Ethical and Responsible Behavior in Applied Empirical Research: Four Essays on Academic Practices in the Social Sciences

Von der Fakultät für Wirtschaftswissenschaften der Rheinisch-Westfälischen Technischen Hochschule Aachen zur Erlangung des akademischen Grades eines Doktors der Wirtschafts- und Sozialwissenschaften genehmigte Dissertation

vorgelegt von

Gernot Pruschak

Berichter: Prof. Dr. rer. pol. Christian Hopp

Univ.-Prof. Dr. rer. pol. Frank Piller

Tag der mündlichen Prüfung: 5. Februar 2021

Diese Disseration ist auf den Internetseiten der Universitätsbibliothek online verfügbar.

Ethical and Responsible Behavior in Applied Empirical Research: Four Essays on Academic Practices in the Social Sciences

Dissertation approved by the School of Business and Economics of RWTH Aachen University to obtain the degree of Doctor of Economics and Social Sciences

submitted by

Gernot Pruschak

Reviewers: Prof. Dr. rer. pol. Christian Hopp

Univ.-Prof. Dr. rer. pol. Frank Piller

Date of the oral exam: 5th of February 2021

This dissertation is accessible online via the website of the university library.

Gernot Pruschak, MSc. BSc.

Ethical and Responsible Behavior in Applied Empirical Research:
Four Essays on Academic Practices in the Social Sciences

Dissertation approved by the School of Business and Economics of RWTH Aachen University to obtain the degree of Doctor of Economics and Social Sciences, February 5 2021.

Research Area for Technology, Innovation, Marketing & Entrepreneurship (TIME)
School of Business and Economics
RWTH Aachen University

E-Mail: pruschak@time.rwth-aachen.de

Abstract

In a world filled with *fake news* and *alternative facts*, public trust in science is of utmost importance. Yet scandals like the cases of Diederik Stapel and Ulrich Lichtenthaler have questioned the integrity of scholars and their research results. To address this issue, several scientists investigated academic (mal)practices like plagiarism, HARKing, authorship misuse and data flexibility. The results were devastating and ignited a credibility crisis, especially in the social sciences. Fortunately, we already can see the light at the end of the tunnel as editors, publishers, research societies and universities have started to introduce techniques and infrastructure that ensure ethical and responsible scholarly behavior. For example, artificial intelligence has enabled plagiarism detection software to not only check for copy-pasting but also for content and reference similarities. Moreover, more and more journals motivate or sometimes even require researchers to pre-register their research hypotheses prior to data collection and/or data analysis to prevent HARKing. In the life sciences, contribution disclosure statements force authors to transparently report the contributions of each researcher involved in a research project. In the social sciences, several articles and editorials highlighted that ensuring replicability by means of transparent reporting and data sharing allows detecting and overcoming flexible and questionable data handling practices.

This thesis builds upon the existing body of literature and provides guidance for those academic (mal)practices that have been covered only rudimentarily in the social sciences. Essay 1 addresses the issues of ghost and honorary authorship, the two most infamous forms of authorship misuses. We show that there exists a strong mismatch between actual and hypothetical authorship assignments: While most social scientists in our survey assigned authorship correctly in three hypothetical scenarios, more than every third paper contained at least one honorary author. We conclude that motivational factors like hierarchical pressure force scholars to include low-contributing supervisors and colleagues in their author lists. To overcome this issue, we call for social science journals to follow their counterparts in the life sciences by enforcing contribution disclosure and implementing whistleblowing platforms.

Essay 2 asks whether scientific collaborations and authorship teams differ across academic disciplines, geographical regions, working experience and job position. The results indicate that the distribution of multi-authored papers varies substantially even within the social sciences. Furthermore, we highlight that language barriers and infrastructural challenges possess effects

on academic collaborations. These findings contain important implications for search and tenure procedures as committees must bear these differences in mind when comparing applicants with different backgrounds.

Essay 3 touches upon the need of promoting replicability and replications. We provide a hands-on step-by-step guide on how to conduct a rigor and robust replication. More specifically, we start by directly replicating the results of Kuhn and Weinberger (2005) showing that, for white men, leadership positions in high school correspond to higher wages eleven years later. We then assess the causality of this effect by employing propensity-score matching and three different forms of instrumental variable techniques. Moreover, we move beyond the original sample and investigate the effects for white females and non-whites. We further include data from a follow-up study conducted 50 years later. Our findings highlight that team captainship combined with club presidency induces higher wages eleven and 50 years later for white men, but this effect does not appaer consistently for white females or non-whites. Therefore, leadership interventions should recognize the leadership skills that are already developed in individuals and identify those areas that are in need of further development. In doing so, it is important to be cognizant of diverse aspirations to lead for individuals of different gender and ethnicities.

