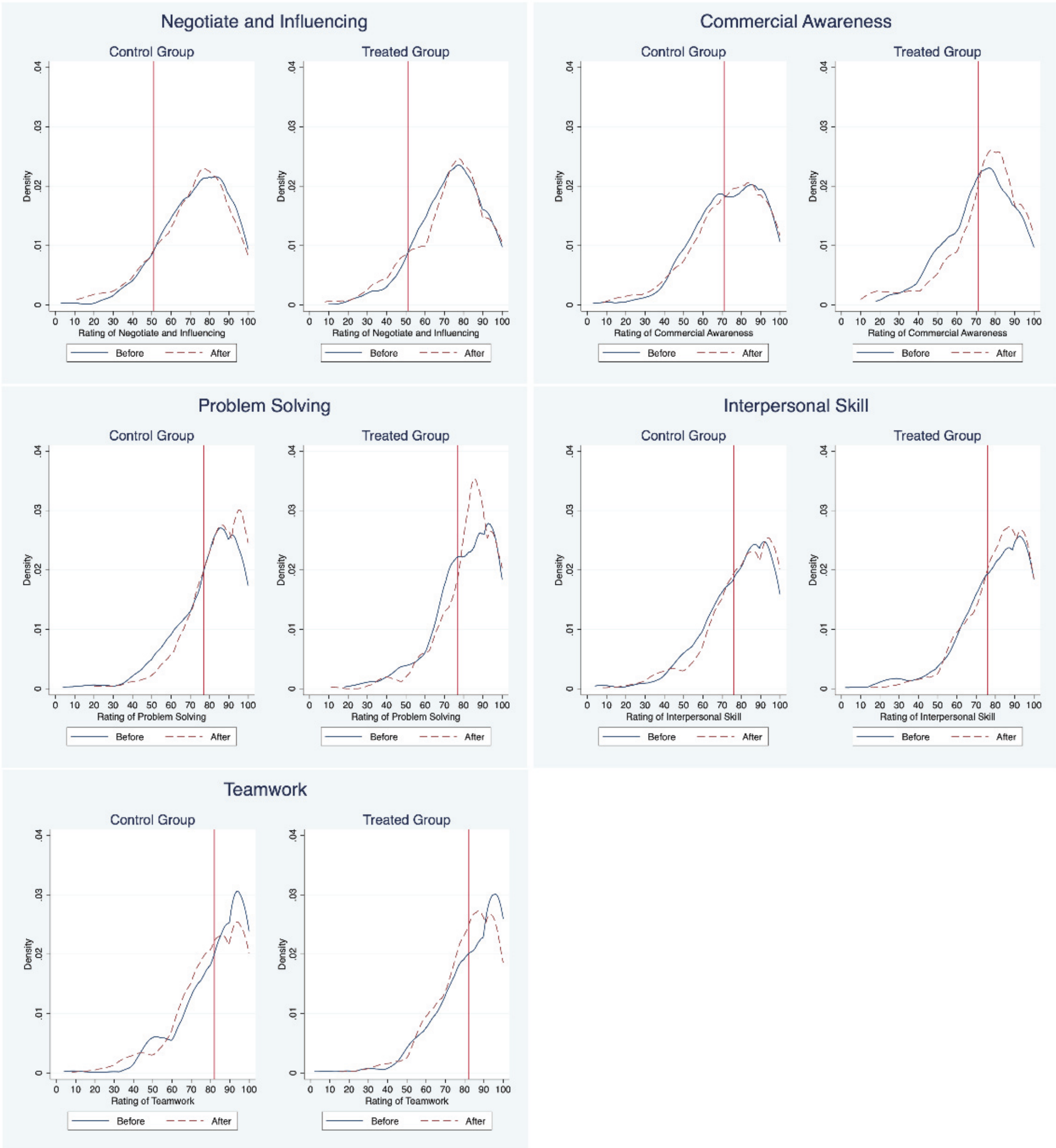
**Acquire these skills to distinguish yourself!**




<<

>>

## *A 1.2 Density of skills perception on employers' demand before and after intervention by treatment status*



## A 1.3 Treatment effects on academic inputs

| | (1) All | (2) Male | (3) Female |
|---|---|---|---|
| | *Panel A: Attendance (%)* | | |
| Treatment | -0.012 | -0.018 | -0.004 |
| | (0.011) | (0.017) | (0.015) |
| Baseline | 0.858*** | 0.801*** | 0.900*** |
| | (0.029) | (0.049) | (0.035) |
| $R^2$ | 0.643 | 0.650 | 0.640 |
| Observations | 769 | 331 | 438 |
| | *Panel B: Attendance (Hours per week)* | | |
| Treatment | -0.016 | -0.036 | 0.061 |
| | (0.069) | (0.112) | (0.084) |
| Baseline | 0.319*** | 0.340*** | 0.299*** |
| | (0.018) | (0.029) | (0.024) |
| $R^2$ | 0.534 | 0.606 | 0.530 |
| Observations | 768 | 330 | 438 |
| | *Panel C: Study Hours per week* | | |
| Treatment | -0.201 | 0.702 | -0.683 |
| | (0.840) | (1.111) | (1.274) |
| Baseline | 0.544*** | 0.576*** | 0.521*** |
| | (0.063) | (0.078) | (0.085) |
| $R^2$ | 0.315 | 0.432 | 0.275 |
| Observations | 595 | 247 | 348 |

*Notes:* All regressions include control for gender, ethnicity, mature student status, social-economic status, tuition fee status, department, tariff quintiles and whether receive intervention in the first and second year. Outcome variables in Panel A and B are obtained from administrative data, therefore, the estimation sample consists of those who attended intervention session in Wave 10. Estimation sample in Panel C consists of those who attended Wave 9, 10 and 11 as study hours were reported in the survey. Attendance (rate) is defined as the total number of attended events in third year over the total number of expected events in third year, while attendance (hours) is specified as average hours spent in attendance per week in third year. A universal cut-off point is applied to define baseline and post-treatment for all participants. For Panel A, everything in third year until the first day of the lab session (22/January/2018) was selected as the baseline. The attendance (hours) during the first term of the third year (October-December 2017) is used as the baseline for regressions in Panel B. The maximum value of study hours per week is 80, capped at 99% percentile. Robust standard errors in parentheses. * p<0.1, ** p<0.05, * p<0.01

## *A 1.4 Attrition*

| | (1) |
|---|---|
| | Probability of being in Wave 14 (Conditional on being in Wave 10) |
| Treatment | 0.013 |
| | (0.036) |
| Female | 0.095** |
| | (0.039) |
| $R^2$ | 0.083 |
| Observations | 770 |

*Notes:* Estimates are obtained from regressing the conditional probability of attending Wave 14 on treatment and other stratifying variables. Robust standard errors in parentheses. * p<0.1, ** p<0.05, * p<0.01

# Reference


Autor, D. H. (2015). Why Are There Still So Many Jobs? The History and Future of Workplace Automation. *Journal of Economic Perspectives*, *29*(3), 3–30. https://doi.org/10.1257/jep.29.3.3

Beaudry, P., Green, D. A., & Sand, B. M. (2016). The Great Reversal in the Demand for Skill and Cognitive Tasks. *Journal of Labor Economics*, *34*(S1), S199–S247. https://doi.org/10.1086/682347

Bettinger, E. P., Long, B. T., Oreopoulos, P., & Sanbonmatsu, L. (2012). The Role of Application Assistance and Information in College Decisions: Results from the H&R Block Fafsa Experiment*. *The Quarterly Journal of Economics*, *127*(3), 1205–1242. https://doi.org/10.1093/qje/qjs017

Biasi, B., & Sarsons, H. (2021). Information, Confidence, and the Gender Gap in Bargaining. *AEA Papers and Proceedings*, *111*, 174–178. https://doi.org/10.1257/pandp.20211019

Bleemer, Z., & Zafar, B. (2018). Intended college attendance: Evidence from an experiment on college returns and costs. *Journal of Public Economics*, *157*, 184–211. https://doi.org/10.1016/j.jpubeco.2017.11.002

Card, D., Cardoso, A. R., & Kline, P. (2016). Bargaining, Sorting, and the Gender Wage Gap: Quantifying the Impact of Firms on the Relative Pay of Women. *The Quarterly Journal of Economics*, *131*(2), 633–686.

Carrell, S. E., & Sacerdote, B. (n.d.). *Late Interventions Matter Too: The Case of College Coaching New Hampshire*.

Casner-Lotto, J., & Barrington, L. (2006). *Are They Really Ready to Work? Employers' Perspectives on the Basic Knowledge and Applied Skills of New Entrants to the 21st Century U.S. Workforce*. Partnership for 21st Century Skills. https://eric.ed.gov/?id=ED519465

Castex, G., & Kogan Dechter, E. (2014). The Changing Roles of Education and Ability in Wage Determination. *Journal of Labor Economics*, *32*(4), 685–710. https://doi.org/10.1086/676018

Coffman, K., Ugalde Araya, M. P., & Zafar, B. (2021). *A (Dynamic) Investigation of Stereotypes, Belief-Updating, and Behavior* [SSRN Scholarly Paper]. Social Science Research Network. https://papers.ssrn.com/abstract=3944430

Cook, P. J., Dodge, K., Farkas, G., Fryer, R. G., Guryan, J., Ludwig, J., Mayer, S. E., Pollack, H. A., & Steinberg, L. (2014). *The (Surprising) Efficacy of Academic and Behavioral*

*Intervention with Disadvantaged Youth: Results from a Randomized Experiment in Chicago* [SSRN Scholarly Paper]. Social Science Research Network. https://papers.ssrn.com/abstract=2386403

Croson, R., & Gneezy, U. (2009). Gender Differences in Preferences. *Journal of Economic Literature*, *47*(2), 448–474. https://doi.org/10.1257/jel.47.2.448

Delavande, A., Del Bono, E., & Holford, A. (2022). Academic and non-academic investments at university: The role of expectations, preferences and constraints. *Journal of Econometrics*, *231*(1), 74–97.

Deming, D. J. (2017). The Growing Importance of Social Skills in the Labor Market*. *The Quarterly Journal of Economics*, *132*(4), 1593–1640. https://doi.org/10.1093/qje/qjx022

Dittrich, M., Knabe, A., & Leipold, K. (2014). Gender Differences in Experimental Wage Negotiations. *Economic Inquiry*, *52*(2), 862–873. https://doi.org/10.1111/ecin.12060

Edin, P.-A., Fredriksson, P., Nybom, M., & Öckert, B. (2022). The Rising Return to Noncognitive Skill. *American Economic Journal: Applied Economics*, *14*(2), 78–100. https://doi.org/10.1257/app.20190199

Exley, C. L., Niederle, M., & Vesterlund, L. (2020). Knowing When to Ask: The Cost of Leaning In. *Journal of Political Economy*, *128*(3), 816–854. https://doi.org/10.1086/704616

Heckman, J. J., & Kautz, T. (2012). Hard evidence on soft skills. *Labour Economics*, *19*(4), 451–464. https://doi.org/10.1016/j.labeco.2012.05.014

Jackson, D., & Chapman, E. (2012). Non-technical skill gaps in Australian business graduates. *Education & Training*, *54*(2–3), 95–113. https://doi.org/10.1108/00400911211210224

Jensen, R. (2010). The (Perceived) Returns to Education and the Demand for Schooling*. *Quarterly Journal of Economics*, *125*(2), 515–548. https://doi.org/10.1162/qjec.2010.125.2.515

Lechner, M., & and Downward, P. (2017). Heterogeneous sports participation and labour market outcomes in England. *Applied Economics*, *49*(4), 335–348. https://doi.org/10.1080/00036846.2016.1197369

Machado, C. (2012). Selection, Heterogeneity and the Gender Wage Gap. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2177203

Manning, A., & Swaffield, J. (2008). The Gender Gap in Early-Career Wage Growth. *The Economic Journal*, *118*(530), 983–1024. https://doi.org/10.1111/j.1468-0297.2008.02158.x

Milner, S., Cousins, W., & McGowan, I. (2016). Does All Work and No Play Make a Dull Graduate? Perceptions of Extra-curricular Activities and Employability. *Journal of Perspectives in Applied Academic Practice*, *4*(1), Article 1. https://doi.org/10.14297/jpaap.v4i1.183

Mueller, G., & Plug, E. (2004). Estimating the Effect of Personality on Male-Female Earnings. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.576123

Nghia, T. L. H. (2017). Developing generic skills for students via extra-curricular activities in Vietnamese universities: Practices and influential factors. *Journal of Teaching and Learning for Graduate Employability*, *8*(1), 22–39.

Persico, N., Postlewaite, A., & Silverman, D. (2004). The Effect of Adolescent Experience on Labor Market Outcomes: The Case of Height. *Journal of Political Economy*, *112*(5), 1019–1053. https://doi.org/10.1086/422566

Reuben, E., Sapienza, P., & Zingales, L. (2015). Taste for Competition and the Gender Gap Among Young Business Professionals. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2677298

Roulin, N., & Bangerter, A. (2013). Students' use of extra-curricular activities for positional advantage in competitive job markets. *Journal of Education and Work*, *26*(1), 21–47. https://doi.org/10.1080/13639080.2011.623122

Saniter, N., & Siedler, T. (2014). Door Opener or Waste of Time? The Effects of Student Internships on Labor Market Outcomes. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2432425

Sarsons, H., & Xu, G. (2021). Confidence Men? Evidence on Confidence and Gender among Top Economists. *AEA Papers and Proceedings*, *111*, 65–68. https://doi.org/10.1257/pandp.20211086

Tchibozo, G. (2007). Extra-curricular activity and the transition from higher education to work: A survey of graduates in the United Kingdom. *Higher Education Quarterly*, *61*(1), 37–56.

Weisberg, Y. J., DeYoung, C. G., & Hirsh, J. B. (2011). Gender differences in personality across the ten aspects of the big five. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00178

# Chapter 2 — The Effect of Sibling Gender on Non-Cognitive Outcomes: Evidence from Australian Children

## 2.1 Introduction

There is a growing recognition of the importance of non-cognitive skills as fundamental determinants of life outcomes. Traits such as personality, motivation, social skills, behavioural patterns, and preference, collectively referred to as non-cognitive skills, are powerful predictors of social and emotional well-being as well as crucial economic outcomes such as education attainment and labour market performance (Heckman et al., 2006; Almlund et al., 2011; Heckman & Kautz, 2012). These non-cognitive traits, including perseverance, conscientiousness, self-control, and interpersonal skills, are not captured by conventional IQ or achievement tests. Instead, they are shaped dynamically during childhood through interactions within families, schools, and broader social environments. These early developmental experiences lay the foundation for future investments in skill acquisition.

In parallel, increasing attention has been paid to the role of gender exposure in shaping developmental outcomes, particularly in educational contexts. Research shows that girls attending single-sex schools outperform their co-educated counterparts in traditionally male-dominated subjects like mathematics and science (Eisenkopf et al., 2015; Lee & Lockheed, 1990). Evidence from randomised experiments suggests that single-sex schooling improves female students' mathematics performance by reducing gender-based stereotypical threat (Eisenkopf et al., 2015). However, comprehensive meta-analyses reveal mixed evidence. While some studies show modest advantages for single-sex education, particularly for girls in maths, controlled studies with rigorous design often find minimal or no significant benefits when selection effects are properly addressed (Pahlke et al., 2014). Whether single-sex schools improve educational performance for girls remains inconclusive, and the specific mechanisms through which exposure to the opposite gender influences child development are not well understood.

The household setting provides a natural and arguably more exogenous environment to investigate gender exposure effects through sibling composition. Unlike educational interventions, sibling gender is largely biologically determined and thus plausibly exogenous. Moreover, sibling interactions tend to be more emotionally intensive and developmentally influential than peer interactions at school. Research has found sibling correlations in non-cognitive traits ranging from 0.22 to 0.46, suggesting that a significant portion of variation in these traits arises from shared sibling influences (Anger & Schnitzlein, 2017). Thus, the sibling context is well-suited to detect and understand potential gender exposure effects, with important implications for developmental science and public policy.

This paper investigates whether exposure to different sibling gender (i.e., male vs. female) shapes children's socio-emotional development. Specifically, I examine whether having a brother, as opposed to a sister, influences the non-cognitive outcomes of 10-11-year-old children in Australia. This starting period of adolescence is particularly important as it marks a stage where behavioural patterns and social skills become more stable and predictive of adolescent and adult outcomes. Children between the ages of 10 and 19 are defined by the World Health Organization (WHO) as in the period of adolescence, where they experience rapid physical and cognitive development from the interaction with the world around them.

Measuring non-cognitive skills at scale is challenging. Standardised assessments of these traits are difficult to design due to variations in effort, motivation, and contextual factors. I use the Strengths and Difficulties Questionnaire (SDQ) as the measure of the non-cognitive outcomes of the study. Originally developed by Goodman (1997) as a clinical screening tool, SDQ has emerged as a reliable and widely adopted instrument for child behavioural and emotional well-being assessment. It aligns closely with constructs like self-regulation, emotional control, interpersonal skills, and behavioural adjustment, making it a practical alternative for investigating sibling effects in large-scale datasets. Its robust psychometric properties and cross-cultural applicability have led to increasing adoption in economics and education research.

The identification strategy relies on the assumption that sibling gender is conditionally exogenous—i.e., random given no parental gender-based fertility preferences. While gender is biologically determined, concerns remain that parental fertility decisions may depend on the gender of existing children. I address this issue through a comprehensive battery of balance tests to assess potential selection biases.

The literature on sibling effects has traditionally focused on family size and birth order. Although recent research has shifted attention toward sibling gender composition, traditional economic outcomes such as educational and labour market outcomes are more often studied. For example, Butcher & Case (1994) reported that women who grew up with only brothers attained more education, but sex composition had no effect on men, potentially due to differential treatment from compensatory parental investment. More recent studies offer a different perspective showing that women with brothers hold more traditional family and gender attitudes and beliefs and are more likely to pursue gender-stereotypical education and careers (Brenøe, 2022; Rao & Chatterjee, 2018). Similarly, Peter et al. (2018) find that women with brothers earn 10-15% less than those with only sisters, suggesting that sibling gender may shape long-term economic outcomes through early socialisation.

Meanwhile, empirical evidence on the effect of sibling gender on non-cognitive outcomes is limited and yields mixed findings. Two main theoretical frameworks offer compelling yet opposing predictions regarding the causal impact of sibling gender on these outcomes. According to social learning theory (Bandura, 1977), children assimilate attitudes and behavioural traits and emotional patterns from siblings through social interactions. Thus, having a sister is expected to show stronger traditional pro-social behaviours due to modelling more nurturing and cooperative behaviours (McHale et al., 2001). However, the theory of sibling differentiation supports sibling rivalry, suggesting that siblings develop contrasting characteristics, interests, and behaviours to differentiate themselves in families to reduce competition and attract resources (Schachter et al., 1978). In this case, having a brother is instead expected to develop better pro-social behaviours to differentiate from more competitive or aggressive tendencies. Despite ongoing debate, there is no clear consensus. While some empirical studies support one theory or the other, others find mixed or limited evidence for both.

Evidence from childhood development is nuanced. An early study by Cyron et al. (2016) documents beneficial effects for boys with sisters in early childhood. They find boys with sisters outperform those with brothers in kindergarten cognitive and non-cognitive domains, but these benefits quickly fade out by first grade. On the other hand, using data from the British Cohort Study, Golsteyn and Magnée (2020) show that boys with younger sisters scored higher on the Big Five personality traits of agreeableness reported by mother at the age of 10 years and 16 years, with the effect being more pronounced for age 16 years. A more recent and large-scale work by Dudek et al. (2022) finds no lasting effects of sibling gender composition on adult personality across nine countries. These imply that timing is crucial as sibling effects may emerge early, they may also diminish once broader peer environments take over.

Meanwhile, the personality is most malleable when children are young (Roberts & DelVecchio, 2000; Sutter et al., 2019). Behavioural patterns and social competencies established at the start of adolescence may be more stable than the early childhood effects documented by Cyron et al. (2016) but potentially more developmentally significant than the adult personality measures examined by Dudek et al (2022).

This paper aims to fill the gap by focusing on children aged 10-11, leveraging rich Australian longitudinal data. Using data from the Longitudinal Study of Australian Children (LSAC), I find that having a brother reduces hyperactivity and inattention problems compared to having a sister. This effect is particularly pronounced among second-born children and is evident across both boys and girls, suggesting a gender-specific mechanism potentially related to the types of play and interaction that brothers facilitate. Girls with brothers document a small but significant increase in the pro-social domain compared to those with sisters, providing strong support for sibling differentiation theory. Second-born children with brothers are reported with a lower total difficulties score, showing a trend toward fewer total behavioural problems. Other SDQ dimensions suggest limited evidence of effects.

This paper contributes to both the sibling composition literature and the non-cognitive development literature. First, it extends the sibling composition literature by providing causal evidence that exposure to a brother, rather than a sister, shapes children's socio-emotional skills during early adolescence. While existing work has largely focused on long-run educational, attitudinal, and labour market outcomes, this study identifies non-cognitive skills as an early-emerging mechanism through which sibling gender composition influences later life outcomes. These findings bridge a gap between economic research on sibling gender and psychological research on socio-emotional development.

Second, the paper advances the non-cognitive development literature by demonstrating that the effects of sibling gender are heterogeneous by birth order, with second-born children driving the results. This highlights the importance of within-family dynamics and shows that family structure plays a meaningful role in shaping non-cognitive traits such as hyperactivity/inattention and pro-social behaviour. This evidence strengthens theoretical accounts of skill formation that emphasise social learning and sibling interactions as key developmental inputs.

The remainder of this paper proceeds as follows. Section 2 describes the LSAC data and presents descriptive statistics. Section 3 outlines the empirical methodology and identification strategy, including tests of the exogeneity assumption. Section 4 presents the main results across SDQ domains. Section 5 explores heterogeneity by child gender and birth order, investigates findings across informants, and discusses mechanisms. Section 6 concludes with implications for policy and future research.

## 2.2 Data: The Longitudinal Study of Australian Children (LSAC)

To understand the effect of sibling's gender on the development of non-cognitive skills, the study examined a group of Australian children using a national survey data called the Longitudinal Study of Australian Children (LSAC). LSAC commenced in 2004 as the first comprehensive national dataset that follows the development and wellbeing of Australian children as they grow up. Two cohorts of Australian children and their families and teachers were surveyed every two years on topics of demographics, home environment, health, education, parenting, cognitive outcomes, emotional outcomes and etc. This dataset aimed at following these kids from early childhood into adulthood and capturing long-term developmental, health and educational outcomes. In the first wave of LSAC in 2004, two cohorts were established. The infant cohort was referred as B-cohort, including children born from March 2003 to February 2004, who were aged 0-1 years old at the time. The other K-cohort interviewed were selected from children born between March 1999 and February 2000, who were aged 4-5 years old in 2004. Children from both cohorts were representative samples of children of these ages across Australia at that time. There was a total of over 10,000 children and their families participated in the first wave.

### 2.2.1 Analytical Sample

Table 2.1 summarises key demographics, school types, household composition, and parental characteristics of the study children, based on data from Wave 4 (K cohort) and Wave 6 (B cohort) of the Longitudinal Study of Australian Children (LSAC). Column (1) presents summary statistics for the full sample of 7,779 children aged approximately 10–11 years, with data collected in 2010 and 2014.

Given the focus on how exposure to a sibling's gender affects non-cognitive development, households without any siblings (approximately 10% of the LSAC sample) were excluded. To isolate the effect of having a sibling of a specific gender, the sample was restricted to study children with exactly one sibling. In this study, siblings are defined as those residing in the same household, regardless of biological relatedness. Nevertheless, about 94% of the study children with one sibling had a full biological sibling. As a result, the analytical sample includes 3,447 children aged 10–11 who reported living with exactly one sibling.

Column (2) reports the summary statistics for this analytical sample, while Columns (3)–(5) compare children based on their sibling's gender (i.e., having a brother vs. a sister) and conduct balance checks across demographic and household characteristics. Girls make up 49.1% of the sample, and the average age is approximately 10.4 years. Over 60% of the children attend government schools, followed by Catholic and private schools. There are no significant differences in school attendance by sibling gender.

Both K cohort (53%) and B cohort (47%) had approximately even contribution to the sample. Conditional on having one sibling, the distribution of sibling gender is roughly even: 50.5% of children have a brother (N = 1,740) and 49.5% have a sister (N = 1,707). About half of the study children reported having a younger sibling living in the household. Among these sibling pairs, about half are full biological siblings, with approximately 94% having younger siblings. The proportion of the children who were living with their biological mother is reported around 98% across groups. Nearly all study children (99.5%)

live with at least one of their biological parents, while the proportion of study children whose parents are reported as married or in a legally recognised de facto relationship in the LSAC is slightly lower at 85%. Roughly 90% of study children reported English as their 4main language spoken at home. These variables demonstrate excellent balance between two sibling-gender group, suggesting no significant differences in household composition.

*Table 2.1 Summary statistics of the study children (SC) and balance checks*

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | All | With one sibling | Has brother | Has sister | p |
| K cohort (2010) | 0.529 | 0.530 | 0.529 | 0.531 | 0.878 |
| Girls | 0.487 | 0.491 | 0.512 | 0.469 | 0.011** |
| Age | 10.365 | 10.366 | 10.360 | 10.372 | 0.456 |
| Government school | 0.626 | 0.612 | 0.603 | 0.620 | 0.308 |
| Catholic school | 0.217 | 0.221 | 0.229 | 0.212 | 0.222 |
| Private school | 0.151 | 0.164 | 0.166 | 0.163 | 0.833 |
| Not attending school | 0.006 | 0.003 | 0.002 | 0.005 | 0.125 |
| *Household* | | | | | |
| #. people in household | 4.566 | 3.950 | 3.958 | 3.943 | 0.386 |
| #. of siblings of SC | 1.615 | | | | |
| Has ≥ one sibling(s) | 0.916 | | | | |
| Brother | 0.661 | 0.505 | | | |
| Sister | 0.638 | 0.495 | | | |
| Biological sibling | 0.875 | 0.939 | 0.940 | 0.937 | 0.721 |
| Younger sibling | 0.548 | 0.480 | 0.490 | 0.469 | 0.205 |
| Live with biological mother | 0.979 | 0.983 | 0.983 | 0.983 | 0.941 |
| Live with ≥ one biological parent(s) | 0.994 | 0.995 | 0.997 | 0.994 | 0.211 |
| English as SC main language at home | 0.906 | 0.898 | 0.896 | 0.900 | 0.709 |
| Parents married/legally de facto | 0.843 | 0.851 | 0.858 | 0.844 | 0.234 |
| *Parents* | | | | | |
| Both ANZ-born | 0.712 | 0.697 | 0.704 | 0.690 | 0.406 |
| Both Overseas-born (non-ANZ) | 0.110 | 0.119 | 0.119 | 0.118 | 0.962 |
| Mixed-origin (ANZ + non-ANZ) | 0.178 | 0.184 | 0.177 | 0.192 | 0.305 |
| 2 × BA+: Bachelor degree or above for both | 0.321 | 0.326 | 0.312 | 0.340 | 0.174 |
| 1 × BA+: Mixed qualification | 0.323 | 0.339 | 0.357 | 0.320 | 0.077* |
| 0 × BA+: Neither holds a bachelor degree or higher | 0.356 | 0.335 | 0.331 | 0.340 | 0.673 |
| 2 × Employed | 0.753 | 0.821 | 0.828 | 0.814 | 0.336 |
| 1 × Employed | 0.217 | 0.160 | 0.152 | 0.168 | 0.248 |
| 0 × Employed | 0.029 | 0.019 | 0.020 | 0.018 | 0.692 |
| Weekly household income < $1500 | 0.301 | 0.277 | 0.274 | 0.28 | 0.672 |
| Weekly household income [$1500, $2500) | 0.330 | 0.338 | 0.338 | 0.337 | 0.975 |
| Weekly household income ≥ $2500 | 0.317 | 0.344 | 0.345 | 0.343 | 0.896 |
| ***N*** | 7779 | 3447 | 1740 | 1707 | 3447 |

*Notes:* The table reports the sample mean of the study childern's demographics, school type, household composition and parents characteristics. The full sample in column (1) consists of children aged 10–11 from the Longitudinal Study of Australian Children (LSAC). Column (2) presents the analytical sample of study children with only one sibling in household. Column (5) reports p-values from two-sided t-tests comparing those with a brother vs. sister among children with exactly one sibling. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. All variables are binary unless otherwise indicated. $N$ indicates the number of non-missing observations.

It is important to note that the parental characteristics recorded in LSAC do not always reflect those of the biological parents. Instead, Parent 1 is defined as the primary caregiver with the most knowledge of the child's life, while Parent 2 is their partner or another adult fulfilling a parental role. In most cases, Parent 1 and Parent 2 are the child's biological mother and father. About 70% of children have both parents born in Australia or New Zealand, 12% have both parents born overseas, and 18% come from mixed-origin households. This largely corresponds with the high rate of English being spoken at home.

It is important to note that the parental characteristics recorded in LSAC do not always reflect those of the biological parents. Instead, Parent 1 is defined as the primary caregiver with the most knowledge of the child's life, while Parent 2 is their partner or another adult fulfilling a parental role. In most cases, Parent 1 and Parent 2 are the child's biological mother and father. About 70% of children have both parents born in Australia or New Zealand, 12% have both parents born overseas, and 18% come from mixed-origin households. This largely corresponds with the high rate of English being spoken at home.

Parental education is relatively evenly distributed: roughly one-third of couples both held at least a bachelor's degree, one-third neither held one, and the remaining third had mixed qualifications (one parent with a degree, one without). Over 80% of families report dual employment. Fewer than 2% of children live in households where both parents are unemployed. Household weekly income is evenly distributed across three brackets: less than \$1,500, \$1,500–\$2,500, and \$2,500 or more, with no significant difference across sibling-gender groups.

P-values in Column (5) of Table 2.1 confirm that almost all parental characteristics, such as country of birth, employment status, and income, are balanced across groups. One marginal imbalance is observed in the share of mixed-education couples ($p = 0.077$), which is not unexpected given the multiple comparisons and does not pose a major concern.

However, a statistically significant imbalance is observed in the gender of the study child: among those with exactly one sibling, more girls are found in the "has brother" group (51.2%) than in the "has sister" group (46.9%), with a p-value of 0.011. This suggests that mixed-gender sibling pairs are more common than same-gender pairs among two-child families. This pattern may raise concerns about selection if sibling gender correlates with the gender of the first-born and thereby influences parental decisions about having a second child. A key identifying assumption in this study is that the gender of the sibling is as good as randomly assigned.

Although sibling gender is determined biologically and is random at birth, this exogeneity assumption warrants scrutiny. Specifically, for a causal interpretation, differences in observed outcomes should stem solely from sibling gender, not from correlated parental preferences or childbearing decisions. For example, if parents with a strong preference for sons are more likely to have a second child when the first-born is a girl, such selection could confound the relationship between sibling gender and non-cognitive outcomes. To address this, Table 2.1 shows that aside from the study child's own sex, observable characteristics are generally balanced between children with a brother and those with a sister. Additional robustness checks are provided in Section 3 to confirm that this imbalance does not bias the main results, thereby supporting the validity of the exogeneity assumption.

### 2.2.2 Measure of Outcomes: The Strengths and Difficulties Questionnaire (SDQ)

In this study, the Strengths and Difficulties Questionnaire (SDQ) is chosen as a psychometrically sound and routinely used instrument to assess the non-cognitive aspects of the study children. Originally designed by Goodman (1997) as a clinical screening tool for psychiatric disorders in children and adolescents aged 4–16, the SDQ has demonstrated consistent reliability and validity across various populations.

The SDQ consists of 25 question items divided into five scales (key attributes), with every five items classified into one scale as follows: pro-social behaviour, hyperactivity/inattention, emotional symptoms, peer relationship problems, and conduct problems. Pro-social is the only attribute directly measuring strength, which assesses children's empathy, cooperation, and willingness to help. The remaining four attributes focus on difficulties, assessing hyperactivity and inattention behaviours, emotional symptoms such as anxiety and fear, peer relationship problems and social difficulties, and conduct problems such as rule-breaking and aggression. Details can be found in Appendix 2 Items of the Strengths and Difficulties Questionnaire (SDQ).

Each item is a brief behavioural statement describing either a positive or problematic trait or behaviour of the study children. Parents, teachers, or the children themselves (if aged 11 and above) serve as the respondent to rate how well each statement applies, using a three-point Likert scale (Not True, Somewhat True, Certainly True). The score for each attribute is calculated by summing the responses to the five corresponding items. Provided at least three of the five items are completed, each key attribute score is converted to a 0–10 numerical scale, with higher scores generally reflecting greater difficulty (except for the pro-social attribute, where higher scores indicate more positive behaviour.[13] A composite score—Total Difficulties—is calculated by summing the four difficulty-focused subscales, yielding a score from 0 to 40.

The SDQ was administered in each LSAC wave: from Wave 2 onward for the B-cohort and from Wave 1 for the K-cohort. Teachers completed the SDQ after children entered school. Respondents include Parent 1 (the primary caregiver), Parent 2 (the partner or second guardian), the child's teacher and the child if eligible. In this study, Parent 1's responses are the primary outcome, as they are the most familiar with the child's daily life. Nevertheless, ratings from Parent 2 and teachers provide valuable complementary perspectives.[14] Teachers, for example, observe children in structured social settings where peer interactions and behavioural regulation are salient.

#### *2.2.2.1 SDQ Scores Between Informants*

Table 2.2 reports the mean SDQ scores and standard deviations across the three informants. Sample sizes for reported outcomes differ by informant type, with Parent 1 (N = 3,447), Parent 2 (N = 2,341), and Teachers (N = 2,789). T-tests compare mean scores between Parent 1 and the other informants.

Compared to Parent 2, Parent 1 reports the highest average SDQ scores for pro-social behaviour (8.602), followed by Parent 2 (8.385) and teachers (7.858). The differences between Parent 1 and Parent 2 ratings and between Parent 1 and teacher ratings are statistically significant at the 1% level. The average scores for pro-social behaviour are

[13] A score can still be computed by proportionally scaling the total of the available items if up to two items are missing. However, if fewer than three items are completed, the score of the attribute is treated as missing.

[14] The response rates of SDQ scores in Wave 4K and Wave 6B of the LSAC were 98%, 84% and 82% for Parent 1, Parent 2, and Teacher respectively.

high across all informants, reflecting strong social, emotional, and moral development among the study children—crucial components of non-cognitive traits such as interpersonal and empathy-related skills.

For attributes framed as difficulties, Parent 1-reported SDQ scores are mostly aligned with Parent 2 ratings, except for emotional problems. Parent 1 reports slightly higher mean scores (1.898) than Parent 2 (1.592), with this discrepancy being statistically significant ($p < 0.01$). This pattern extends to the composite Total Difficulties Score, reflecting a higher level of overall behavioural or emotional issues reported by Parent 1.

There are no statistically significant differences between the two parent informants for hyperactivity/inattention, peer problems, and conduct problems, suggesting consistency in parental perceptions. The higher pro-social score indicates that Parent 1 perceives children as more considerate of others' feelings, helpful, kind, and empathetic compared to Parent 2's assessments. Meanwhile, the higher emotional problems score suggests Parent 1 more frequently reports signs of emotional distress, such as clingy, unhappy, fearful, or anxiety-prone behaviour.

*Table 2.2 Average Strengths and Difficulties Questionnaire (SDQ) scores reported by various informants*

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | Parent 1 | Parent 2 | p | Teacher | p |
| Pro-social | 8.602 | 8.385 | 0.000*** | 7.858 | 0.000*** |
| | (1.623) | (1.668) | | (2.194) | |
| Hyperactivity/Inattention | 3.160 | 3.104 | 0.360 | 2.398 | 0.000*** |
| | (2.386) | (2.217) | | (2.635) | |
| Emotional symptoms | 1.898 | 1.592 | 0.000*** | 1.192 | 0.000*** |
| | (1.932) | (1.748) | | (1.782) | |
| Peer problems | 1.432 | 1.405 | 0.534 | 1.262 | 0.000*** |
| | (1.649) | (1.590) | | (1.725) | |
| Conduct problems | 1.155 | 1.097 | 0.108 | 0.680 | 0.000*** |
| | (1.381) | (1.317) | | (1.383) | |
| *Total difficulties score* | 7.645 | 7.199 | 0.001*** | 5.531 | 0.000*** |
| | (5.346) | (4.908) | | (5.637) | |
| ***N*** | 3447 | 2341 | | 2789 | |

*Notes:* Each cell displys the sample mean of SDQ scores with the standard deviation in parentheses below. SDQ consists of 25 items with every 5 items being classified into a scale. The SDQ score ranges from 0 to 10 for each of the 5 scales if at least 3 items were completed for the scale. Pro-social is the only scale for strengths; the Total Difficulties Score are summed from the remaining four scales. Columns (1) and (2) show parent-reported SDQ scores for Parent 1 and Parent 2, respectively. Parent 1 is defined as the parent who knows the study child best. Parent 2 is defined as Parent 1's partner or another adult in the home with a parental relationship to the study child. Column (3) presents p-values from two-sided t-tests comparing the means of Parent 1 and Parent 2 reports. Column (4) reports SDQ scores as assessed by teachers. Column (5) presents p-values from two-sided t-tests comparing the means of Parent 1 and Teacher reports. *** p < 0.01, ** p < 0.05, * p < 0.1.

These discrepancies likely reflect contextual differences and familiarity. Parent 1's role, often filled by the biological mother, allows for a more comprehensive understanding of the child's behaviour across settings, making their ratings more sensitive to internalising issues such as anxiety or peer struggles.

Teacher-reported scores provide an important external perspective, with notable differences in means compared to parent reports. Average SDQ scores are consistently lower across all subscales, suggesting fewer observed behavioural difficulties in school environments. This could be because some issues are less visible in school settings due to stricter institutional norms.

Both informant-specific perceptions and context-specific behavioural expressions of children contribute to these discrepancies in how behavioural and emotional difficulties are reported by different informants. Among the four difficulty attributes, hyperactivity/inattention consistently receives the highest mean scores regardless of informant type, indicating this as the most prevalent challenge that is most easily detected by both parents and teachers.