The thorough analyses in Essay 3 were only possible because we had access to the data employed in the original article. Unfortunately, Essay 4 shows that this constitutes rather an exception than the norm. Investigating data sharing among innovation management researchers, we find that only about one third of their datasets are publicly available. We specifically focus on innovation management researchers because we originally expected them to be specifically prone to open data due to them advocating the advantages of openness to firms and other stakeholders. Our results indicate that the identified personal incentives to open data sharing might not outweigh the burden of open data places on individual researchers. Consequently, we call for academic impeti that give more credit to data sharing and for journal policies enforcing data sharing as a mandatory requirement for publication.

Overall, this thesis highlights that scholars, journals, research societies and universities must change their habits, incentives and conducts especially in the areas of scientific authorship and data practices to secure scientific integrity. Therefore, we outline detailed solutions for guiding this change to ensure ethical and responsible behavior in applied empirical research. This is the only way we can regain public trust in science.

Dedication

For Beatrix, Wilfried and Marlene.

Thank you for your never-ending encouragement, patience and love.

Acknowledgements

I want to sincerely and deeply thank Prof. Christian Hopp for acquiring me and giving me the opportunity to conduct my PhD studies at the RWTH Aachen University. Your limitless support, wisdom and working attitude made you a perfect role model, much valued mentor and highly important collaborator. I had the honor to work with you on various research projects including three essays published in this thesis and all projects were enlightening and skill enhancing. Together with Prof. Christian Hopp, I want to thank Prof. Frank Piller for their numerous efforts and valuable advices as my supervisors and reviewers. My sincere gratitude also goes to Jermain Kaminski who was not only a highly esteemed and appreciated doctoral colleague at the RWTH Aachen University but also greatly supported me by answering all my administrative and formatting question for this dissertation. Moreover, I am lucky to call the distinguished scholars Prof. Christian Hopp, Prof. Frank Piller, Jermain Kaminski and Prof. Gloria Barczak co-authors in the last essay of this thesis. Their contributions and comments enhanced this thesis substantially.

I am extremely grateful for all my supporters at the RWTH Aachen University. Specifically, I would like to sincerely thank Prof. Torsten-Oliver Salge and Prof Daniel Wentzel for joining my dissertation committee. Moreover, my earnest gratitude goes to all colleagues, especially Alexander Speil and Doris Torres for sharing office space, time and fun; Irina Liebing for her administrative support during my time in NRW; Leona Brust, Ekatarina Korneeva, Ruth van Bracht and all other colleagues with whom I attended workshops, seminars and conferences; Dr. Georg Fischer on behalf of all external lecturers who inspired me with their wisdom, knowledge and logic; All students I taught and worked with and especially the student who submitted a plagiarized paper in the bachelor seminar *The Economy of Science* in the winter semester 2017/18 because this made ethics in science my dissertation topic.

Going back and forth from my time at the RWTH Aachen University, I am sincerely grateful for all the years I spent working at the University of Vienna. I want to genuinely thank Prof. Fabel and Prof. Vetschera who gave me the opportunity to get a taste of academic life when they hired me as student assistant in 2014. In fact, without them I would have probably never ended up in academia. Furthermore, Prof. Fabel and Prof. Vetschera also acquired me after my working RWTH Aachen University contract had ended thus giving me the chance to finish my PhD studies at a research institution. In addition to thanking them for leading me and keeping me on the academic path, I am also extremely grateful that they let me present my research projects in their research seminar. Their comments and inputs on earlier versions of the essays included in this thesis significantly enhanced the whole dissertation. Moreover, I want to thank al my colleagues at the University of Vienna, specifically Aysegül Engin, Oksana Galak, Christian Hamböck, Cigdem Keles, Dana Minarikova, Haris Seferagic, Prof. Josef Windsperger and Yingchao Zhang, for their inputs, comments and collaborations as well as our interesting and entertaining office, kitchen and lunch talks and afterwork drinks. Furthermore, I want to thank Elitsa Nestorova and Elke Pendl for their tremendous support in the last six years. Struggling through the administrative hurdles of a public institution was not always successful but always fun with them. My deep gratitude also goes to Iurii Berlach, Alina Chugreeva, Mounir Derboghossian Özge Gürcay, Veronika Keuschnigg, Matthias Pfisterer, Halina Schikner and Maren Suffel for helping me with organizing conferences, lectures and exams during my time at the University of Vienna. Representative for fellow students during my Bachelor and Master studies, I want to thank Katarina and Julia for our common lecture visits, project collaborations and lasting friendship.

Writing of friendship, it is an honor to count on friends like Fabian, Felix, Nicolaus and Raphael who I first met more than 15 years ago. The lively discussions with Fabian about various topics always encouraged me to increase my knowledge. Felix's LAN-parties were always a welcomed distraction from school, studies and work. Nicolaus' more than perfect English challenged me to raise my own English level. Raphael's eagerness to beat me in geography and economics tests sparked my interest into business already in high school. On behalf of all my teachers and lecturers who formed and shaped my interests and skills, I want to thank Prof. Elisabeth Schmidt for being my form teacher for eight years and providing me with relevant mathematical skills and Prof. Katrin Pürer for proliferating my English to a level that allows me to write this dissertation.