While these difficulties may manifest differently across environments, caregiving roles are crucial in shaping perceptions of non-cognitive traits. Parent 1, as the primary caregiver (typically the biological mother or someone highly engaged in the child's life), is most aware of daily behaviours. This role enables them to make assessments based on the complete scope of behaviours, providing more reliable SDQ ratings of children overall.

#### *2.2.2.2 Standardised SDQ Scores*

To facilitate interpretation and cross-group comparisons, all SDQ subscale scores are standardised (mean = 0, SD = 1) across the full analytical sample. Table 2.3 presents standardised scores by gender and birth order.

Girls are rated significantly higher on pro-social behaviour than boys (0.182 vs. -0.175, $p < 0.01$), with higher pro-social scores reflecting stronger interpersonal skills and empathetic traits. Girls also receive higher average scores on emotional symptoms, suggesting greater propensity toward internalising difficulties and emotional distress, including sadness and anxiety. The mean score difference in emotional symptoms is statistically significant (0.064 vs. -0.062, $p < 0.01$), despite being small in magnitude.

Conversely, boys score higher in in hyperactivity/inattention, peer problems, and conduct problems ($p < 0.01$), resulting in higher Total Difficulties scores (0.113 vs. -0.117, $p < 0.01$). This pattern aligns with established findings in psychological and developmental studies showing that girls are more prone to internalising problems due to biological, hormonal, and social factors, while boys are more likely to externalise difficulties through hyperactivity/inattention and conduct issues.

*Table 2.3 Standardised Strengths and Difficulties Questionnaire (SDQ) from Parent 1 by gender & birth order*

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| | All | Boy | Girl | p | First | Second | p |
| Pro-social | 0.000 | -0.175 | 0.182 | 0.000*** | 0.109 | -0.118 | 0.000*** |
| Hyperactivity | 0.000 | 0.197 | -0.205 | 0.000*** | 0.022 | -0.024 | 0.020* |
| Emotional problems | 0.000 | -0.062 | 0.064 | 0.000*** | -0.062 | 0.067 | 0.169 |
| Peer problems | 0.000 | 0.099 | -0.103 | 0.000*** | -0.066 | 0.072 | 0.000*** |
| Conduct problems | 0.000 | 0.064 | -0.066 | 0.000*** | -0.020 | 0.022 | 0.222 |
| Total Difficulties Score | 0.000 | 0.113 | -0.117 | 0.001*** | -0.038 | 0.041 | 0.020* |
| ***N*** | 3447 | 1756 | 1691 | | 1794 | 1653 | |

*Notes:* Standardised mean of SDQ scores based on parent-reported data (Parent 1) are shown, with the overall mean for each SDQ scale normalised to zero. Positive values indicate above-average scores, and negative values indicate below-average scores relative to the full analytical sample. Pro-social is the only positively framed scale, while the other subscales (Hyperactivity, Emotional Problems, Peer Problems, and Conduct Problems) contribute to the Total Difficulties Score. Columns (2)–(4) compare boys and girls, while Columns (5)–(7) compare first-born and second-born children. P-values are derived from two-sided t-tests for mean differences. *** p < 0.01, ** p < 0.05, * p < 0.1.

Columns (5)–(7) present comparisons of average standardised SDQ scores between first-born and second-born children, revealing heterogeneity by birth order. First-born children receive significantly higher scores on pro-social behaviour (0.109 vs. -0.118) and more favourable ratings (lower scores) on peer problems (-0.066 vs. 0.072), with both differences significant at the 1% level. This advantage reverses slightly for hyperactivity/inattention scores, where second-born children show marginally better outcomes with significantly lower scores (-0.024) at the 5% level. However, emotional problems and conduct problems scores do not differ significantly by birth order. Consequently, the Total Difficulties score is slightly lower for first-born children (-0.038 vs. 0.041, $p = 0.02$), although the effect size is small.

In summary, the SDQ effectively captures dimensions of non-cognitive development. The descriptive evidence confirms with established developmental patterns by gender, with girls scoring higher in pro-social attributes and showing more signs of emotional distress, while boys manifest more behavioural difficulties. Birth order differences are more subtle, with some evidence that second-born children experience slightly more behavioural difficulties, though not across all sub-domains. Parent 1's responses are used as the main outcome variables, supplemented by ratings from Parent 2 and Teachers to triangulate behavioural assessments.

## 2.3 Empirical Strategy

### 2.3.1 Model Specification

To estimate the effect of sibling gender—specifically, having a brother versus a sister at home—on children's non-cognitive outcomes, I employ the following linear specification estimated separately for each outcome $k \in \{$ pro-social, hyperactivity/inattention, emotional symptoms, peer problems, conduct problems, total difficulties$\}$:

$$y_i^k = \alpha^k + \beta^k BrotherAtHome_i + \delta^k Sex_i + \vartheta^k BirthOrder_i + \boldsymbol{X}_i{}'\gamma^k + \varepsilon_i^k$$

Here, $y_i^k$ represents the standardised Strengths and Difficulties Questionnaire (SDQ) score for child $i$ on outcome $k$. Higher scores on pro-social behaviour indicate greater social adjustment (e.g. more considerate, empathetic, and helpful), while higher scores on the remaining subscales—including Total Difficulties—reflect greater behavioural or emotional difficulties.

The key treatment variable $BrotherAtHome_i$ equals 1 if child $i$ has a male sibling residing at home, and 0 if the sibling is female.[15] $Sex_i$ is a binary indicator for the child's

[15] "Brother at home" is defined as having a male sibling who co-resides in the same household as the study child at the time of the survey. In LSAC, biological siblings almost universally live together during early childhood, with 94% of co-residing siblings being biological, and cases of biological siblings living apart are extremely rare and usually arise only under exceptional circumstances (e.g., parental separation with split custody or kinship

own gender (1 = girl, 0 = boy). $BirthOrder_i$ is determined by whether the child has a younger sibling residing in the household. $\boldsymbol{X_i}$ is a vector of covariates including demographic information, school type, household structure, and family characteristics. These include indicators for cohort year, school sector, full biological sibling status, presence of biological parents, and household income.

Parent 1's SDQ ratings serve as the primary outcome measures. Parent 1 refers to the caregiver most familiar with the child's daily life and is not necessarily the biological mother. Accordingly, $\boldsymbol{X_i}$ also includes Parent 1-specific controls: country of birth, employment status, educational attainment, household income, and an indicator for whether Parent 1 is the biological mother. These variables help account for heterogeneity in caregiving roles. If $BrotherAtHome_i$ is as-good-as-random, then $\beta^k$ identifies the average causal effect of having a brother versus a sister on outcome $\boldsymbol{k}$.

## 2.3.2 Identification Checks: Exogeneity of Sibling Gender

A critical assumption in this analysis is that, within two-child families, sibling gender is quasi-random. While a child's gender is biologically determined at conception, the validity of this assumption depends on the absence of parental selection into sibling gender composition. That is, if the decision to have a second child depends on the gender of the first, the exogeneity assumption may be violated. For example, parents with a particular preference may systematically end up with mixed-gender pairs, in which a family with a first-born girl is systematically different from those with a first-born boy, and those differences also affect SDQ scores.

Table 2.1 already shows that household and parental characteristics are well-balanced across sibling-gender groups. However, the gender of the study children is significantly associated with the gender of the sibling: study children with a brother are more likely to be girls. This imbalance—about four percentage points ($p$ = 0.011)—

care). To address the small number of cases where co-residence might reflect family dynamics, I include controls for family structure and parental characteristics.

deviates from what pure randomness would predict. If sibling genders were independent, the proportion of mixed-gender and same-gender pairs would each be approximately 50%.

Table 2.4 investigates potential selection into sibling gender composition by conditioning on the gender of the first-born child. The logic here is straightforward: If selection were present, we would expect observable differences between female first-born (F1) and male first-born (M1) families. Columns (1)–(3) of Table 2.4 report balance tests within F1 and M1 families. Sibling type (biological or younger), parental education levels, employment status, household income, and other parental and household characteristics are all statistically indistinguishable. One minor imbalance is observed in Parent 1's country of birth, with F1 families more likely to have Parent 1 born in Australia or New Zealand (difference = 2.5 p.p., $p$ = 0.083).

However, substantial differences in the gender composition of study children and their siblings are observed: in F1 families, 73% of study children are girls; in M1 families, only 26% are girls. Similarly, the likelihood of having a sister is 74% among F1 families versus 26% in M1 families. Instead of signalling selection, these imbalances reflect arithmetic constraints when fixing the first-born's gender and then drawing one child from a two-child household. For instance, in two-child families where the first-born is female, the study child can only be a girl if the sibling pairing is either GG or GB. Since each of the four sibling pairings (GG, GB, BG, BB) is equally likely under random assignment, the conditional probability of the second child being a girl in an F1 family is 75%. The observed ratios (e.g. 0.73/0.26) align closely with this benchmark. This suggests that when the gender of the first-born is accounted for, the gender distribution of the second-born child in the sample roughly mirrors a random biological process, indicating weak evidence of mixed-gender stopping.

Moreover, even in the absence of selection, there is a mechanical correlation between own gender and sibling gender in two-child households. For any two-child families A girl can only have a brother if in a mixed-gender pair (GB), and vice versa. Because mixed-gender pairs contribute one girl to the "brother" group and one boy to the "sister" group, the gender of the study children is not independent of the sibling's gender. As a result, girls comprise exactly 50% of the "has-brother" group and less than 50% of the "has-sister" group, not due to selection but due to simple combinatorics.

Hence, the observed imbalance in the gender of study children by sibling gender in Table 2.1 is consistent with random gender assignment under known arithmetic constraints and is not necessarily indicative of selection into sibling gender. Table 2.4 supports this by showing no systematic parental selection based on first-born gender. Together, these findings lend strong support to the assumption that sibling gender is quasi-random, validating its use as an exogenous treatment in this study.

*Table 2.4 Additional balance checks for selection into sibling's gender by the parents: by first-born gender*

| | (1) | (2) | (3) |
|---|---|---|---|
| | Gender of the first-born child of SC's family | | |
| | Female (F1) | Male (M1) | p |
| K cohort (2010) | 0.526 | 0.536 | 0.555 |
| Girls | 0.728 | 0.259 | 0.000*** |
| Age | 10.370 | 10.361 | 0.588 |
| Government school | 0.613 | 0.605 | 0.649 |
| Catholic school | 0.217 | 0.227 | 0.484 |
| Private school | 0.166 | 0.165 | 0.973 |
| Not attending school | 0.004 | 0.002 | 0.329 |
| ***Household*** | | | |
| #. people in household | 3.943 | 3.961 | 0.330 |
| Brother | 0.260 | 0.744 | 0.000*** |
| Sister | 0.740 | 0.256 | 0.000*** |
| Biological sibling | 0.941 | 0.934 | 0.386 |
| Younger sibling | 0.478 | 0.502 | 0.155 |
| Live with biological mother | 0.984 | 0.983 | 0.714 |
| Live with ≥ one biological parent(s) | 0.995 | 0.996 | 0.555 |
| English as SC main language at home | 0.898 | 0.897 | 0.949 |
| Parents married/legally de facto | 0.847 | 0.857 | 0.427 |
| ***Parents*** | | | |
| Both ANZ-born | 0.708 | 0.693 | 0.400 |
| Both Overseas-born (non-ANZ) | 0.109 | 0.125 | 0.174 |
| Mixed-origin (ANZ + non-ANZ) | 0.183 | 0.181 | 0.894 |
| 2 × BA+: Bachelor degree or above for both | 0.341 | 0.309 | 0.119 |
| 1 × BA+: Mixed qualification | 0.324 | 0.350 | 0.222 |
| 0 × BA+: Neither holds a bachelor degree or higher | 0.335 | 0.342 | 0.748 |
| 2 × Employed | 0.821 | 0.821 | 0.986 |
| 1 × Employed | 0.160 | 0.160 | 0.969 |
| 0 × Employed | 0.019 | 0.019 | 0.957 |
| Weekly household income < $1500 | 0.275 | 0.277 | 0.916 |
| Weekly household income [$1500, $2500) | 0.337 | 0.339 | 0.914 |
| Weekly household income ≥ $2500 | 0.346 | 0.343 | 0.842 |
| *Parent who has the most knowledge of SC (Parent 1)* | | | |
| Age | 42.113 | 41.850 | 0.133 |
| Indigenous | 0.012 | 0.010 | 0.642 |
| Female | 0.962 | 0.958 | 0.603 |
| Biological parent | 0.993 | 0.996 | 0.214 |
| Biological mother | 0.955 | 0.955 | 0.965 |
| ANZ born | 0.809 | 0.786 | 0.083* |
| BA+ | 0.387 | 0.387 | 0.966 |
| Employed | 0.852 | 0.837 | 0.236 |
| Usual weekly income less than $1000 | 0.305 | 0.302 | 0.893 |

| | | | |
|---|---|---|---|
| *Parent 1's partner or another adult fulfilling a parental role (Parent 2)* | | | |
| Age | 44.359 | 43.995 | 0.104 |
| Indigenous | 0.009 | 0.008 | 0.736 |
| Female | 0.032 | 0.034 | 0.735 |
| Biological parent | 0.799 | 0.792 | 0.601 |
| Biological father | 0.771 | 0.765 | 0.680 |
| ANZ born | 0.676 | 0.676 | 0.999 |
| BA+ | 0.302 | 0.289 | 0.434 |
| Employed | 0.946 | 0.952 | 0.484 |
| Usual weekly income less than $1000 | 0.704 | 0.702 | 0.911 |
| | | | |
| ***N*** | 1648 | 1725 | 3373 |

*Notes:* This table reports sample means for study children (SC) with exactly one sibling, drawn from Wave 4 of the K cohort and Wave 6 of the B cohort. Columns (1)–(2) compare SC demographics, school type, household composition, and parental characteristics between families whose first-born child is a girl (F1) and those whose first-born is a boy (M1). Column (3) reports two-sided p-values for testing the equality of group means. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Girls = 1 if the SC is female; Biological sibling = 1 if the sibling is fully biologically related to the SC; Live with ≥1 biological parent(s) = 1 if at least one biological parent resides in the household. BA+ = 1 if the highest qualification attained by a parent is at least a bachelor's degree. All variables are binary unless otherwise indicated. Approximately 2% of the analytical sample (N = 74) report the sibling as a same-age twin, resulting in a slightly smaller number of non-missing observations than the earlier N = 3447.

# 2.4 Results

This section presents the empirical findings from examining how sibling gender affects child behavioural and emotional outcomes, using six scales from the Strengths and Difficulties Questionnaire (SDQ). Parent 1-reported scores are used as the main outcomes, while SDQ scores reported by other informants (Parent 2, teachers) are also compared and discussed.

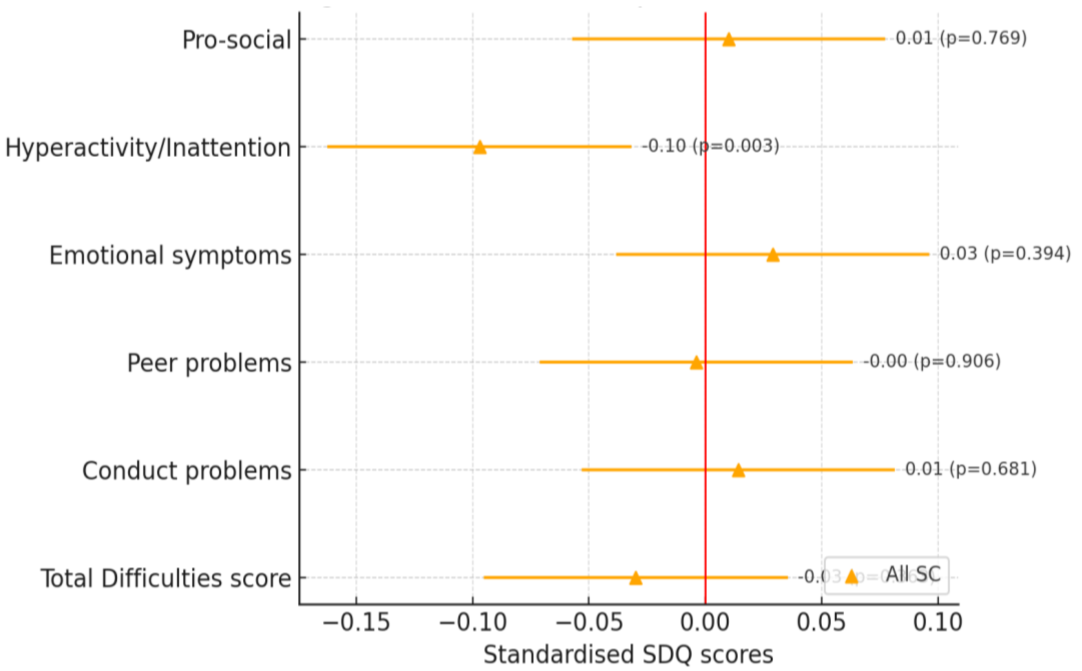
## 2.4.1 Average Effect on SDQ Scores

Figure 2.1 plots the average treatment effect of having a brother on each SDQ scale for the full analytical sample of study children with exactly one sibling. Estimates are derived from regressions specified in Section 3.1, including controls for child, household, and parents' characteristics. The results suggest no statistically significant differences in pro-social behaviours, emotional symptoms, conduct problems, and total difficulties between children having a brother at home versus a sister. However, a small but significant negative effect is observed on hyperactivity/inattention (-0.1 standard deviations, $p = 0.003$). A negative value indicates better outcomes on the difficulties scales. This decrease indicates that having a brother as opposed to a sister results in, on average, fewer behavioural problems perceived by the parent.

One possible explanation is gender-based behavioural expectations. Parents may be less sensitive to and may hold lower behavioural expectations for aggressive behaviour when boys are part of the family (whether being the study child or the sibling). In that case, a girl with a brother is perceived as behaving better by comparison—even if the behaviour is the same as those with a sister.

Another potential explanation is that male siblings tend to engage in distinct forms of interaction that may offer unique developmental benefits, particularly for children with attention difficulties. Compared to sisters, brothers are more likely to participate in rough-and-tumble play and high-energy physical activities. These forms of interaction have been shown—at least in parent-child contexts—to support self-regulation and reduce aggression (Flanders et al., 2009; Flanders et al., 2020). Moreover, recent work by Feinberg et al. (2012) documenting how sibling interactions serve as crucial contexts for developing self-regulation skills as well as other crucial skill. Although direct evidence on sibling rough-and-tumble play is limited, these findings suggest a plausible mechanism. In contrast to the more verbal, cooperative, and structured play often observed between sisters, interactions with brothers are typically more physical, dynamic, and unpredictable. Such play may demand greater impulse control, attentional flexibility, and real-time social adaptation—skills that are particularly relevant for children exhibiting hyperactivity and inattention. As such, having a brother could serve as a natural training ground for developing these capacities through everyday play.

*Figure 2.1 Parent 1-reported SDQ on SC*



### 2.4.2 Heterogeneity by Gender and Birth Order

To further investigate, I disaggregate the treatment effects by the study child's own gender and birth order (as shown in Figure *2.2*). Second-born children exhibit a trend with increased scores in pro-social ratings (+0.07 standard deviations, $p = 0.119$), although it is not statistically significant. The effect is more evident for girls, with a significant positive coefficient (+0.09 standard deviations, $p < 0.05$) on the pro-social score, suggesting having a brother enhanced pro-social behaviour in daily life. No significant effects on pro-social behaviour are found for other groups.