Furthermore, I want to greatly thank my dancing school clique Jessica, Lea, Lisa, Simon and Stephanie for all the fun we have had together and my dancing partner Klara for coping with my frequent (academic) travels. Another shout-out goes to my ESN friends, specifically to Claudia, Katharina, Jens, Lisa, Marco, Mario, Patrick, Sandra and Silvia, for challenging and inspiring me. In addition, I want to thank Tobias and Klavdija for the interesting (lunch) discussions. Of course, such a list can only be incomplete. Consequently, I want to apologize in advance for missing anyone out. Nevertheless, I am deeply grateful to all the people who brought and accompanied me to this point.

Another sincere and infinitely large thank you goes to Anna for spell-checking these 200+ pages, for inspiring me in lively discussions to investigate new aspects, for suffering from my vagaries, for coping with me coming home late because I just discovered some new effects and for her endless motivation and happy smiles.

I am deeply grateful to my family for enabling my academic career. I want to thank Amina, Medina, Astrid and Nadir as well as Christina, Michael, Isabella and Leodegar for the pleasant and engaging hours we have spent together. I want to sincerely thank my godmother Ingrid for spellchecking all my theses, for taking me out skiing, for fostering my interests in science from early age on and for an uncountable number of sleepovers. My greatest gratitude goes to my parents Beatrix and Wilfried and my sister Marlene. I want to greatly thank my mother for all the hours she spent stdying with me and checking my homework. This built the basis for my success in school which enabled me to go to university. I am extremely grateful that my father motivated me to start business studies and encouraged me to take up my first academic job as student assistant. Another deep thank you goes to my sister for enduring me in our shared flat despite my unusual living habits. Overall, thank you for all your love, encouraging words and support!

Last, I want to thank Aladdin, Arielle, Baloo, Belle, Merida, Moana, Rapunzel, Pocahontas, Simba, Tarzan and of course Anna, Elsa, Olaf, Christopher and Sven for their musical support and greatly appreciated evening distractions.

Some people are worth melting for!

Table of Contents

1 Introduction

Hic in omnibus fere sermonibus […] ita disputant ut nihil affirmet ipse refellat alios, nihil se scire dicat nisi id ipsum, eoque praestare ceteris, quod illi quae nescient scire se putent, ipse se nihil scire id unum sciat.

Marcus Terentius Varro – Roman Polymath

(Cicero, 45 BC)

The quote above attributes the famous phrase "I know that I do not know"[1] to Socrates. While there exists no record of Socrates ever using this exact phrase, we find several metaphors referring to this phrase in Plato´s (399 BC) *Apology of Socrates* recording Socrates´ trial for corrupting the minds of the Athenian youth. This resulted from the Oracle at Delphi calling Socrates the wisest inhabitant of Athens (Cicero, 45 BC). Socrates himself did not want to believe this and tried to find wiser coevals (Plato, 399 BC). However, although discussing with many distinguished Athenians, he always found ways of disproving their so-called knowledge (Plato, 399 BC). This led Socrates to conclude that the Oracle at Delphi "declared him to be the wisest of all men because all wisdom consists solely in not thinking that you know what you do not know" (Cicero, & Rackham, 1933: 425-427). Whereas this statement introduced fundamental principles for philosophy and science (Vlastos, 1985), it publicly humiliated many notable Athenians, most of them also jury members in the trial. As a consequence, this jury found Socrates guilty and sentenced him to death (Fine, 2008).

Fortunately, the interest in discussing knowledge and philosophizing about the truth of statements and observations did not end with Socrates' death. Yet scientists challenging well-established concepts and perceptions still lived dangerous lives. The reaction of the catholic church to the (re-)discovery of heliocentrism during Renaissance represents a prime example. Scientists promoting Copernicanism (e.g. Johannes Kepler, Galileo Galilei and Giordano Bruno) were banned from publishing their writings and faced prison or even death for breaking this rule (Becker, 2007; Pedersen, 1983). Nevertheless, researchers continued to strive for new findings and new osbervations.

[1] The phrase is often wrongly translated as "I know that I know nothing" but Socrates only claimed that he could disprove specific knowledge but did not say that he "knows nothing" (Fine, 2008).

By doing so, scientists have mainly used three basic forms of inferences to generate new results: Abduction, Deduction and Induction (Jensen, 2008). Abduction occurs when people refer to the nearest and best fitting justification to explain a new observation (Jensen, 2008). As a case in point, ancient Greeks attributed lightnings to Zeus' anger because without knowing of electricity this seemed to be a plausible cause (Dowden, 2006). Deduction refers to the process of applying general theorems to specific cases to receive new implications (Jensen, 2008). For example, if we knew that all cats are cute, we would then be able to deduct that Johnny (my tomcat) is cute. Induction is exactly the opposite of deduction. Hereby, knowledge generation comes from inferring from specific cases to general theorems (Jensen, 2008). To invoke the same example, if we knew that Johnny is cute, we could induct that all cats are cute. Of course, this procedure might lead to erroneous conclusions (Sloman, & Lagnado, 2005) because Johnny´s cuteness might (or actually is very likely to) differ from other cats and thus (at least a few) other cats might not be cute.