The finding that having a brother enhances pro-social behaviour in girls most likely supports sibling differentiation theory. Because boys typically display lower levels of pro-social behaviour than girls, social learning theory would predict the opposite pattern, that girls with brothers should exhibit less pro-sociality due to observing and imitating their brothers' lower levels of nurturing behaviour or caregiving tendencies. Yet the evidence here instead suggests that siblings may intentionally cultivate contrasting traits to establish distinct family roles and reduce intra-household competition. In this context, girls with brothers may strategically develop stronger pro-social tendencies to differentiate themselves from their brothers' more competitive or less socially oriented behavioural patterns.

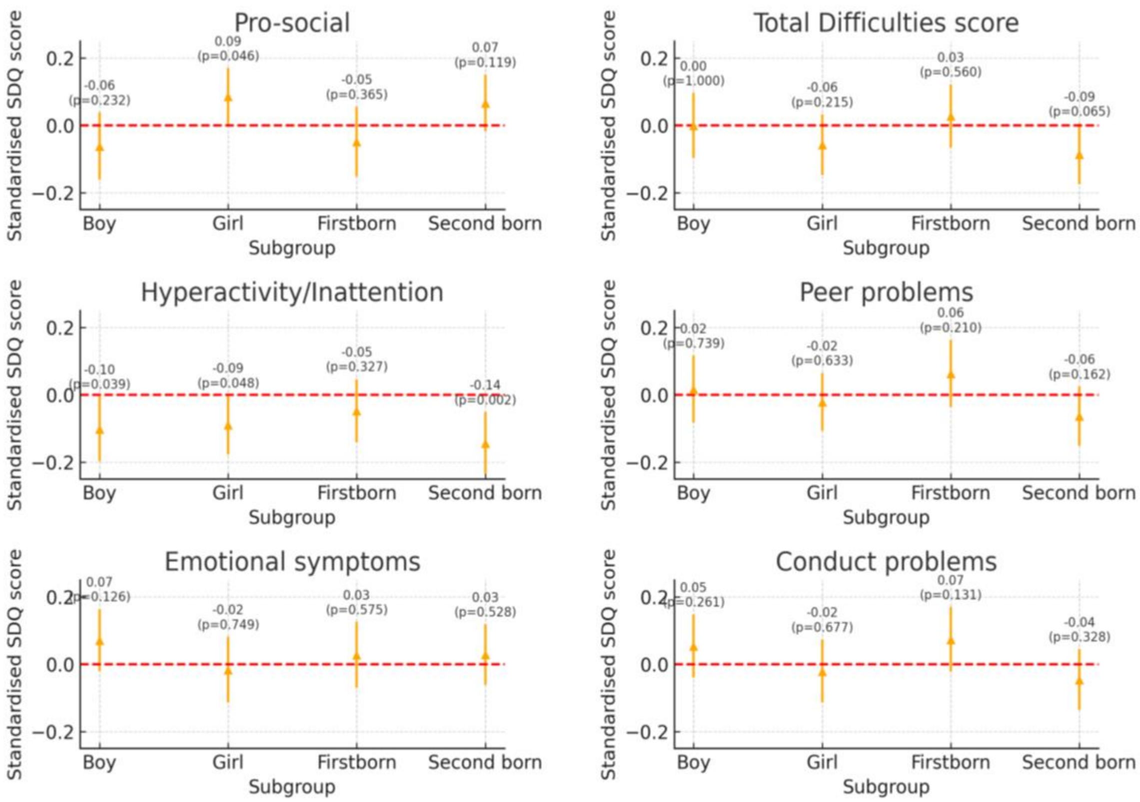
Interestingly, the effect of having a brother is consistently beneficial across all sub-groups in the domain of hyperactivity/inattention. Children with brothers tend to have fewer reports of experiences with restlessness, fidgeting, or inattention, with the strongest and most significant reduction observed among second-born children (-0.14 standard deviations, $p < 0.05$). The decrease in hyperactivity/inattention scores is statistically significant for boys, girls, and second-born children, but not for first-born children. Since both girls and boys benefit, it implies a non-gender-specific mechanism.

Across all sub-groups, the estimated effect of having a brother on emotional symptoms, peer problems, conduct problems, and total difficulty scores is small and mostly statistically insignificant, with no meaningful differences between boys versus girls or first-born versus second-born children. The positive coefficients for boys on emotional symptoms and conduct problems suggest that having a brother might be associated with slightly more internalising difficulties and conduct problems, but the effects are not significant. However, there is evidence that second-born children, in particular, benefit from having a brother as they exhibit a significant reduction in total difficulties scores (-0.12 standard deviations, $p = 0.065$).

The stronger effect among second-born children supports the social learning explanation. Younger children may observe and imitate older siblings' behaviour, learning how to manage attention and impulse control. On the other hand, first-born children exhibit the opposite pattern, showing no significant differences between having a brother or sister across all scales, which can be explained by social learning theory in that they have no older siblings to model from.

The reduction in hyperactivity/inattention for children with brothers may reflect the regulatory benefits of physical play typically occurring in male sibling interactions, which explains why effects persist for both boys and girls with brothers.

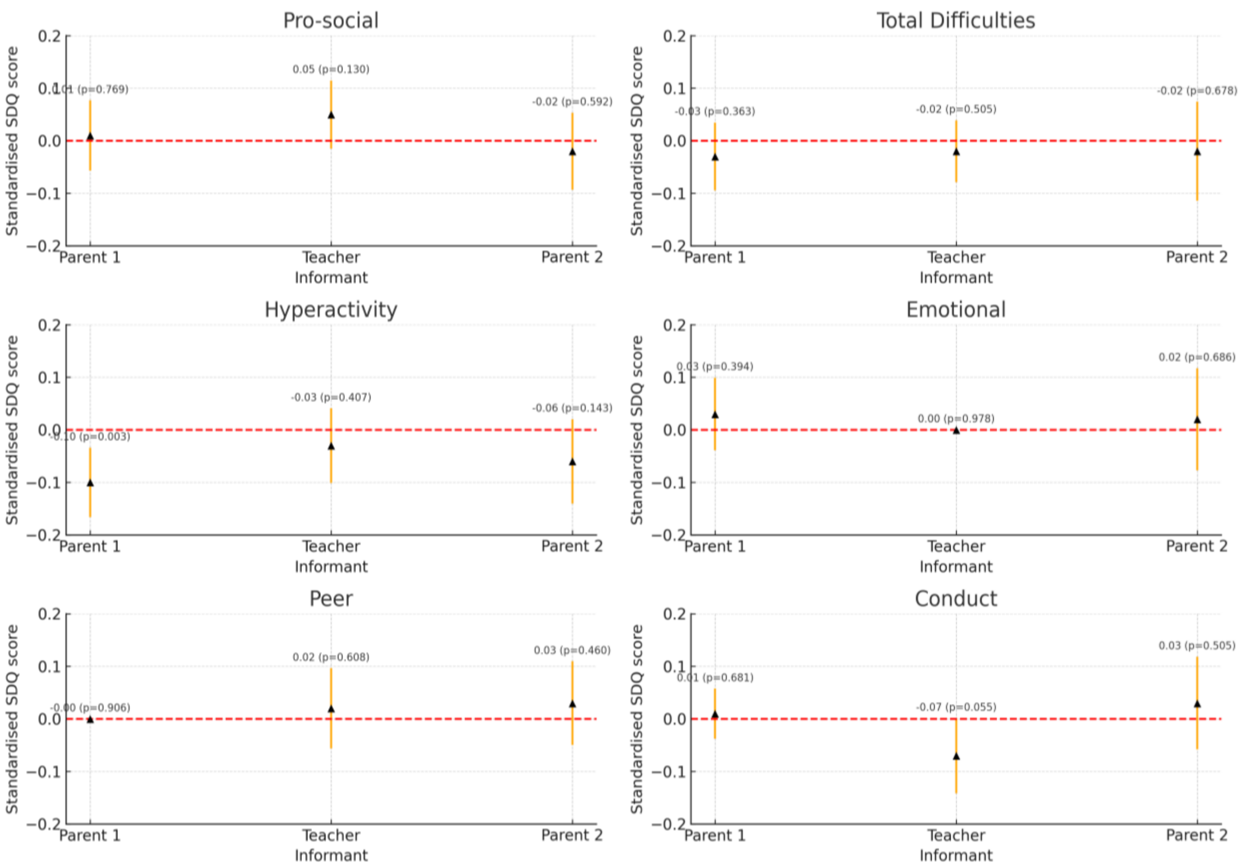
*Figure 2.2 Parent-1 reported SDQ by gender and birth order*




## 2.4.3 Heterogeneity by Informants

Figure 2.3 compares the estimated effects of having a brother across three informants—Parent 1, teachers, and Parent 2—for each SDQ domain. Overall, the patterns are consistent, suggesting that the estimated treatment effects are not driven by any single reporting source. For instance, in the domain of hyperactivity/inattention, all three informants report negative effects (i.e., fewer behavioural problems when the child has a brother), though the magnitude and significance vary, with Parent 1's report reaching statistical significance. For pro-social behaviour, teacher reports suggest a small benefit (+0.05 standard deviations), while reports from parents show smaller or slightly negative effects. In total difficulties and peer problems, the effects are modest and mostly negative but not statistically significant across all informants. Emotional symptoms and conduct scores also remain stable, with estimates close to zero, reinforcing the notion that having a brother does not significantly affect internalising behaviours. One notable exception is teacher-reported conduct problems, where the effect approaches statistical significance in a negative direction (-0.07 standard deviations, $p$ = 0.055), suggesting that teachers may perceive fewer conduct issues in children with brothers. Taken together, this triangulation across informants enhances the robustness of the findings and affirms the relative consistency in the direction of effects, especially in the externalising domains.

*Figure 2.3 Effects on SDQ reported by various informant*



## 2.5 Discussion

The findings of this study both complement and challenge existing studies on sibling gender effects. While previous studies focus primarily on educational and labour market outcomes (Butcher & Case, 1994; Brenøe, 2022), I provide direct evidence on non-cognitive development. The findings suggest that while the average effect of having a brother is small, substantial heterogeneity exists. Childers who are second born are more favourably affected by having a brother across several behavioural dimensions, followed by girls, whereas first born are unaffected.

Despite huge heterogeneous effects observed by gender, by birth order and by informants, the finding of having a brother reduces hyperactivity/inattention problems by 0.10 standard deviations remains robust. This effect is particularly pronounced for second-born children (0.14 standard deviations). Moreover, having a brother is consistently associated with lower scores on the hyperactivity/inattention for both boys and girls, suggesting a gender-specific beneficial sibling dynamics and interaction with brother.

The effects on hyperactivity/inattention—a key predictor of educational achievement and labour market success (Currie & Stabile, 2006)—suggest sibling gender shapes crucial non-cognitive skills and have lasting effects for later-life outcomes.

While the finding of having a brother enhances pro-social behaviour among girls support sibling differential theory, the heterogeneous effects by birth order provide compelling support for social learning theory. Second-born children, who can observe and model their older siblings' behaviours, show the strongest benefits from having a brother. This aligns with Bandura's (1977) social learning theory. The absence of effects for first-born children—who lack older sibling role models—further reinforces this interpretation.

The findings have several policy implications. The beneficial effects of brothers on hyperactivity/inattention suggest that single-gender educational environments may inadvertently limit opportunities for cross-gender social learning that promotes behavioural regulation. This provides micro-level evidence relevant to debates about single-sex schooling (Eisenkopf et al., 2015).

For early childhood interventions targeting attention and behavioural problems, the results suggest considering family composition when designing programs. Children without brothers—particularly second-born children—may benefit from structured opportunities for the types of physical, regulatory play typically occurring in mixed-gender sibling interactions.

The persistent effects across informants also highlight the importance of multi-source assessment in identifying children's non-cognitive needs. Relying solely on parent or teacher reports may miss important behavioural patterns that manifest differently across contexts.

Nevertheless, the paper still suffers several limitations. First, while I argue sibling gender is quasi-random conditional on family size, unobserved parental preferences could still bias results. Despite the balance checks in Table 2.4 show no meaningful differences in observable characteristics across first-born gender, potential gender-selective behaviour and stopping rule indicated by the slight over-representation of mixed-gender pairs have not been completely ruled out and more robust design should be adopted. Second, the SDQ, while psychometrically validated, still captures a limited range of non-cognitive skills. Future research could examine other important traits like grit, self-efficacy, or creativity.

## 2.6 Conclusion

This study provides novel evidence that sibling gender composition significantly influences non-cognitive skill development in childhood. Using nationally representative Australian data and exploiting the quasi-random assignment of sibling gender, I find that having a brother reduces hyperactivity and inattention problems, with particularly strong effects for second-born children. Girls with brothers benefit from pro-social domain, while second-born children with brothers show trend toward fewer total behavioural difficulties.

The findings open several directions for future research. Longitudinal follow-up could examine whether early differences in attention and behaviour translate into educational and labour market outcomes. Cross-cultural comparisons could investigate how societal gender norms moderate sibling effects. Additionally, examining mechanisms through detailed time-use data or experimental interventions could provide deeper insights into how sibling interactions shape skill development.

As policymakers increasingly recognize non-cognitive skills as crucial for life success and amenable to intervention, understanding their developmental origins becomes paramount. The evidence suggests family composition—specifically sibling gender—represents an important but understudied influence on these critical skills. Recognizing these effects can inform more effective, targeted interventions to promote optimal child development and reduce inequalities stemming from family structure differences.

# Appendix 2 Items of the Strengths and Difficulties Questionnaire (SDQ)

1. Pro-social Behaviours
   a) Considerate of other people's feelings
   b) Shares readily with other children (treats, toys, pencils, etc.)
   c) Helpful if someone is hurt, upset or feeling ill
   d) Kind to younger children
   e) Often volunteers to help others (parents, teachers, other children)

2. Hyperactivity/Inattention
   a) Restless, overactive, cannot stay still for long
   b) Constantly fidgeting or squirming
   c) Easily distracted, concentration wanders
   d) Thinks things out before acting
   e) Good attention span, sees chores or homework through to the end

3. Emotion Symptoms
   a) Often complains of headaches, stomach aches or sickness
   b) Many worries, often seems worried
   c) Often unhappy, depressed or tearful
   d) Nervous or clingy in new situations, easily loses confidence
   e) Many fears, easily scared

4. Peer Problems
   a) Rather solitary, tends to play alone
   b) Has at least one good friend
   c) Generally liked by other children
   d) Picked on or bullied by other children
   e) Gets on better with adults than with other children

5. Conduct Problems
   a) Often loses temper
   b) Generally well behaved, usually does what adults request
   c) Often fights with other children or bullies them
   d) Often lies or cheats
   e) Steals from home, school or elsewhere

# Reference

Almlund, M., Duckworth, A. L., Heckman, J., & Kautz, T. (2011). Personality psychology and economics. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the economics of education* (Vol. 4, pp. 1–181). Elsevier. https://doi.org/10.1016/B978-0-444-53444-6.00001-8

Anger, S., & Schnitzlein, D. D. (2017). Cognitive skills, non-cognitive skills, and family background: Evidence from sibling correlations. *Journal of Population Economics, 30*(2), 591–620. https://doi.org/10.1007/s00148-016-0625-9

Bandura, A. (1977). *Social learning theory*. Prentice Hall.

Brenøe, A. A. (2022). Brothers increase women's gender conformity. *Journal of Population Economics, 35*(4), 1859–1896. https://doi.org/10.1007/s00148-021-00830-9

Butcher, K. F., & Case, A. (1994). The effect of sibling sex composition on women's education and earnings. *The Quarterly Journal of Economics, 109*(3), 531–563.

Conley, D., & Glauber, R. (2006). Parental educational investment and children's academic risk: Estimates of the impact of sibship size and birth order from exogenous variation in fertility. *Journal of Human Resources, 41*(4), 722–737.

Cyron, L., Schwerdt, G., & Viarengo, M. (2016). The effect of opposite sex siblings on cognitive and noncognitive skills in early childhood. *ADBI Working Paper 627*. Asian Development Bank Institute.

Dudek, T., Brenøe, A. A., Feld, J., & Rohrer, J. M. (2022). No evidence that siblings' gender affects personality across nine countries. *Psychological Science, 33*(9), 1574–1587.

Eisenkopf, G., Hessami, Z., Fischbacher, U., & Ursprung, H. W. (2015). Academic performance and single-sex schooling: Evidence from a natural experiment in Switzerland. *Journal of Economic Behavior & Organization, 115*, 123–143.

Feinberg, M. E., Solmeyer, A. R., & McHale, S. M. (2012). The third rail of family systems: Sibling relationships, mental and behavioral health, and preventive intervention in childhood and adolescence. *Clinical child and family psychology review*, *15*, 43-57.

Flanders, J. L., Leo, V., Paquette, D., Pihl, R. O., & Séguin, J. R. (2009). Rough-and-tumble play and the regulation of aggression: An observational study of father–child play dyads. *Aggressive Behavior, 35*(4), 285–295.

Flanders, J. L., Simard, M., Paquette, D., Parent, S., Vitaro, F., Pihl, R. O., & Séguin, J. R. (2010). Rough-and-tumble play and the development of physical aggression and emotion regulation: A five-year follow-up study. *Journal of Family Violence, 25*(4), 357–367.

Golsteyn, B. H., & Magnée, C. A. (2020). Does sibling gender affect personality traits? *Economics of Education Review, 77*, 102016.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry, 38*(5), 581–586.

Heckman, J. J., & Kautz, T. (2012). Hard evidence on soft skills. *Labour Economics, 19*(4), 451–464. https://doi.org/10.1016/j.labeco.2012.05.014

Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics, 24*(3), 411–482. https://doi.org/10.1086/504455

Lee, V. E., & Lockheed, M. E. (1990). The effects of single-sex schooling on achievement and attitudes in Nigeria. *Comparative Education Review, 34*(2), 209–231.