1.1 Falsifiability

David Hume (1748) was the first modern scientist to recognize and define this so-called *Problem of Induction*. The question that arose was how often does someone need to observe specific cases so that they can induct that there exists an underlying theory (Hume, 1748). To overcome the *Problem of Induction*, Karl Popper (1935) formalized the concept of *Falsifiability*. Hereby, scientists generate a postulation and then try hard to proof it wrong (Popper, 1935). If they are successful, they reject the original postulation and instead come up with a new one (Popper, 1935). Going back to the example above where we inducted that all cats are cute, we now could test this postulation using the concept of *Falsifiability*. To do so, we would look for as many cats as possible and test whether they are cute. In case we find at least one cat that is not cute, we can reject the postulation that all cats are cute. Instead, we could, in case we find a high share of cute cats, postulate that most cats are cute. This postulation again could then be tested using the concept of *Falsifiability* (although, based on my experience, I personally doubt that there exist more grizzled than cute cats).

1.1.1 An example of falsifiability in today's research

The concept of *Falsifiability* relates back to Socrates because it requires scientists to constantly doubt existing theories and results. This paradigm guides especially the physical sciences (Hendrick, 1990). The 2017 Nobel Prize for the detection of gravitational waves represents a perfect case in point: More than a century ago, Albert Einstein (1916) postulated the existence of gravitational waves based on his theory of relativity. Over the following decades, several mathematicians, physicists and computer scientists proved the algebraic derivation and were also able to observe gravitational waves in simulation studies (Arnowitt et al., 1962; Misner, 1960; Hahn, & Lindquist, 1964; Smarr, 1979). Yet scientists were unable to observe them in nature and therefore (in accordance with the concept of *Falsifiability*) still questioned their existence (Thorne, 2017). To overcome the doubt, Kip Thorne, Ronald Drever and Rainer Wess found the Laser Interferometer Gravitational-Wave Observatory (LIGO) consisting of three installations of L-shaped vacuum tubes with several mirrors worth millions of dollars. Two of them were errected in Hanford, Washington and another one was assembled in Livingston, Louisiana (Cho, 2017). The efforts paid off: LIGO was most likely able to detect gravitational waves stemming from the merger of two black holes on the 14th of September, 2015 (Castelvecchi, & Witze, 2016). Yet there still existed doubts on the validity of this detection as "only the LIGO detectors were observing at the time of GW150914. The Virgo detector was being upgraded, and GEO 600, though not sufficiently sensitive to detect this event, was operating but not in observational mode" (Abott et al., 2016: 3). While the data collected by LIGO accorded with the results expected from previous simulations, various scientists doubted the results (e.g. Hossenfelder, 2017; Creswell et al., 2017; Brooks, 2018). The fog only cleared on the 17th of August, 2017, when LIGO as well as VIRGO (the European version of LIGO located next to Pisa) simultaneously captured gravitational waves from the inspiral of two neutron stars (Abott et al., 2017). This confirmation pathed the way for the LIGO and VIRGO collaborations to receive the 2017 Nobel Prize in Physics (Botner et al., 2017).

Such confirmations do not only constitute a pre-requisite for receiving the Nobel Prize in Physics. In fact, "replication by an independent investigator […] has long been considered essential in the physical and biological sciences, and findings are typically not accepted until they have been duplicated by at least one independent investigator" (Neher, 1967: 261-262). Accordingly, replicability represented the most important aspect according to natural scientists in a survey among professors employed at an undisclosed Big Ten Úniversity (Chase, 1970). However, social scientists ranked replicability only third among ten aspects (Chase, 1970). Unsurprisingly, replication attempts addressing findings from social scientific research articles did not always work out successfully (e.g. Camerer et al., 2016; Duvendack et al., 2015; Federer et al., 2018 Open Science Collaboration, 2015). In the light of these findings, several scholars stated that the social sciences find themselves in a credibility crisis. Section 1.2 covers the credibility crisis in-depth.

1.1.2 Falsifiability and null hypothesis significance testing

The concept of *Falsifiability* builds the basis for null hypothesis significance testing (NHST). Hereby, scholars first generate a null hypothesis. Then they employ statistical tests aiming at rejecting the null hypothesis. If the statistical tests are significant, the scholars can reject the null-hypothesis and instead accept an alternative hypothesis (Gigerenzer, & Murray, 1987). For example, we could postulate the null hypothesis that winter and summer days do not differ in the amount of snowfall. We then observe the snowfall in Kitzbühel for every day in a given year. After the year passed, we run a two-tailed two-sample t-test splitting all recorded days into summer (e.g. after 15^{th} of April and before 15^{th} of October) and winter (e.g. before 15^{th} of April and after 15^{th} of October) days. If the t-test is significant, we can reject the null hypothesis that winter and summer days do not differ in the amount of snowfall in Kitzbühel and accept the alternative hypothesis that they differ.

NHST can not only be applied to natural phenomena like snowfall but is in fact one of the widest spread scholarly approaches in empirical and experimental research in the social sciences (Harlow, 2016). However, many scholars from various disciplines have criticized the common usage of NHST. Cohen (1994) highlighted the problem of false positives in psychological tests assessing schizophrenia. While those tests have an accuracy of more than 95% (implying significance on the usual threshold of $p < 0.05$), about 60% of those identified as schizophrenic are so called-false positives, people not suffering from schizophrenia whose test is positive (Cohen, 1994). This derives from the distribution of schizophrenia in the population. Only about 2% of the world population suffer from this syndrome (Cohen, 1994). Consequently, the share of wrongly identified cases among the whole population is $98\% * 5\% = 4.9\%$ while the share of correctly identified cases among the whole population is $2\% * 95\% = 1.9\%$ (Cohen, 1994).

Cumming (2008) also critized NHST. He ran simulations using means and standard deviations of children's verbal abilities in two school districts as well as the number of students per district as base data. After 25 runs, the simulations returned p-values for two-tailed two-sample t-tests ranging from $p < 0.001$ to $p = 0.76$ (Cumming, 2008). Based on this, Amrhein et al. (2017) argued that p-values are hardly replicable. In fact, the heavy usage of p-values represents a core source of the credibility crises in the social sciences (Amrhein et al., 2017). The following section elaborates further on this crisis.

1.2 Credibility Crisis in the Social Sciences

The events of fall and winter 2011-12 had a long-lasting effect on social scientific research: Shortly before that period, Stapel and Lindenberg (2011) published a *Science* article (presumably) showing that trashy environments induce discriminative and stereotyping traits. According to the paper, Stapel conducted the corresponding experiment at the Utrecht train station at a row of six chairs with an Afro-Dutch confederate sitting on the first chair. Stapel and Lindenberg (2011) reported that if the platform was trashy (as a result of a strike of the cleaning personnel), the Caucasian participants (random people waiting for their train) took on average a seat further away from the Afro-Dutch confederate than if the platform was clean (two weeks later after the strike had ended). While this finding was surprising yet intuitive, the problem with this study was, as it turned out several months later, that there existed no combinations of six chairs at the platforms of the Utrecht train station (Zwart, 2017). In September 2011, PhD students of Diederik Stapel contacted multiple department heads at the University of Tilburg alleygating him of scientific fraud as he only supplied them the overall data but was not willing to supply them with the original interview sheets and questionnaires for their projects (Bhattacharjee, 2013). In fact, Diederik Stapel "took over" the experiment conduction in nearly all his research projects from 2004 onwards and never supplied his colleagues or subordinates with the original data (Bhattacharjee, 2013). The PhD students' allegation led to a commission investigating all of Stapel's research. Overall, they found evidence for data fabrication in 55 papers and 10 dissertations (Enserink, 2012). Currently, Diederik Stapel has 58 retractions ranking him 5th in the overall and 1st in the social science ranking of *Retraction Watch* (Degen, 2015).

Only months after the first public allegations against Diederik Stapel and far before the publication of the final report of his fraudulent actions, another scandal hit the social science. In early 2012, a group of about 20 business scholars including editors from *Research Policy* and *Organization Science* started to investigate the works of Ulrich Lichtenthaler, back then professor at the WHU Mannheim, because it seemed that he had submitted several papers containing similar research design, analyses, results and implications to multiple journals (Storbeck, 2012). In addition the self-plagiarism allegations, the group also charged him with statistical irregularities. As a case in point, Lichtenthaler and Ernst (2009) reported significant coefficients despite their standard deviations exceeding their values. As for now, 16 out of 43 publications from Ulrich Lichtenthaler were retracted (Hermanns, 2014). In addition, the WHU Mannheim invoked his habilitation (West, 2013) and he was forced to quit his job in 2014 (Retraction Watch, 2014).

While these two infamous cases raised serious concerns about ethical research conduct and the trustworthiness of social scientific results, they only represent the tip of the iceberg (Karabag and Berggren, 2012). In fact, Hopp and Hoover (2019) showed that more than a third out of 196 editors at management journals have dealt at least once with submissions containing fabricated or falsified data and nearly every second of them have dealt at least once with submissions including “changed or omitted data points” (Hopp, & Hoover, 2019: 1554). Unfortunately, data malpractices do not constitute the only form of misconduct. In fact, research design flaws (e.g. unreported research design changes to increase funding and publication chances or purposely not reporting technical details of the statistical analyses), plagiarism and authorship disputes occur even more often than data fabrication or falsification (Martinson et al., 2005)

As those instances undermined the trust in scientific results and researchers, it is not surprising that many scholars from various backgrounds stated that the social sciences face a credibility crisis (e.g. ; Bergh et al., 2017; Byington, & Felps, 2017; Earp, & Trafimow, 2015; Firth et al., 2014; Gall et al., 2017). As a case in point, Osherovich (2011) highlighted that even venture capitalists do not trust published scientific results when making their investment decisions. Consequently, the credibility crisis represents a serious problem in a world filled with polarizing terms like *fake news* and *alternative facts* because it undermines public trust in scientists and their results (Hendriks et al., 2016). To overcome this problem, we need to identify the most serious and most prevalent forms of academic malpractices. Subsequently, we must find and implement adequate solutions to raise scholars’ credibility as well as the credibility of the research process (Ioannidis, 2012).