McHale, S. M., Updegraff, K. A., Helms-Erikson, H., & Crouter, A. C. (2001). Sibling influences on gender development in middle childhood and early adolescence: A longitudinal study. *Developmental Psychology, 37*(1), 115–125. https://doi.org/10.1037/0012-1649.37.1.115

Pahlke, E., Hyde, J. S., & Allison, C. M. (2014). The effects of single-sex compared with coeducational schooling on students' performance and attitudes: A meta-analysis. *Psychological Bulletin, 140*(4), 1042–1072. https://doi.org/10.1037/a0035740

Peter, N., Lundborg, P., Mikkelsen, S., & Webbink, D. (2018). The effect of a sibling's gender on earnings and family formation. *Labour Economics, 54*, 61–78. https://doi.org/10.1016/j.labeco.2018.06.006

Rao, N., & Chatterjee, T. (2018). Sibling gender and wage differences. *Applied Economics, 50*(15), 1725–1745. https://doi.org/10.1080/00036846.2017.1374536

Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin, 126*(1), 3–25. https://doi.org/10.1037/0033-2909.126.1.3

Schachter, F. F., Gilutz, G., Shore, E., & Adler, M. (1978). Sibling deidentification judged by mothers: Cross-validation and developmental studies. *Child Development, 49*(2), 543–546. https://doi.org/10.2307/1128728

Sutter, M., Zoller, C., & Glätzle-Rützler, D. (2019). Economic behavior of children and adolescents—A first survey of experimental economics results. *European Economic Review, 111*, 98–121

# Chapter 3 — Trapped in the Cycle: How Mental Accounting and Asymmetric Friction Affect Online Gambling Behaviours

Joint work with Hanlin Lou, University of New South Wales

## 3.1 Introduction

Gambling is a significant public policy concern worldwide, particularly in Australia. According to data from H2 Gambling Capital, a UK-based gambling consultancy, Australia records the highest per capita gambling losses globally (Letts, 2018). The estimated annual gambling loss per adult is USD $958, exceeding that of Hong Kong, the second highest, by 20%. Nearly one in four Australian adults has gambled at least once in the past 12 months, primarily on lottery purchases and betting on horse races and sports (Australian Gambling Research Centre, 2023). Among these individuals, almost half are classified as being at some risk of gambling harm according to the Problem Gambling Severity Index (PGSI). Excessive gambling can impose severe economic and social costs on individuals and families. However, with advancements in technology and digitalisation, an increasing range of gambling products is available via online platforms, enhancing accessibility and encouraging rapid and continuous play.

Online gambling platforms such as The Lott and Sportsbet are widely used in Australia.[16] In contrast to land-based gambling, such as casinos and poker machines, where gamblers can exchange chips for cash at the cashier, online platforms typically require players to use an account linked to their bank account or cards for deposits and withdrawals. This online gambling account effectively functions as a virtual wallet, which can influence spending behaviour through the lens of mental accounting. According to Thaler (1985), individuals' consumption decisions often violate the principle of fungibility, as they create separate mental accounts and treat money differently depending on its designated "account." For example, an individual might receive a $500 tax refund and mentally label it as "bonus money," spending it freely on entertainment, even though they would be more cautious with the same $500 from their regular salary. Similarly, in the context of online gambling, funds deposited into a gambling account may be psychologically separated from other household funds, making players more willing to spend them on bets or lottery tickets.

Lottery tickets and bets can be purchased using either funds deposited into these online accounts or via instant payment methods, such as bank cards, Apple Pay, Google Pay, PayID, and PayPal, with deposits and purchases typically processed immediately. In contrast, withdrawing winnings is considerably less seamless. Winnings are credited exclusively to the online gambling account, requiring users to initiate a withdrawal to transfer funds back to their personal bank account or card. Several barriers make withdrawals more cumbersome than deposits. For instance, unverified accounts may allow deposits but require identity verification for withdrawals. While deposits are processed instantly, withdrawals often involve delays ranging from a few hours to several business days. In some cases, users must re-link their bank accounts for withdrawals, even if they were previously linked. Additionally, deposited funds cannot be withdrawn unless they have been used for gambling and converted into winnings, so the account balance and withdrawable balance may differ. [17]

These additional steps, including extra effort, waiting time, and restrictions, generate non-pecuniary costs unique to the withdrawal process, resulting in an asymmetric

---

[16] The most popular gambling products in Australia are lotteries and scratch cards, followed by race betting, which includes horse and greyhound racing. Sports betting ranks third in popularity. The Lott and Sportsbet are two common examples of online gambling operators offering these products.

[17] This withdrawal condition applies universally across online lottery and betting platforms. However, only some platforms display both total account balance and the withdrawable balance separately.

friction system in online gambling. Nowadays, businesses often design systems that make it exceptionally easy for customers to sign up for services, while imposing significant friction when they attempt to cancel or unsubscribe. For instance, individuals can set up energy or internet services online within minutes, yet disconnecting these services may require waiting for half an hour in a queue to speak with a human customer service representative. The practices of converting cash into restricted monetary forms and introducing asymmetric friction primarily benefit businesses but reduce consumer welfare. Individuals engaging in online gambling are particularly more vulnerable to these business tactics, highlighting the importance of studying asymmetric frictions in online gambling. While non-pecuniary costs may be trivial in the context of online gambling for large-prize winners, they can be substantial for small-prize winners, especially for those receiving $20, a common prize level among most players.[18]

Motivated by these institutional features, this study explores the role of mental accounting and asymmetric friction in the online gambling domain. Understanding these dynamics is critical for evaluating the behavioural impact of online gambling and for informing effective policy interventions.

We conducted an online experiment comprising two stages with cross-randomisation. We conducted an online experiment comprising two stages with cross-randomisation. The first stage aimed to identify the role of mental accounting, while the second stage examined how asymmetric frictions in the withdrawal process influence online gambling behaviour. A sample of 592 Australian adults aged 18 or above was recruited via Prolific.[19] We stratified randomisation by gender to avoid severe imbalances between groups. Gender differences have been extensively documented in domains such as overconfidence, risk attitudes, and financial decision-making (Sunden & Surette, 1998; Schubert et al., 1999; Barber & Odean, 2001; Croson & Gneezy, 2009; Cueva et al., 2019; Thaler, 2021). Given the intrinsic risk and uncertainty involved in gambling, exploring gender differences in decision-making is particularly relevant.

---

[18] For example, in a recent Powerball draw (Draw No. 1502), a total of 988,213 winners received a prize in Division 8 or Division 9, with an average payout of less than AUD $20.

[19] There were over 6,000 matching subjects on Prolific who had been active for the past 90 day at the time of our experiment.

This study makes three key contributions. First, it contributes to the literature on mental accounting. The concept of mental accounting, first introduced by Thaler (1985), explains that individuals' consumption decisions often violate the principle of fungibility, as they create separate mental accounts and treat money differently based on subjective criteria such as its source or intended use. Since then, mental accounting has been widely studied and empirically verified in numerous studies by economists and behavioural researchers, collectively advancing understanding of its effects on consumer behaviour and financial decision-making (Thaler, & Jason, 1990; Heath & Soll, 1996; Gourville & Soman, 1998; Prelec & Loewenstein, 1998; Thaler, 1999; Soman, 2001; Thaler & Benartzi, 2004; Hastings & Shapiro, 2013). Later studies by Kooreman (2000) and Beatty et al. (2014) extended this concept, demonstrating that spending decisions depend not only on total wealth but also on its composition. Sui et al. (2021) further identify mental accounting as a key factor in unintentional overspending, demonstrating that overspending from earned income, excessive credit card use (where balances are rarely paid off), and spending beyond expectations for a 'normal' year are all influenced by wealth allocation.

Despite gambling being a distinct form of consumption or investment, few studies have examined the role of mental accounting in gambling behaviour. If individuals are influenced by mental accounting when gambling online, they may be more likely to spend money stored in online gambling accounts than funds held in bank accounts. This mechanism may also explain why gambling platforms design systems that effectively trap money within these accounts.

In Stage 1 of the experiment, participants were assigned hypothetical gambling-specific and bank accounts with randomly allocated credit balances, keeping total wealth constant. They then participated in a gamble-equivalent game, deciding how many credits to wager, with payoffs dependent on remaining balances. Results indicate that individuals exhibit mental accounting in online gambling, treating money differently depending on where it is stored. Participants with higher balances in gambling accounts placed larger bets despite equal total wealth. No significant gender differences were observed, but individuals without experience in high-risk financial activities such as frequent share trading, derivatives, or cryptocurrencies exhibited stronger mental accounting biases.

Second, the study contributes to the literature on responsible gambling interventions by highlighting the role of asymmetric frictions in the withdrawal process. Existing research on responsible gambling interventions primarily focuses on pre-gamble and in-game measures. Pre-gamble interventions include transaction blocks that prevent bank transfers to gambling accounts and deposit limits that cap allowable spending over a set period. In-game interventions involve on-screen notifications, spending reminders, pop-up warnings, and enforced breaks in play (Harris & Griffiths, 2017; Ivanova et al., 2019; Auer et al., 2020; The Behavioural Insights Team, 2021; Edson et al., 2021).

However, few studies or policies address post-gamble interventions, particularly withdrawal frictions. Existing evidence is largely qualitative or anecdotal. For example, Hing et al. (2014) conducted interviews with problem gamblers, finding that withdrawal difficulties increased gambling duration and expenditure. In 2020, the UK Gambling Commission banned reverse withdrawals due to evidence that they exacerbate problem gambling (Gambling Commission, 2021). Newall and Rockloff (2022) argue that online gambling designs, including reverse withdrawal, align with the concept of "sludge" that impairs decision-making with excessive friction. Auer and Griffiths (2023) find that personalised messages encouraging withdrawals can reduce gambling-related harm.

Despite these insights, empirical evidence on how withdrawal frictions influence gambling behaviour remains limited. Unlike frictionless deposits, withdrawals often involve delays and additional requirements, creating non-pecuniary costs that can deter individuals from accessing winnings, particularly smaller amounts. Consequently, funds may remain in gambling accounts longer, potentially increasing gambling participation and expenditure. This paper provides the first experimental evidence quantifying the impact of withdrawal barriers on gambling behaviour.

In Stage 2, participants were assigned two hypothetical accounts with identical total endowments and allocations. They played four rounds of gamble-equivalent games, deciding whether to transfer funds between accounts. Any untransferred funds in the gambling-specific account were automatically used in the game. The control group faced no transfer restrictions, while the treatment group encountered a time-costly withdrawal task. Payoff depended on the final credit balance. Results show that withdrawal barriers increase the probability of inaction with respect to funds in the gambling-specific account, driven by reductions in both deposits and withdrawals. This inaction led to higher gambling-specific account balances and passive gambling participation, potentially perpetuating a cycle of risk exposure. Older individuals, particularly those aged 50 and over, were more affected, retaining higher balances and increasing bet sizes.

Third, the study provides policy-relevant insights by linking psychological mechanisms with institutional features of the gambling industry. The interaction of mental accounting and asymmetric frictions illustrates how platform design can influence gambling behaviour. Our findings suggest that withdrawal frictions reduce the likelihood of withdrawing funds and, when combined with mental accounting, contribute to higher gambling participation. Prior research indicates that, although the quantitative magnitudes may vary depending on context, experience, and stakes, the qualitative patterns of risky and uncertain choice observed in laboratory settings generally reflect real-world decision-making (Camerer & Hogarth, 1999; Levitt & List, 2007; Schindler & Pfattheicher, 2017). Accordingly, regulators could mitigate gambling-related harm by removing withdrawal barriers, for example, requiring winnings to be transferred directly to bank accounts via near-instant payment systems such as Osko® or PayID. Understanding these mechanisms highlights the potential for exploitation and provides guidance for designing interventions and regulatory policies that better protect consumers.

The remainder of this paper is structured as follows: Section 2 details the experimental design and hypotheses, Section 3 presents the empirical methodology and results, and Section 4 concludes.

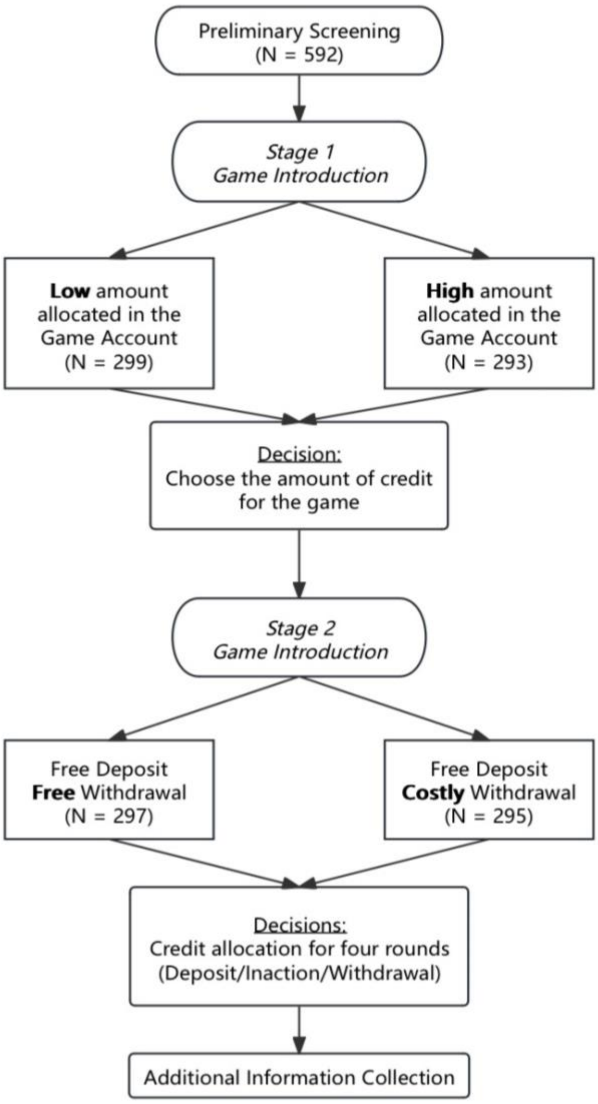
## 3.2 Design of the Experiment

To examine the presence of mental accounting and assess the impact of withdrawal barriers on gambling behaviour in an online setting, the design of the experiment plays a crucial role in replicating real-life online gambling conditions. We implemented a two-stage online experiment with cross-randomisation via Qualtrics. Figure 3.1 presents the flowchart of the experimental design, which we describe below, and the full set of instructions and questions is provided in Appendix 3 Qualtrics Survey.

The first stage investigates mental accounting and its effects on gambling behaviour. As discussed earlier, mental accounting refers to the tendency of individuals to categorise and treat money differently depending on its source, label, or intended use, thereby displaying non-fungibility. An online experimental setting allows us to assess whether individuals exhibit mental accounting by varying the framing of funds or the labelling of accounts and observing their betting decisions. Specifically, we examine whether gambling behaviour differs when money is allocated to an account explicitly earmarked for gambling versus a general account. This leads to our first hypothesis:

***Hypothesis 1:*** *Subjects with a higher balance in a gambling-specific account will be more likely to increase their bets compared to those with a lower balance, even though their total wealth remains the same.*

*Figure 3.1 Flow chart*




The second stage evaluates the impact of withdrawal barriers on both withdrawal and gambling behaviour. We introduce a time-costly task in the withdrawal process to simulate real-world friction. By imposing this task before withdrawal completion, we test the following hypothesis:

***Hypothesis 2:*** *Subjects who face a time-costly task in the withdrawal process will be less likely to withdraw their funds from the gambling-specific account and more likely to gamble compared to those who do not.*

To minimise experimenter effects and enhance realism, we employed contextual priming by instructing subjects to imagine a specific scenario: "Please imagine that you have been invited by a company based in Australia to conduct game testing." The experiment was framed as an incentivised game testing study, with subjects being informed that they could receive bonus payments based on their game performance. To avoid bias and ensure that subjects were not influenced by negative connotations, we used euphemistic terms, replacing "gamble" with "game" and avoiding words such as "bet," "stake," "win," or "lose." The experimental currency was credits, with a conversion rate of 50 credits equalling 5 cents AUD in bonus payments. Additional information on subjects' demographics, preferences and financial were collected at the end of the experiment with a final attention check question.

### 3.2.1 The First Stage

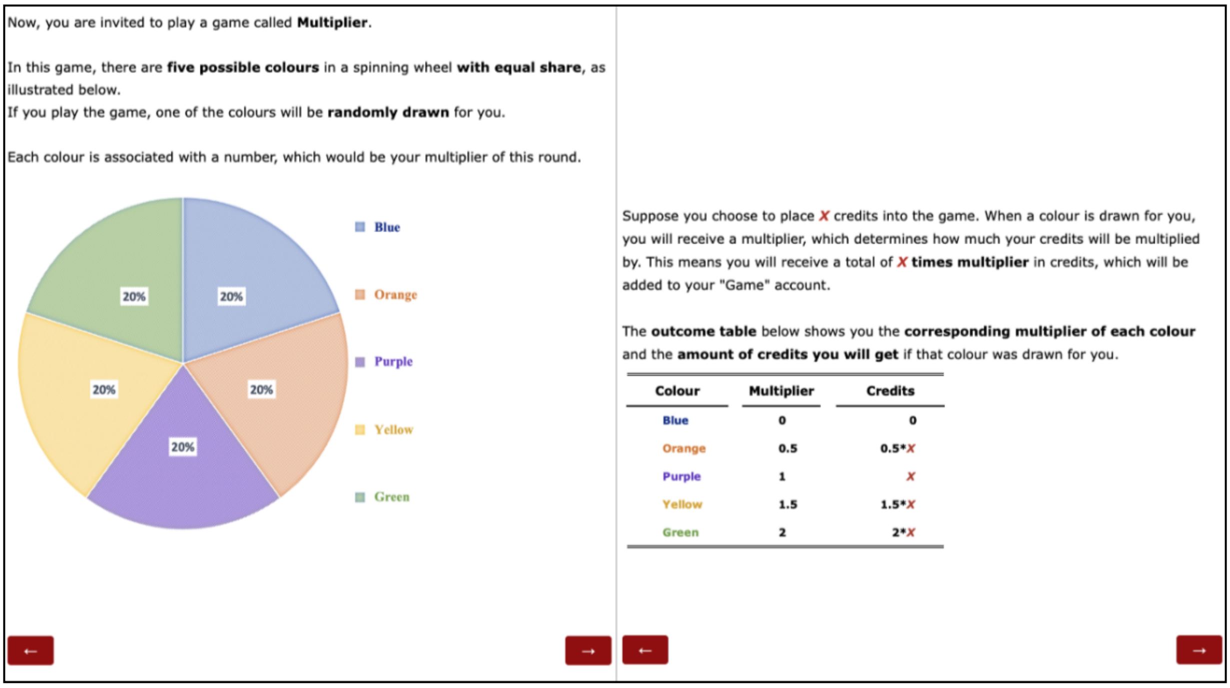
Subjects were assigned two hypothetical accounts: a "Transaction" account and a "Game" account. They were instructed that only credits in the "Game" account could be used for participation in the game, while the "Transaction" account served as a source for topping up the "Game" account and redeeming bonus payments at the end of the experiment. Participants played a game called *Multiplier*, a modified version of the *Big Six Wheel*.

The game wheel was divided into five equal-coloured sections, each colour $c \in \{\text{blue}, \text{orange}, \text{purple}, \text{yellow}, \text{green}\}$ corresponding to a multiplier value $m \in \{0, 0.5, 1, 1.5, \text{and } 2\}$.[20] All five colours had an equal probability of being drawn. If a

[20] By setting the expected value of the multiplier to 1, subjects neither gain nor lose money on average over the long run. This neutrality helps avoid systematic bias, ensuring that any observed effects on behaviours are due to the experimental conditions or their preferences, rather than an inherent advantage or disadvantage in the game's setup. If the expected value is set to be less than 1, subjects will generally experience net losses, potentially leading to more conservative behaviour. Conversely, if the expected value is greater than 1, it would encourage riskier behaviour.

subject placed $x$ credits in the game, their return was $mx$ credits. For example, if a subject wagered 100 credits from the “Game” account and landed on green, they would receive 200 credits, thereby increasing their balance by 100 credits.  provides a screenshot of the game introduction.