This thesis contributes to overcoming the credibility crisis in the social sciences by analyzing the existing literature on ethical and responsible behavior focusing specifically on academic (mal)practices in the social sciences. The following sections highlight which areas have already been thoroughly covered and where more attention and action is needed. The four essays following the introduction tackle the latter topics. They analyze important but so far only rudimentarily covered forms of (un-)ethical scholarly behaviors and point out solutions to overcome them. Hereby, we generate important insights on why and how academic misconduct occurs and highlight stakeholder action plans that universities, journal editors and publishers could follow to overcome the credibility crisis in the social sciences.

1.2.1 Scholars' credibility

To increase public trust in scientific expertise, we first need to clarify who actually those experts are. Usually, if we look for scholarly knowledge we go to databases and search engines like Google Scholar, SCOPUS or the Web of Science (Jacso, 2005). We then take the most appropriate, most cited and/or newest (peer-reviewed) articles and cite them by writing phrases like "Hopp and Pruschak (2018) showed…" or "The conclusions from Beck et al. (2020) allow us…" (Golden-Biddle et al., 2006). In other words, we specifically name the experts in the citations. However, this ideal scenario in which the listed authors are also the experts for that topic does not always apply. As a case in point, Martinson et al. (2005) highlighted that authorship issues occur about ten times more often than data fabrication or data falsification. Consequently, we might cite sometimes authors who contributed very little to a research project and thus might not be experts for the topic in question (Kumar, 2018). This issue emerged with the stark increase in multi-authored papers (Fleischman, & Schuele, 2009): Whereas from all papers published in the American Economic Review in 1950 only 8 percent included more than one author, 80 percent of all papers published in the American Economic Review in 2010 included more than one author (Hamermesh, 2015; Hudson, 1996). Having multiple authors working on the same research paper allows scientists to divide the workload. Hence, Adam Smith's (1776) concept of the division of labor also works in academia: Dividing research projects into individual tasks allows specialization. Scientists with talents in collecting and/or analyzing data can focus on these affairs while skilled writers might engage more in drafting the actual paper (Leahey, & Reikowsky, 2008). Not surprisingly, the increase in co-authored papers also came along with a boost in academic productivity resulting in more and more publications being published in less and less time (Engels et al., 2012; Lee, & Bozeman, 2005). However, we can no longer assume that all authors of a research article are experts in all its aspects due to their different specializations (Leahey, & Reikowsky, 2008).

To identify whose expertise we can rely on, journals like *Nature*, *Science*, *PLoS ONE* and *PNAS* introduced so-called contribution disclosures or authorship statements (Sauermann, & Haeussler, 2017). Hereby, each author (and also important contributors) have to describe the tasks they executed for the research project in question (Hwang et al., 2003). Whereas such authorship statements represent the norm in life science journals and have become more frequent in the natural sciences (Sauermann, & Haeusler, 2017), we did not find a single example of a social science journal requiring this kind of statements. In fact, authorship assignments overall have been discussed and investigated to a much smaller degree in the social sciences compared to life and natural sciences (Marusic et al., 2011). Notable exceptions are Hamilton et a. (1997), Manton and English (2006) and Manton and English (2008). These three studies address cases of so-called honorary authorship.[2] Especially Manton and English (2008) highlighted the need for thorough investigations of authorship practices in the social sciences. They showed that nearly 10% of their respondents selected from business faculty members in the U.S. gave author credits to a person that did not contribute anything to the research paper. More than 35% of their respondents had assigned authorship for "very little work" at least once (Manton, & English, 2008). To understand the underlying forces of such authorship malpractices and to establish guidelines to avoid them, Essay 1 addresses honorary authorship as well as its infamous sibling ghost authorship.[3] The essay investigates motivational and circumstantial factors that foster authorship malpractices and provides hands-on solutions like authorship contribution forms and whistleblowing platforms.

Why is authorship such a contested topic? The primary answer to this question is that publication and citation credits represent the "currency" in the academic job market (Figa-Talamanca, 2007). The more publications scholars have in top-tier journals and the more often those are cited, the higher are their chances of successfully mastering search and tenure procedures (Park, & Gordon, 1996). Yet as already discussed we must be very careful with giving all credits to all authors. Indeed, the ongoing trend towards more and more multi-authored research papers has required new measurement techniques for assessing researchers' productivity (Carpenter, Cone, & Sarli, 2014). To cope with the changing sizes of author teams, Cole and Cole (1973: 33) introduced a technique called "straight counts" that only awards publication and citation credits for the first author with all other authors not receiving those benefits (Cole,

[2] Honorary authorship emerges if an author receives authorship despite not contributing substantially to a research paper (da Silva, & Dobranszki, 2016).

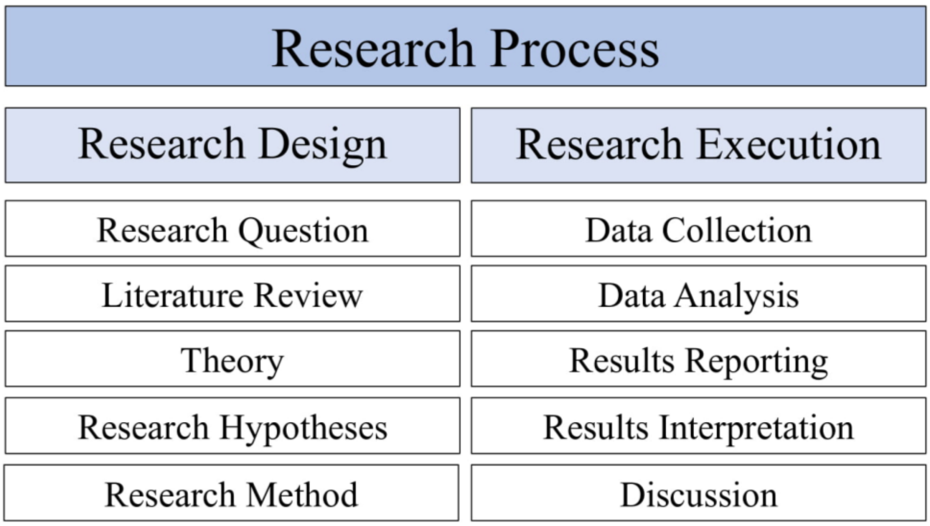
[3] Ghost authors are researchers who contribute substantially to a research article but do not receive authorship (da Silva, & Dobranszki, 2016).

& Cole, 1973). Similar to this approach, Shaw (1967) and Stallings and Singhai (1970) introduced techniques that give the first author a certain ratio of the counts and all other authors together the remaining ratio of the counts (e.g. for a paper with three authors the first author would receive half of the counts and the other two authors would receive a quarter each). Yet these approaches relied on estimates as without contribution disclosures one cannot know the involvement of each author in the research process (Lindsey, 1980). In addition, these techniques did not seem reasonable for fields like Accounting or Economics in which authors are not ranked by contribution but alphabetically (Chan et al., 2009; Engers et al., 1999). To better assess researchers' impact and productivity, various editorials and articles have followed up on the discussions of author order and how to distribute publication and citation counts among co-authors (e.g. Balkin et al., 2018; Endersby, 1996; Floyd et al. 1994; Macfarlane, 2017; Sauermann, & Haeussler, 2017). Nevertheless, existing research has only rudimentarily covered the basic underlying question of who and why actually comes together to form those author teams so far. Henriksen (2016), Lariviere et al. (2006) and Ossenblok et al. (2014) represent notable exceptions. Two of those studies investigated and compared the average number of authors across research fields (Henriksen, 2016; Ossenblock et al., 2014) while the third looked at language and regional differences (Lariviere et al., 2006). However, all three articles lacked theoretical frameworks explaining their effects and none of them included multivariate analyses. Essay 2 extends their findings and interpretations. Hereby, we apply the well-known economic concept of transaction costs towards author collaborations and test this application empirically addressing research field, regional, working experience and job position effects in a single model. The essay offers deep insights into authorship practices in the social sciences and shows that search and tenure commissions need to overcome solely authorship-based decision criteria like publication and citation credits and focus more on the skill sets of the candidates.

1.2.2 Research process credibility

The inevitable first step in overcoming the credibility crisis is increasing public trust in scientific experts. Nevertheless, ensuring the credibility of the research process itself is the ultimate goal in order to restore general trust in the social sciences (Byington, & Felps, 2017). Yet before digging into the integrity of research processes, we need to clarify what type of research processes we refer to. Theory and literature-guided confirmatory research represents the primary scientific process in management (Tukey, 1980), economics (Hakim, 1987) and psychology scholarship (Hershey et al., 2006). Figure 1.1 exemplarily highlights the archetypical experimental and empirical research process in the social sciences.[4] In line with Bhattacherjee (2012), Saunders et al. (2016) and Tharenau et al. (2012), the research process consists of two phases: The research design phase and the research execution phase. During the research design phase, scholars first generate a research question and conduct a literature review to assess the corresponding theory. Based on this, they apply deduction (Chapter 1) to generate research hypotheses that postulate specific effects and choose an appropriate research method. After completion of the research design phase, the research execution phase starts. Hereby, scientists collect data and analyze them with the aim of finding support for the previously established research hypotheses. They report the results and interpret them in line with the underlying theory and existing literature in the discussion (Bhattacherjee, 2012; Saunders et al., 2016; Tharenau et al., 2012). In the following, this thesis addresses the most prevalent forms of (un-)ethical behavior and academic (mal)practices encountered in both phases.