*Figure 3.2 Screenshots of game explanation*

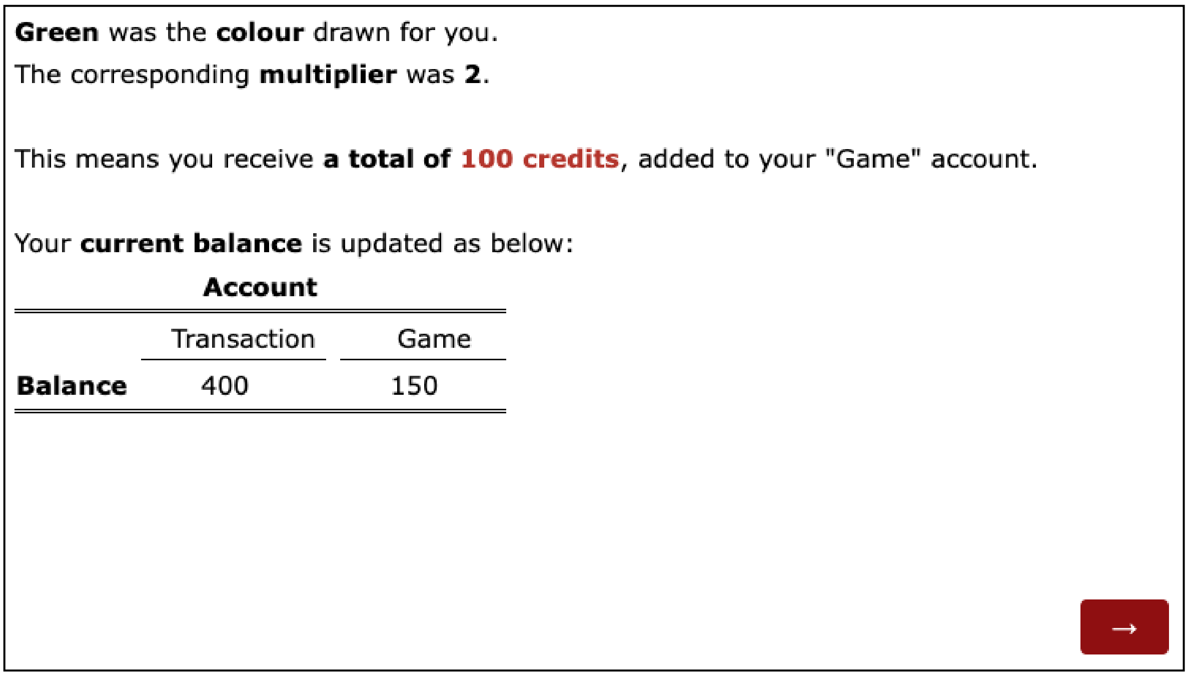



Subjects were stratified by gender and randomly assigned to either the “High Game Balance” or “Low Game Balance” group. Both groups received an initial endowment of 500 credits, but the allocations differed: the “High Game Balance” group had 400 credits in the “Game” account and 100 credits in the “Transaction” account, while the “Low Game Balance” group had the opposite allocation.

Subjects were asked to decide how many credits to use for a single round of the game. They could opt out by placing 0 credits. They were also allowed to enter a value exceeding their “Game” account balance, with any shortfall automatically transferred from the “Transaction” account. Thus, the betting range was between 0 and 500 credits.

After making a decision, subjects were shown the drawn colour, the corresponding multiplier value, and their updated account balances, as depicted in Figure 3.3. At the end of Stage 1, any remaining credits in the “Game” account were transferred to the “Transaction” account, which determined the final payoff.

*Figure 3.3 Screenshot of game outcome*




## 3.2.2 The Second Stage

Stage 2 followed a similar setup to Stage 1. All subjects were given “Transaction” and “Game” accounts, each initially endowed with 500 credits. They participated in *Multiplier Plus* for four rounds, a variation of *Multiplier* where multiplier values varied between rounds. In Rounds 1 and 4, the colour multipliers were set at $m \in \{0.5, 0.75, 1, 1.25, 1.5\}$, with slight modifications in Rounds 2 and 3.[21]

A crucial change in Stage 2 was the introduction of a **default participation mechanism**, where any remaining credits in the “Game” account were automatically used in the next game round unless actively withdrawn. This design ensured that withdrawal decisions in the experiment were incentivised, mirroring real-world online gambling conditions. In reality, funds left in an online gambling account are typically restricted in

21 The expected value of the multiplier is set to be 1 in all four rounds, however, there was one round of high risk with the multiplier value ranges from 0, 0.5, 1, 1.5, 2], and one round of low risk with the multiplier value ranges between [0.75, 0.875, 1, 1.125, 1.25]. We also randomised the order of high risk round and low risk round within each group. This design was exploratory, and we do not discuss it further. See complete survey in Appendix 3 Qualtrics Survey for the detailed multiplier value of each colour in each round.

fungibility and liquidity until withdrawn. Without actively requesting a withdrawal, these funds cannot be converted into tangible money. However, since experimental payouts were consolidated into a single lump-sum payment at the end, implementing the default participation mechanism allowed subjects to perceive differences between the two accounts.

Unlike Stage 1, subjects in Stage 2 chose whether to transfer credits between accounts rather than deciding a bet amount. At the beginning of each round, they selected one of three options:

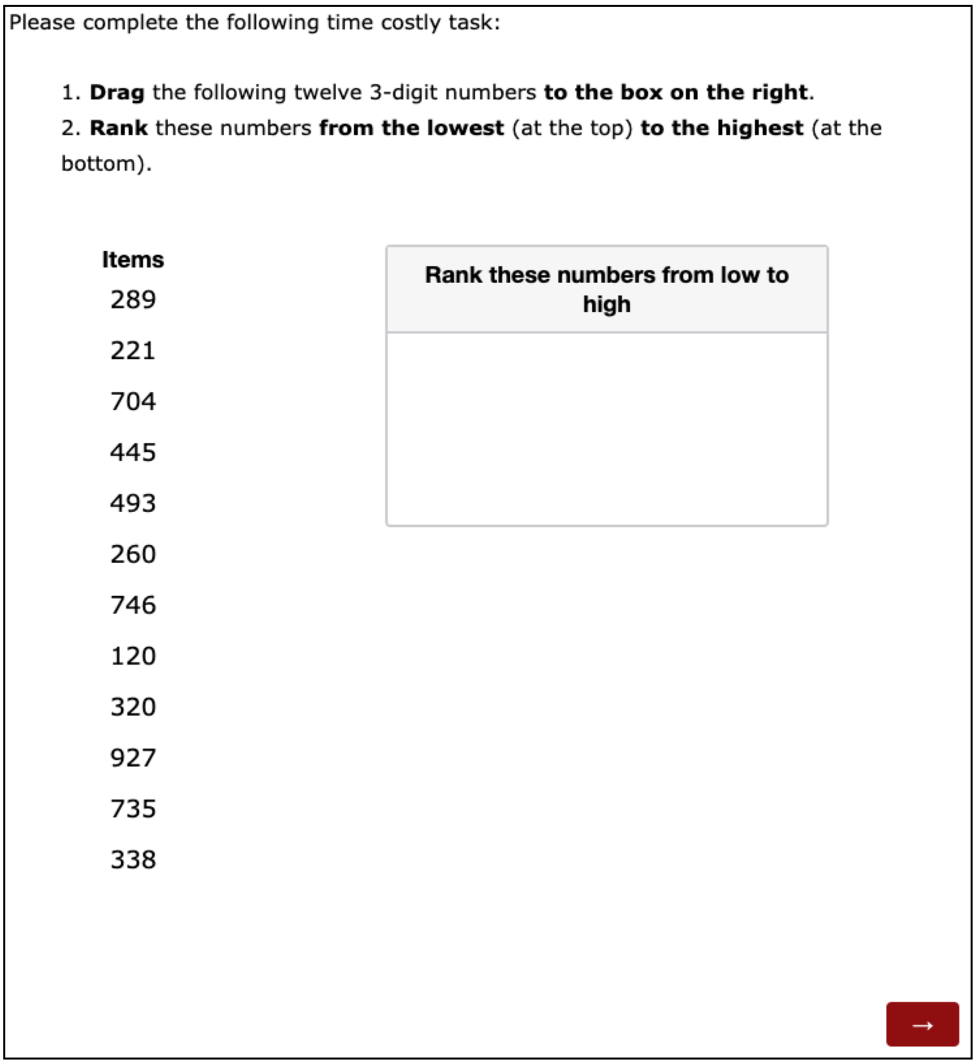
1. *Withdraw funds (transfer credits from the "Game" account to the "Transaction" account, reducing the available gambling balance).*
2. *Deposit funds (transfer credits from the "Transaction" account to the "Game" account, increasing the available gambling balance).*
3. *Do nothing (maintain the current balance).*

For those selecting withdrawal or deposit, a follow-up screen required them to specify the transfer amount. The final amount available for gambling in each round was thus determined by their choices.

Subjects were randomly assigned to either the control or treatment group. In the control group, transfers in both directions were task-free, mirroring Stage 1. In the treatment group, deposits remained task-free, but withdrawals required completion of a time-costly task. After selecting withdrawal, treated subjects had to correctly arrange twelve three-digit numbers in ascending order before specifying the withdrawal amount (see Figure 3.4).

Once all decisions were made, the total bet amount was displayed, followed by the randomly drawn colour and corresponding multiplier value (see Figure 3.5). Updated balances for both accounts were shown, carrying forward to the next round. The experiment concluded after four rounds, with final payoffs determined by summing the "Game" and "Transaction" account balances.

*Figure 3.4 Screenshot of the time-costly task*



Please complete the following time costly task:

1. **Drag** the following twelve 3-digit numbers **to the box on the right**.
2. **Rank** these numbers **from the lowest** (at the top) **to the highest** (at the bottom).

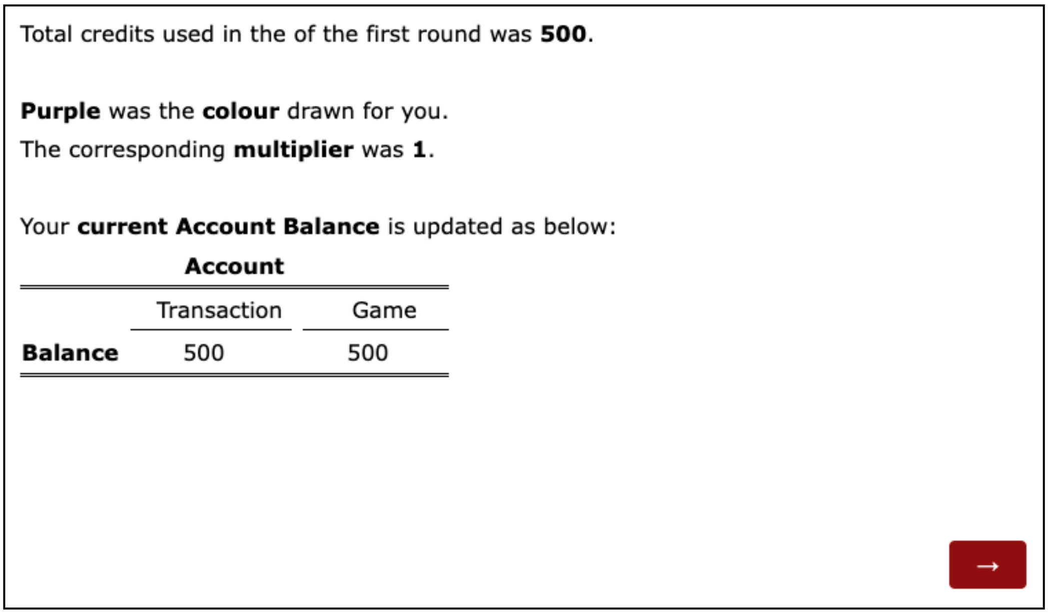
**Items**

289
221
704
445
493
260
746
120
320
927
735
338

**Rank these numbers from low to high**

→

*Figure 3.5 Screenshot of game outcome*



Total credits used in the of the first round was **500**.

**Purple** was the **colour** drawn for you.
The corresponding **multiplier** was **1**.

Your **current Account Balance** is updated as below:

**Account**

| | Transaction | Game |
|---|---|---|
| **Balance** | 500 | 500 |

→

### 3.2.3 Sample

We recruited experimental subjects from Prolific, a UK-based online platform that provides a diverse pool of participants for academic studies. A preliminary screening ensured that subjects met the inclusion criteria: residing in Australia, aged 18 or above, and providing clearly defined gender information. After accounting for incomplete responses, the final sample consisted of 592 subjects. Participants received a base payment of £1.80 ($3.50 AUD) upon completing the experiment, with an estimated completion time of 10 to 15 minutes.[22]

*Table 3.1 Descriptive statistics*

| | **Mean** | **SD** | **Min** | **Max** |
|---|---|---|---|---|
| *Age* | 35.508 | 11.99 | 18 | 86 |
| *Female* | 0.537 | 0.499 | 0 | 1 |
| *College* | 0.715 | 0.452 | 0 | 1 |
| *Married* | 0.497 | 0.5 | 0 | 1 |
| *Employed* | 0.755 | 0.43 | 0 | 1 |
| *Patient* | 15.438 | 11.253 | 1 | 32 |
| *Risk seeking* | 6.693 | 2.828 | 3 | 15 |
| *Financial literacy* | 2.4 | 0.75 | 0 | 3 |
| *Gamblified investing* | 0.652 | 0.883 | 0 | 3 |
| ***N*** | | | | 592 |

*Notes*: Mean, standard deviations, minimum, and maximum values of key demographic variables and characteristics of 592 subjects recruited online via Prolific

Table 3.1 summarises the demographic characteristics and other collected information following Stage 2. The average age of subjects was 35.5 years (SD = 12), with the oldest participant being 86 years, covering a broad age range. The gender distribution was relatively balanced, with females slightly outnumbering males. Educational attainment was high, with 71.5% of the sample holding a college degree. Half of the sample were married, and nearly three-quarters were currently employed.

We elicited subjects' time preferences (*Patient*) using the staircase approach from Falk et al. (2018). In this method, subjects made multiple binary choices between a smaller

[22] The actual median completion time was 13 minutes and 49 seconds.

immediate reward and a larger delayed reward, allowing us to assess their patience. The resulting variable, *Patient*, is a numerical measure of an individual's preference for delayed rewards over immediate gratification. It is scaled from 1 to 32, where higher values indicate a stronger preference for waiting for larger future rewards, and lower values indicate a preference for immediate rewards. A value of 1 means the participant chose today for all options.

Risk preference was assessed using a standard measure, and, consistent with previous empirical studies, subjects exhibited low levels of risk-seeking behaviour. The average financial literacy score was 2.4 out of 3, indicating that most participants demonstrated adequate financial literacy. Risk preferences were measured using adapted experimental tasks (Eckel & Grossman, 2002; Eckel & Grossman, 2008), where subjects selected their preferred gamble from a set of five under three decision environments. Financial literacy was assessed using the "Big Three" questions introduced by Lusardi and Mitchell (2011).

A key variable of interest is *Gamblified investing*. We aimed to examine whether the effects of withdrawal barriers varied among individuals with different levels of gambling intensity. Instead of directly asking participants how often they gamble in their spare time—an approach that may lead to response bias due to social desirability concerns—we employed an indirect measure based on investment behaviours (McNeeley, 2012). Three investment-related questions were designed following the definition of *gamblified* financial products by Newall and Weiss-Cohen (2022), as well as findings from Mills and Nower (2019) and Delfabbro et al. (2021).[23] Investing and gambling share similarities, as both involve risk-taking and are motivated by financial gain. However, unlike traditional investments, gambling does not offer the potential for long-term profits.

Recent evidence suggests that the expected financial advantage of investing does not hold for specific investment products, such as high-frequency stock trading, high-risk derivatives, and cryptocurrencies. These products are highly *gamblified* as they adopt design features commonly seen in gambling, including frequent engagement and the prospect of large, lottery-like wins. Such products typically result in financial losses for

---

[23] Three questions designed for capturing gamblified-investing tendency are available in Appendix 3 Qualtrics Survey.

most investors and disproportionately attract individuals at risk of gambling-related harm. We assigned a *Gamblified Investing* score of 0 to subjects with no history of engagement in these activities, while a score of 3 represented the highest level of involvement. The relatively low average score in our sample suggests that only a limited number of participants engaged in high-intensity gambling-like investment behaviours.

The stratified randomisation process was conducted via Qualtrics. The combination of two manipulations at each stage produced four experimental groups: (1) *Low amount in 'Game' and No task for withdrawal*, (2) *High amount in 'Game' and No task for withdrawal*, (3) *Low amount in 'Game' and Task for withdrawal*, and (4) *High amount in 'Game' and Task for withdrawal*. Table 3.2 presents the balance checks, with the last column reporting p-values from ANOVA tests. The results indicate no significant differences between group means, confirming that randomisation was successful in achieving balance across key characteristics.

*Table 3.2 Balance checks*

| | Low Game + Task free | High Game + Task free | Low Game + Task | High Game + Task | ANOVA p-value |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| *Age* | 35.662 | 34.517 | 37.324 | 34.540 | 0.877 |
| | (12.104) | (11.248) | (14.071) | (10.137) | |
| *Female* | 0.552 | 0.517 | 0.524 | 0.553 | 0.960 |
| | (0.499) | (0.501) | (0.501) | (0.499) | |
| *College* | 0.740 | 0.727 | 0.669 | 0.720 | 0.473 |
| | (0.440) | (0.447) | (0.472) | (0.451) | |
| *Married* | 0.545 | 0.455 | 0.503 | 0.480 | 0.403 |
| | (0.500) | (0.500) | (0.502) | (0.501) | |
| *Employed* | 0.773 | 0.776 | 0.772 | 0.700 | 0.158 |
| | (0.420) | (0.418) | (0.421) | (0.460) | |
| *Patient* | 15.578 | 14.524 | 15.538 | 16.067 | 0.557 |
| | (11.117) | (11.382) | (11.542) | (11.043) | |
| *Risk seeking* | 6.760 | 6.713 | 6.441 | 6.847 | 0.994 |
| | (2.870) | (2.901) | (2.821) | (2.734) | |
| *Financial literacy* | 2.383 | 2.483 | 2.366 | 2.373 | 0.611 |
| | (0.725) | (0.711) | (0.780) | (0.782) | |
| *Gamblified investing* | 0.623 | 0.720 | 0.531 | 0.733 | 0.643 |
| | (0.833) | (0.907) | (0.825) | (0.953) | |
| ***N*** | 154 | 143 | 145 | 150 | 592 |

*Notes*: Means of key demographic variables and characteristics across four groups based on a two-stage design are presented in columns (1) to (4), with standard deviations in parenthese. Column (5) reports the p-values from ANOVA F-tests for mean differences between groups.

# 3.3 Results

## 3.3.1 Evidence of Mental Accounting in Stage 1

We test ***Hypothesis 1*** by comparing gambling behaviour between subjects assigned to the *High Game Balance* (400 credits in the *Game* account) and *Low Game Balance* (100 credits in the *Game* account) groups, while ensuring total wealth remains constant at 500 credits. We focus on two outcome variables: (1) the probability of participating in the game (i.e., placing a positive bet) and (2) the amount wagered by those who chose to participate.

*Figure 3.6 Effects of mental accounting in Stage 1*

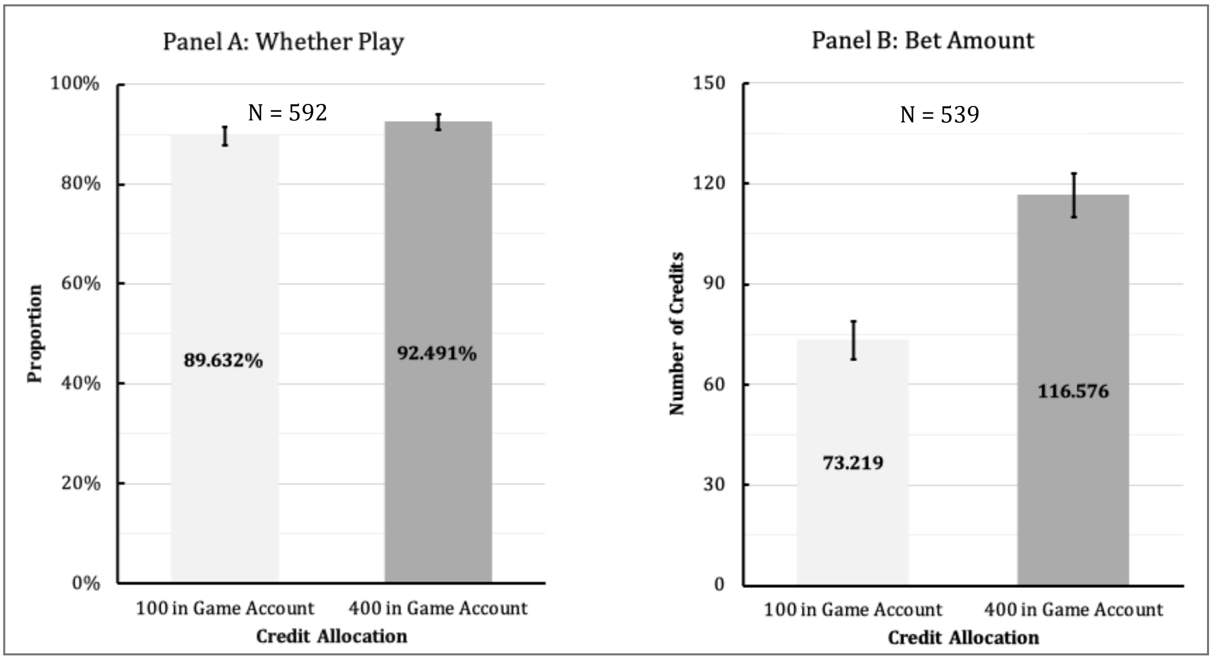



Figure 3.6, Panel A, shows that varying the initial allocation of funds between accounts did not significantly affect participation probability. The proportion of participants engaging in the game was 3.2 percentage points higher among those with 400 credits in the *Game* account compared to those with 100 credits (92.5% vs. 89.6%), but this difference was not statistically significant (two-sample t-test: $p$ = 0.223). While mental accounting can influence whether individuals engage in certain financial activities, its primary effect is typically observed in how money is allocated rather than in the decision to participate.

Panel B of Figure 3.6 provides strong evidence for mental accounting. Among participants who placed a bet, those with a higher balance in the *Game* account wagered an average of 43 more credits than those with a lower balance (116.6 vs. 73.2 credits; two-sample t-test: $p < 0.001$). This difference represents an 8.67% increase in spending, despite both groups having identical total wealth. We also estimated the effect using the following OLS regression specification $y_i = \beta_0 + \gamma HighGame_i + \delta X_i + \epsilon_i$, where $\gamma$ captures the effect of mental accounting on gambling behaviours. $X_i$ includes the demographics as well as other controls such as time preference, risk preference and financial literacy. Table *3.3* yields similar results.

***Result 1:*** *Subjects treated funds in the two accounts differently. Money allocated to the gambling-specific account was more readily spent, leading to increased bet amounts.*

*Table* 3.3 *Regression outputs for Stage 1*

| | Dependent variable | |
|---|---|---|
| | **Whether Play** | **Play Amount** |
| *High Balance in Game* | 0.032 | 41.249 *** |
| | (0.023) | (9.171) |
| *Age* | -0.001 | -0.806 ** |
| | (0.001) | (0.349) |
| *Female* | 0.036 | -32.288 *** |
| | (0.022) | (9.691) |
| *College* | -0.034 | 9.286 |
| | (0.026) | (10.568) |
| *Employed* | 0.054 * | 11.256 |
| | (0.031) | (11.508) |
| *Married* | 0.056 ** | -7.013 |
| | (0.025) | (9.562) |
| *Patience* | -0.001 | 0.174 |
| | (0.001) | (0.424) |
| *Risk seeking* | -0.001 | 1.835 |
| | (0.005) | (1.752) |
| *Financial literacy* | 0.01 | 6.008 |
| | (0.015) | (6.464) |
| *Intercept* | 0.871 *** | 79.39 *** |
| | (0.070) | (23.795) |
| ***N*** | 592 | 539 |

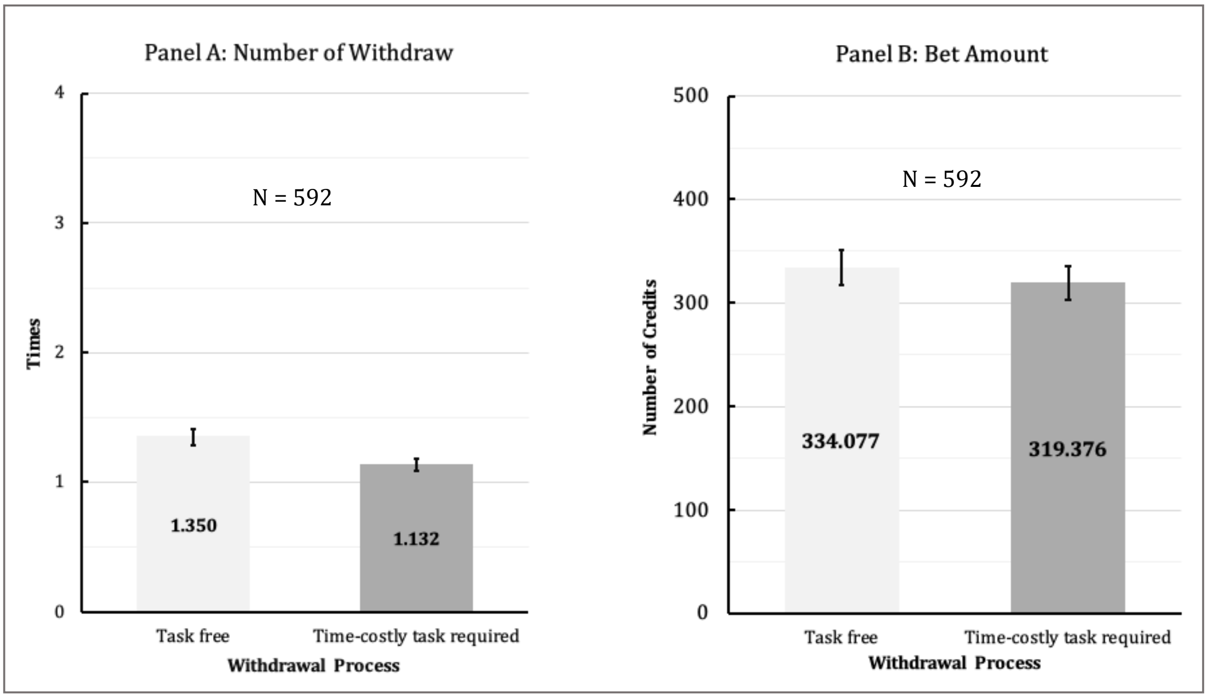
Notes: Standard errors are in parenthese, clustered at the individual level. For regression on the dependent variable Play Amount, we only includes those who chose to participate for both group (i.e. have a greater than 0 amount). *p < 0.1, **p < 0.05, ***p < 0.01.

## 3.3.2 Gambling Behaviours in Stage 2

We next examine the impact of withdrawal barriers on withdrawal and gambling behaviours. The analysis begins by comparing the average number of withdrawals across four rounds between the treatment group (required to complete a time-costly task for withdrawal) and the control group (task-free withdrawal). Figure 3.7, Panel A, shows that introducing a withdrawal barrier significantly reduced withdrawal frequency. On average, subjects in the treatment group withdrew 0.218 times fewer than those in the control group (1.132 vs. 1.35 times; two-sample t-test: $p = 0.006$).

A second key outcome is the amount wagered per round. Figure 3.7, Panel B, compares the average bet size between groups for all players. No significant difference was found between the two groups (334.08 vs. 319.38 credits; two-sample t-test: $p = 0.524$), suggesting that the withdrawal barrier did not directly influence bet size.

*Figure 3.7 Effects of withdrawal barriers in Stage 2*




### *3.3.2.1 Panel Data Analysis*

To better understand decision-making over multiple rounds and improve the precision of our estimates, we converted the data into panel format, yielding a sample size of $N = 2{,}368$. In each round, subjects chose one of three actions:

1. Withdraw funds (transfer from the *Game* to the *Transaction* account)

2. Deposit funds (transfer from the *Transaction* to the *Game* account)
3. Do nothing (leave balances unchanged)

Since these choices are nominal and not ordinal, we employed a multinomial logistic regression to model the log-odds of selecting *Withdraw* $(W)$ or *Deposit* $(D)$ relative to *Inaction* $(I)$:

$$Ln\left(\frac{\pi_i^W}{\pi_i^I}\right) = \beta_0^W + \gamma^W WithdrawalTask_i + \zeta^W BalanceGap_i + \eta^W Round1_i + \delta^W X_i + \epsilon_i^W$$

$$Ln\left(\frac{\pi_i^D}{\pi_i^I}\right) = \beta_0^D + \gamma^D WithdrawalTask_i + \zeta^D BalanceGap_i + \eta^D Round1_i + \delta^D X_i + \epsilon_i^D$$

The parameters of interest are captured by $\gamma^W$ and $\gamma^D$, while controls for demographics, time preference, risk preference and financial literacy are shown as $X_i$.[24] $BalanceGap$ measures the initial difference in account balances between the *Game* and *Transaction* prior to decision making, in thousands of credits. $Round1$ is a binary variable indicating 1 if observation $i$ occurred in the first round. Standard errors of the regression are clustered at the individual level. Since the coefficients estimated in multinomial logistic regression represent changes in log-odds, not actual probabilities, it can be difficult to interpret directly. Thus, average marginal effects, which translate the changes in log-odds into changes in probabilities, are calculated and presented in Table *3.4*.

For every additional thousand credits (equivalent to $1 AUD) in the *Game* account relative to the *Transaction* account, the likelihood of withdrawing from the *Game* account increases by 10.5 percentage points ($p < 0.001$). This suggests that a higher proportion of funds in the *Game* account, compared to the total balance at the time of decision-making, increases the likelihood of withdrawal. This finding aligns with the common expectation that individuals are more inclined to transfer larger balances from online gambling accounts

[24] Note that we do not control for gamblified investing behaviours in our main regression specifications for both Stage 1 and Stage 2. This is because such indirect measure of gambling intensity is not as standardised and robust as the others. However, we do incorporate this variable in Section 4.3.2 when analysing the heterogeneity in treatment effects.

or non-bank payment apps to traditional bank accounts for enhanced security and satisfaction.

No significant effects are observed for education level, employment status, time preferences, or financial literacy. However, as expected, a higher level of risk-seeking behaviour is associated with a lower likelihood of withdrawal ($p = 0.002$), while females are more likely to withdraw than males ($p = 0.005$). Additionally, subjects are significantly more likely to withdraw funds during the first round of the game than in subsequent rounds ($p < 0.001$).

*Table* 3.4 *Average marginal effects on the probabilities of observing withdrawal, inaction and deposit*

| | **Withdrawal** | **Inaction** | **Deposit** |
|---|---|---|---|
| | (1) | (2) | (3) |
| *Task for withdrawal* | -0.053 ** | 0.117 *** | -0.065 *** |
| | (0.021) | (0.023) | (0.020) |
| *Balance gap between 'Game' & 'Transaction'* | 0.105 *** | -0.049 ** | -0.056 *** |
| | (0.019) | (0.024) | (0.019) |
| *First round* | 0.259 *** | -0.250 *** | -0.009 |
| | (0.016) | (0.025) | (0.020) |
| *Age* | -0.001 | 0.002 | -0.001 |
| | (0.001) | (0.001) | (0.001) |
| *Female* | 0.062 *** | -0.017 | -0.046 ** |
| | (0.022) | (0.024) | (0.020) |
| *College* | 0.000 | -0.006 | 0.007 |
| | (0.023) | (0.027) | (0.023) |
| *Employed* | -0.030 | 0.015 | 0.015 |
| | (0.025) | (0.028) | (0.026) |
| *Patient* | 0.000 | 0.001 | -0.001 |
| | (0.001) | (0.001) | (0.001) |
| *Risk seeking* | -0.012 *** | 0.002 | 0.010 *** |
| | (0.004) | (0.004) | (0.003) |
| *Financial literacy* | 0.014 | -0.010 | -0.004 |
| | (0.016) | (0.017) | (0.015) |
| ***N*** | | | 2,368 |

*Notes*: Standard errors are in parenthese, clustered at the individual level. The sample size of 2,368 is obtained by pooling observations from all four rounds for the same subjects in Stage 2. The average marginal effects of the coefficients in the multinomial logistic regression are reported. Each column shows the average change in the probability of choosing a particular action reagding the "Game" account in Stage 2 across all observations. For example, column (1) presents the average change in the probability of choosing to withdraw funds before each round of the game in Stage 2. *p < 0.1, **p < 0.05, ***p < 0.01.

As shown in the first column of Table *3.4*, the marginal treatment effect on the probability of withdrawal is, on average, 5.3 percentage points across all observations ($p = 0.011$). This indicates that when subjects are required to complete a time-costly task to withdraw funds, their likelihood of withdrawing decreases. Since unwithdrawn funds are automatically used in the game by default at this stage, withdrawing effectively represents an active decision to reduce the amount available for gambling.

Our experimental results also reveal a significant decrease in the probability of deposit for treated subjects, as shown in column (3) ($p < 0.01$). This suggests that participants are less inclined to add funds to the *Game* account—an action akin to increasing their bet—when no task is required to withdraw.

Given the simultaneous decline in both withdrawal and deposit probabilities, our analysis focuses on the marginal treatment effect on inaction to provide a clearer interpretation of gambling behaviour. In our experiment, inaction signifies continued gambling participation. When subjects neither withdraw nor deposit funds, these funds remain in the gambling-specific account. Consequently, the default participation mechanism increases the likelihood of passive gambling engagement.

Since total probability remains constant, reductions in deposit and withdrawal probabilities are reallocated to inaction. As a result, the observed 11.8 percentage point increase in inaction probability in column (2) precisely reflects the combined reduction in deposit and withdrawal probabilities ($p < 0.01$). As discussed, a higher probability of inaction implies increased retention of funds in the gambling-specific account when a withdrawal task is introduced. This retention, in turn, fosters greater gambling engagement and increases the risk of financial loss:

***Result 2****: When facing barriers in the withdrawal process, individuals are significantly more likely to leave funds unwithdrawn in the gambling-specific account, leading to continued gambling participation and a higher risk of financial loss.*

#### *3.3.2.2 Panel Data Analysis Excluding the First Round*

Given the significant impact of $Round1$ on withdrawal decisions, we consider the possibility that subjects in the treatment group may only become fully aware of the "non-pecuniary cost" of withdrawing funds after the first round of the game. This suggests that the first round may serve as a learning phase, with informed decisions occurring in subsequent rounds. To test the robustness of the results from Section 3.3.1, we exclude observations from the first round and re-estimate the parameters using a multinomial logistic regression, yielding an updated sample size of $N$ = 1,776:

$$Ln\left(\frac{\pi_i^W}{\pi_i^I}\right) = \beta_0^W + \gamma^W WithdrawalTask_i + \zeta^W BalanceGap_i + \delta^W X_i + \epsilon_i^W$$

$$Ln\left(\frac{\pi_i^D}{\pi_i^I}\right) = \beta_0^D + \gamma^D WithdrawalTask_i + \zeta^D BalanceGap_i + \delta^D X_i + \epsilon_i^D$$

The average marginal treatment effects on the probabilities of withdrawal, inaction, and deposit are calculated and presented in Table *3.5*, columns (1), (3), and (5), respectively.

The results remain consistent with the previous findings. Males and more risk-seeking individuals continue to exhibit a lower likelihood of withdrawing funds compared to their counterparts. Additionally, a lower balance in the *Game* account relative to the *Transaction* account at the time of decision-making is consistently associated with a reduced probability of withdrawal.

Participants are significantly more likely to leave funds in the *Game* account unwithdrawn when faced with the time-costly withdrawal task ($p$ = 0.005). The effect size observed without the inclusion of the first round is slightly larger than the main result, suggesting that the "non-pecuniary barrier" becomes more salient to subjects after experiencing the first round.

To further examine the role of prior game outcomes, we introduce two new binary variables, *Win* and *Lose*, into the regression specification outlined in Section 3.3.2. The variable *Win* equals 1 if the subject received a multiplier greater than 1 in the previous round, while *Lose* equals 1 if the subject received a multiplier smaller than 1. The results, presented in Table 4, columns (2), (4), and (6), confirm that the main findings remain unchanged.

Specifically, we find that the probability of withdrawal increases significantly when subjects experience a win in the previous round and decreases following a loss. In particular, subjects who win more credits than their initial bet amount exhibit a 19.1 percentage point increase in the likelihood of withdrawal ($p$ < 0.001). This behaviour may be driven by a desire to secure gains, as withdrawing funds reduces the risk of losing them in subsequent gambling rounds.

*Table 3.5 Average marginal effects excluding the first round*

| | Withdrawal | Inaction | Deposit | Withdrawal | Inaction | Deposit |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| *Task for withdrawal* | -0.064 *** | 0.136 *** | -0.072 *** | -0.066 *** | 0.136 *** | -0.071 *** |
| | (0.023) | (0.027) | (0.022) | (0.021) | (0.027) | (0.021) |
| *Balance gap between 'Game' & 'Transaction'* | 0.092 *** | -0.040 | -0.052 *** | 0.051 *** | -0.035 | -0.016 |
| | (0.018) | (0.026) | (0.020) | (0.018) | (0.027) | (0.021) |
| *Age* | -0.002 * | 0.002 | 0.000 | -0.002 | 0.002 | 0.000 |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| *Female* | 0.047 * | 0.004 | -0.050 ** | 0.042 * | 0.003 | -0.045 ** |
| | (0.025) | (0.029) | (0.023) | (0.023) | (0.029) | (0.023) |
| *College* | -0.015 | -0.001 | 0.016 | -0.012 | 0.001 | 0.011 |
| | (0.025) | (0.033) | (0.027) | (0.023) | (0.033) | (0.025) |
| *Employed* | -0.016 | 0.016 | -0.001 | -0.010 | 0.014 | -0.003 |
| | (0.027) | (0.035) | (0.028) | (0.026) | (0.035) | (0.027) |
| *Patient* | -0.001 | 0.002 | -0.001 | -0.001 | 0.002 | -0.001 |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| *Risk seeking* | -0.014 *** | 0.001 | 0.013 *** | -0.012 *** | 0.001 | 0.010 *** |
| | (0.004) | (0.005) | (0.004) | (0.004) | (0.005) | (0.004) |
| *Financial literacy* | 0.023 | -0.015 | -0.008 | 0.022 | -0.015 | -0.007 |
| | (0.017) | (0.021) | (0.017) | (0.016) | (0.021) | (0.016) |
| *Won in the last round* | | | | 0.191 *** | -0.124 *** | -0.067 ** |
| | | | | (0.024) | (0.032) | (0.028) |
| *Lost in the last round* | | | | -0.051 * | -0.081 ** | 0.132 *** |
| | | | | (0.030) | (0.033) | (0.024) |
| ***N*** | | | | | | 1,776 |

*Notes* : Standard errors are in brackets, clustered at the individual level. The sample size of 1,776 is obtained by pooling observations from the last three rounds of Stage 2 for the same subjects, excluding the first round. The average marginal effects of the coefficients in the multinomial logistic regression are reported. Columns (1)-(3) have the same specification as the regression reported in Table 3, except for the automatic omission of the dummy variable *First round* . Columns (4)-(6) add two additional dummy variables, *Won in the last round* and *Lost in the last round* , to the set of controls, along with the automatic omission of *First round* . All results represent the average change in the probability of choosing a specific action regarding the "Game" account before each round of the game in Stage 2 across the last three rounds. *p < 0.1, **p < 0.05, ***p < 0.01.

## 3.3.3 Heterogenous Treatment Effects

Since gambling is a topic with strong policy relevance, understanding the differentiated responses between groups is crucial when designing effective interventions for withdrawal or deposit restrictions. In this section, we further explore the heterogeneity of mental accounting and the effect of withdrawal barriers on gambling behaviour. Each panel from A to C in Table 3.6 presents the estimated treatment effects for groups split by gender, gamblified investing experience, and age at 50, respectively. Whenever applicable, differences in treatment effects between sub-samples are also reported.

### *3.3.3.1 Gender*

We begin by examining the heterogeneous effects of mental accounting by gender. We observe an increase in betting expenditure for both men and women who chose to participate in the game in Stage 1. As shown in column (1) of Panel A, men spend an estimated 45.94 more credits when the gambling-specific account holds relatively larger funds, while women spend 35.08 more credits. Although the estimated effect of mental accounting for men is 10.86 credits higher than for women, the difference is not statistically significant.

Meanwhile, columns (3) to (5) suggest that men are more sensitive to the added friction in withdrawal than women. Despite the lack of significant effects on betting expenditure in Stage 2, requiring a time-costly task for withdrawal significantly discourages both men and women from adjusting their gambling-specific account balance ($p < 0.01$). We find an increase in inaction of 14.7 percentage points for men—almost twice as large as for women. Given that any funds left in the account are automatically used in the game, a higher increase in inaction for men suggests they are more likely to have their current funds automatically reinvested, potentially exposing them to a greater compounded risk. Moreover, men's overall inaction is driven by reductions in both withdrawals ($p < 0.05$) and deposits ($p < 0.01$), whereas women only show a significant change in deposit probability. The absence of a significant change in women's withdrawal probability suggests they are less deterred from withdrawing, even when a task is required. Women's tendency to maintain their pattern of withdrawing funds as needed may indicate that their gambling decisions are more planned, suggesting greater financial self-control.

*Table 3.6 Heterogenous treatment effects on the Stage 1 and Stage 2 outcomes*

| | Stage 1: Effect of Mental accounting | | Stage 2: Effect of Withdrawal barriers | | | | |
|---|---|---|---|---|---|---|---|
| | Bet amount | ***N*** | Withdrawal | Inaction | Deposit | Bet amount | ***N*** |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| *Panel A: Gender* | | | | | | | |
| Female | 35.080 *** | 293 | -0.039 | 0.089 *** | -0.051 ** | -6.511 | 318 |
| Male | 45.942 *** | 246 | -0.067 ** | 0.147 *** | -0.080 *** | -1.168 | 274 |
| Δ | 10.862 | **539** | | | | 5.343 | **592** |
| *Panel B: Gamblified investing behaviours* | | | | | | | |
| Never | 62.129 *** | 321 | -0.060 ** | 0.125 *** | -0.065 *** | -3.301 | 355 |
| Attempted | 7.632 | 218 | -0.051 | 0.120 *** | -0.068 ** | -3.437 | 237 |
| Δ | 54.498 *** | **539** | | | | 12.640 | **592** |
| *Panel C: Mature workforce* | | | | | | | |
| Aged under 50 | 41.833 *** | 473 | -0.050 ** | 0.107 *** | -0.057 *** | -10.225 | 518 |
| Aged 50 and over | 49.322 *** | 66 | -0.118 * | 0.205 *** | -0.087 | 52.061 ** | 74 |
| Δ | 7.489 | **539** | | | | 62.286 *** | **592** |

*Notes* : Each panel (A–C) displays the estimated treatment effects for groups split by gender, gamblified investing experience, and age at 50, respectively. Within each panel, the first two rows show estimates for the corresponding sub-samples, while the third row (Δ) presents the estimated difference in effects from the interaction model using the full sample. Column (1) reports the effects of having a relatively higher balance in the "Game" account on the bet amount, that is, the effect of mental accounting measured in Stage 1. Column (2) shows the sample size used for these estimates, which includes only observations with a non-zero bet amount in Stage 1. Columns (3) to (5) report the average marginal treatment effects estimated from the multinomial logistic regression, using pooled data from all four rounds in Stage 2. These values represent the average change in the probability of selecting a specific action regarding the "Game" account before each round. Column (6) reports the average bet amount over the four rounds in Stage 2, and Column (7) provides the corresponding sample size. Significance levels are indicated as follows: *p < 0.1, **p < 0.05, ***p < 0.01.

#### *3.3.3.2 Gamblified Investing Experience*

It is important to account for individuals' gambling intensity and experience when examining mental accounting bias and behavioural changes in gambling. As mentioned in Section 3.3.1, we use experience with gamblified financial products as a proxy variable for gambling intensity and experience. According to Newall and Weiss-Cohen (2022), gamblified financial products such as derivatives and cryptocurrencies promote frequent trading and the allure of gambling-like big wins. These design elements can lead to financial losses for most investors and attract individuals vulnerable to gambling-related harm. Gamblified financial products closely resemble online gambling platforms. Participants who have engaged in gamblified investing are generally more experienced gamblers and likely gamble more often in real life.

Although prior experience with gamblified financial products does not seem to result in different response patterns to withdrawal frictions, it plays a crucial role in shaping mental accounting, as indicated by results in column (1) of Panel B of Table 3.6. We observe higher gambling expenditure among participants who have never invested in gamblified products. For inexperienced participants, having more funds allocated to the gambling-specific account dramatically increases their bet sizes by 62.13 credits—equivalent to a sizable increase of 12.43%, given the total wealth of 500 credits ($p < 0.01$).

This suggests they may perceive funds in the gambling-specific account as "house money," leading to looser budgeting and higher risk-taking. In contrast, for those with experience in gamblified investing, the allocation of funds between their gambling-specific account and their bank account does not significantly alter their bet amounts. The observed difference of 55 credits in gambling expenditure suggests that inexperienced gamblers may be more susceptible to mental accounting bias ($p < 0.01$). One possible explanation is that exposure to these highly gamblified financial products may help individuals develop strategies to mitigate mental accounting bias, whereas inexperienced individuals remain more affected by the allocation of wealth, leading to greater spending.

#### *3.3.3.3 Age at 50*

Finally, considering that younger and older individuals often differ in risk tolerance, financial priorities, and cognitive capacities that influence online gambling behaviour, we examine differential treatment effects by splitting the sample at age 50. Digital literacy is particularly relevant in the context of online gambling, as individuals' technological familiarity can significantly affect how they interact with online gambling platforms. Since individuals aged 50 and over are the key demographic group targeted by the Australian Government's digital literacy and online safety initiatives, we follow convention and use age 50 as the cutoff.[25]

We find that participants aged under 50 and those aged 50 and above respond similarly to mental accounting bias, showing an increase in gambling expenditure when the balance is relatively higher in the gambling-specific account ($p < 0.01$). However, added friction in withdrawal appears to have a stronger behavioural impact on older participants. Those aged 50 and over exhibit a 20.5 percentage point increase in inaction ($p < 0.01$), compared to a 10.7 percentage point increase for those under 50 ($p < 0.01$). Older individuals are substantially more likely to leave their account balance unchanged when faced with friction, potentially due to lower digital familiarity or a greater sensitivity to additional procedural steps.

For the older group, the friction leads to an 11.8 percentage point reduction in withdrawals ($p < 0.1$), whereas the younger group shows a smaller 5 percentage point reduction ($p < 0.01$). This suggests that effort costs are more likely to deter older participants from withdrawing funds from their gambling-specific accounts ($p < 0.1$). Since the change in deposits among the older group is not statistically significant, the inaction appears to be primarily driven by reluctance to withdraw rather than a deliberate reduction in deposits.

The treatment effect on average bet amount is observed only in the older group, with an increase of 52.06 credits ($p < 0.05$). This implies that additional friction not only discourages active management through withdrawals and deposits but also results in higher amounts being wagered automatically in subsequent rounds as funds remain in the gambling account.

Though the number of older gamblers in the sample is relatively small—which might raise concerns about statistical power—the fully interactive regression using the full sample shows a significant difference in treatment effects on average bet amounts between the two age groups. Specifically, older participants aged 50 and over placed 62.29 more

[25] The Australian government initiative "Be Connected" launched in October 2017 aims to improve the digital literacy for older Australians aged 50 years and older.

credits on average (an increase of 6.23% relative to initial wealth) when faced with withdrawal friction compared to their younger counterparts ($p < 0.01$). This robust finding suggests that the observed effect among older participants is not merely a byproduct of a small sample but reflects a true difference in how withdrawal friction influences betting behaviour across age groups.

# 3.4 Conclusion

Online gambling platforms in Australia require users to maintain gambling-specific accounts that serve as virtual wallets for credited winnings, yet impose excessive friction on withdrawal requests. This design makes it easy to start gambling but challenging to stop, as the barriers hinder the removal of funds. Our study demonstrates that these features contribute to higher gambling participation and larger bets. Specifically, we show that allocating funds to dedicated gambling accounts increases betting expenditure, highlighting a mental accounting effect in this context. The experimental findings indicate that when withdrawal friction is introduced, individuals are more likely to do nothing—leaving funds unwithdrawn—and thereby inadvertently increasing their gambling activity. Moreover, the results suggest that interventions aimed at reducing gambling-related harm should consider differences in experience with gamblified financial products and age. Inexperienced gamblers appear more susceptible to mental accounting, leading to greater spending, while withdrawal friction most effectively influences the gambling behaviour of older individuals aged 50 and over. Importantly, this friction seems to reinforce pre-existing behavioural tendencies rather than inducing entirely new changes.

The simultaneous reduction in both withdrawals and deposits observed in our experiment merit further discussion. One interpretation is that the estimated reduction in the probability of deposits may be overestimated, while the effect on withdrawals is underestimated under experimental conditions. In our experiment, participants are required to make rapid and frequent deposit and withdrawal decisions, which accentuates the salience of the barriers and results in reduced deposit behaviour. In contrast, real-life withdrawal barriers primarily affect withdrawal decisions, so our experimental design might inflate the deposit reduction. Furthermore, while unwithdrawn funds in our experiment are exposed to the risk of loss, funds in actual gambling accounts are accessible without immediate risk, making participants even more reluctant to withdraw.

Our results may also reflect the presence of two types of agents. Naïve, myopic agents may refrain from withdrawing due to friction and thus continue habitual gambling with funds left in the account; for them, removing the barrier could increase the likelihood of withdrawal, which might be beneficial. In contrast, sophisticated agents may already avoid depositing funds because they anticipate the friction, meaning that removing the barrier could inadvertently lead to increased gambling activity through more frequent deposits.

Given that the general public is likely predominantly composed of naïve agents, we argue that, from a policy perspective, eliminating withdrawal barriers and the segregation of funds into gambling-specific accounts could be an effective, low-cost regulatory approach to promote safer gambling behaviour. For example, ensuring that winnings are directly transferred to bank accounts without friction could mitigate excessive gambling participation and its associated harms. Alternatively, making withdrawal barriers more salient in real-life contexts might encourage individuals to consider the costs before depositing and gambling further. Our findings are not only applicable to online gambling platforms but can also offer insights for other online services that employ virtual accounts and asymmetric designs, such as petrol cards and gym memberships.

# Appendix 3 Qualtrics Survey

Please click here to open the full survey PDF via Google drive.

# Reference


Auer, M., & Griffiths, M. D. (2023). Nudging online gamblers to withdraw money: The impact of personalized messages on money withdrawal among a sample of real-world online casino players. *Journal of Gambling Studies*, 1-18.

Auer, M., Hopfgartner, N., & Griffiths, M. D. (2020). The effects of voluntary deposit limit-setting on long-term online gambling expenditure. *Cyberpsychology, Behavior, and Social Networking*, 23(2), 113-118.

Barber, B. M., & Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *The Quarterly Journal of Economics*, 116(1), 261-292.

Beatty, T. K., Blow, L., Crossley, T. F., & O'Dea, C. (2014). Cash by any other name? Evidence on labeling from the UK Winter Fuel Payment. *Journal of Public Economics*, 118, 86-96.

The Behavioural Insights Team. (2021). Applying behavioural insights to design better safer gambling tools. Part 1: Anchoring: Commitment devices.

Camerer, C. F., & Hogarth, R. M. (1999). The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework. *Journal of Risk and Uncertainty*, 19(1/3), 7–42.

Croson, R., & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, 47(2), 448-474.

Cueva, C., Iturbe-Ormaetxe, I., Ponti, G., & Tomás, J. (2019). Boys will still be boys: Gender differences in trading activity are not due to differences in (over) confidence. *Journal of Economic Behavior & Organization*, 160, 100-120.

Delfabbro, P., King, D., Williams, J., & Georgiou, N. (2021). Cryptocurrency trading, gambling and problem gambling. *Addictive Behaviors*, 122, 107021.

Eckel, C. C., & Grossman, P. J. (2002). Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and Human Behavior*, 23(4), 281-295.

Eckel, C. C., & Grossman, P. J. (2008). Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. *Journal of Economic Behavior & Organization*, 68(1), 1-17.

Edson, T. C., Tom, M. A., & LaPlante, D. A. (2021). Examining player engagement with and attitudes toward a gambling play management system. *Journal of Gambling Studies*, 1-22.

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., & Sunde, U. (2018). Global evidence on economic preferences. *The Quarterly Journal of Economics*, 133(4), 1645-1692.

Gambling Commission. (n.d.). Online games design and reverse withdrawals. https://www.gamblingcommission.gov.uk/print/online-games-design-and-reverse-withdrawals

Gourville, J. T., & Soman, D. (1998). Payment depreciation: The behavioral effects of temporally separating payments from consumption. *Journal of Consumer Research*, 25(2), 160-174.

Harris, A., & Griffiths, M. D. (2017). A critical review of the harm-minimisation tools available for electronic gambling. *Journal of Gambling Studies*, 33, 187-221.

Hastings, J. S., & Shapiro, J. M. (2013). Fungibility and consumer choice: Evidence from commodity price shocks. *Quarterly Journal of Economics*, 128(4), 1449-1498.

Heath, C., & Soll, J. B. (1996). Mental budgeting and consumer decisions. *Journal of Consumer Research*, 23(1), 40-52.

Hing, N., Cherney, L., Blaszczynski, A., Gainsbury, S. M., & Lubman, D. I. (2014). Do advertising and promotions for online gambling increase gambling consumption? An exploratory study. *International Gambling Studies*, 14(3), 394-409.

Ivanova, E., Magnusson, K., & Carlbring, P. (2019). Deposit limit prompt in online gambling for reducing gambling intensity: A randomized controlled trial. *Frontiers in Psychology*, 10, 639.

Kooreman, P. (2000). The labeling effect of a child benefit system. *American Economic Review*, 90(3), 571-583.

Letts, C. (2018, November 20). Australians world's biggest gambling losers. ABC News. https://www.abc.net.au/news/2018-11-20/australians-worlds-biggest-gambling-losers/10495566

Levitt, S. D., & List, J. A. (2007). What Do Laboratory Experiments Measuring Social Preferences Reveal about the Real World? *The Journal of Economic Perspectives*, 21(2), 153–174.

Lusardi, A., & Mitchell, O. S. (2011). Financial literacy around the world: An overview. *Journal of Pension Economics & Finance*, 10(4), 497-508.

McNeeley, S. (2012). Sensitive issues in surveys: Reducing refusals while increasing reliability and quality of responses to sensitive survey items. In L. Gideon (Ed.), *Handbook of survey methodology for the social sciences* (pp. 377-396). Springer.

Mills, D. J., & Nower, L. (2019). Preliminary findings on cryptocurrency trading among regular gamblers: A new risk for problem gambling? *Addictive Behaviors*, 92, 136-140.

Newall, P. W., & Rockloff, M. J. (2022). Promoting safer gambling via the removal of harmful sludge: A view on how behavioural science's 'nudge' concept relates to online gambling. *Addiction*, 117(5), 1398-1404.

Newall, P. W., & Weiss-Cohen, L. (2022). The gamblification of investing: How a new generation of investors is being born to lose. *International Journal of Environmental Research and Public Health*, 19(9), 5391.

Prelec, D., & Loewenstein, G. (1998). The red and the black: Mental accounting of savings and debt. *Marketing Science*, 17(1), 4-28.

Samuelson, W., & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty*, 1(1), 7-59.

Schindler, S., & Pfattheicher, S. (2017). The frame of the game: Loss-framing increases dishonest behavior. *Journal of Experimental Social Psychology*, 69, 172–177.

Schubert, R., Brown, M., Gysler, M., & Brachinger, H. W. (1999). Financial decision-making: Are women really more risk averse? *American Economic Review*, 89(2), 381-385.

Soman, D. (2001). Effects of payment mechanism on spending behavior: The role of rehearsal and immediacy of payments. *Journal of Consumer Research*, 27(4), 460-474.

Sui, L., Sun, L., & Geyfman, V. (2021). An assessment of the effects of mental accounting on overspending behaviour: An empirical study. *International Journal of Consumer Studies*, 45(2), 221-234.

Sunden, A. E., & Surette, B. J. (1998). Gender differences in the allocation of assets in retirement savings plans. *The American Economic Review*, 88(2), 207-211.

Thaler, M. (2021). Gender differences in motivated reasoning. *Journal of Economic Behavior & Organization*, 191, 501-518.

Thaler, R. H. (1985). Mental accounting and consumer choice. *Marketing Science*, 4(3), 199-214.

Thaler, R. H. (1999). Mental accounting matters. *Journal of Behavioral Decision Making*, 12(3), 183-206.

Thaler, R. H., & Benartzi, S. (2004). Save More Tomorrow™: Using behavioral economics to increase employee saving. *Journal of Political Economy*, 112(S1), S164-S187.

Thaler, R. H., & Johnson, E. J. (1990). Gambling with the house money and trying to break even: The effects of prior outcomes on risky choice. *Management Science*, 36(6), 643-660.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211(4481), 453-458.