Figure 1.1: Archetypical experimental and empirical research process in the social sciences



Note: Own figure based on Bhattacherjee (2012), Saunders et al. (2016) and Tharenau et al. (2012)

[4] The archetypical research process varies from research approach to research approach. On a general note, the social sciences distinguish between theoretical/literary-based research and experimental/empirical data-based research (Lancaster, 2005). The research process in Figure 1.1 mainly applies to experimental and empirical data-based research because theoretical- and literary-based research do usually not include data collection and data analysis (Lancaster, 2005).

1.2.1.1 Academic malpractices in research design

HARKing (Hypothesizing After the Results are Known) represents the most common form of unethical research design changes in the social sciences (Hollenbeck, & Wright, 2017). According to the archetypical research process discussed above, scholars should generate research hypotheses from the existing theory before they conduct the data collection. Only afterwards, they should test the hypotheses using the gathered data. However, researchers face incentives to present as many significant statistical tests as possible because of the *publication bias.* The *publication bias* constitutes the scientific term for the fact that over the last 70 years, economic, management and psychology journals have shown a growing tendency towards preferentially publishing significant results (Fanelli, 2012; Harrison et al., 2017; Rosenthal, 1979; Sterling, 1959). Consequently, finding support for more hypotheses increases scholars' publication chances (Fanelli, 2012; Harrison et al., 2017; Rosenthal, 1979; Sterling, 1959). Yet even the best a priori defined research hypotheses sometimes do not transfer into significant statistical results due to various reasons like noise in the data or inappropriate research designs (Kerr, 1998). To overcome the "danger" of receiving insignificant results, some scholars reversed the research process by first collecting as well as analyzing the data and only afterwards formulating ex post hypotheses that suit the results, thus engaging in HARKing (Leung, 2011). This way they could ensure that their papers contained many significant results, which in turn increased their publication chances (Lipton, 2005). Unsurprisingly, HARKing is very common in the social sciences with 27% of surveyed psychologists admitting that they have employed this technique in their past research (John et al., 2012). However, HARKing is highly unethical because it suggests to editors, reviewers and readers that the scholars conducted a rigor theory and literature guided confirmatory research process while in fact they employed the theory that best matched the results (Bosco et al., 2016). This is problematic as we lose the valuable insights theory deduction might provide. Consequently, we rely solely on induction for both, identifying the underlying theory as well as the new implications. It is therefore no wonder that a substantial number of research articles, commentaries and editorials in many top-tier social scientific journals have already raised, addressed and investigated HARKing as well as its consequences and have proposed solutions to this problem (e.g. Bekkers, 2012; Cox et al., 2018; Mazzola, & Deuling, 2013; Murphy and Aguinis, 2019; Rubin, 2019; Vancouver, 2018). The preferred solution seems to be study preregistration (Burlig, 2018; Gonzales, & Cunningham, 2015; Yamada, 2018). Hereby, scholars either register their study at an online registry like *AsPredicted* (2020)*, AEA Registry* (2020) or *Open Science Framework Preregistration* (2020) or only submit a registered report outlining in detail the whole research project but not including results of the data collection and analysis to journals (e.g. *Nature Human Behaviour,* 2017; *The Leadership Quarterly,* 2020; *Journal of Development Economics,* 2020).

Plagiarism constitutes another academic malpractice that can occur in the research design phase. In fact, "classic" plagiarism, copy-pasting from others, is the best known form of scientific misconduct as nearly every undergraduate lecture on scientific paper writing includes warnings that students must quote anything that they copy-paste to avoid plagiarism (Jackson, 2006). However, while "classic" plagiarism appears from time to time in students' seminar and term papers (Selwyn, 2008), it occurs only very rarely in social science journals (Sun, 2013). This derives from the fact that most of the journals already have employed plagiarism detection software (Luparenko, 2014). Yet two other forms of plagiarism exist more often: Copying and rephrasing ideas and results from others without referencing to them as well as the act of self-plagiarism (Lehman, & Ramanujam, 2009). Honig and Bedi (2012) estimated that more than 13% out of 279 papers presented at the 2009 Academy of Management copied more than 5% of their content from other papers without referencing them. This corresponds to Hopp's and Hoover's (2017) result that nearly 60% of 165 surveyed management journal editors had to deal with at least one case of plagiarism per year. Whereas copying from others without properly acknowledging them through citations or quotations constitutes academic misconduct for sure, there exists an ongoing debate on whether self-plagiarism, copying from own work, also counts as academic misconduct (Bretag, & Mahmud, 2009). For example, Thurman et al. (2016) argued that using similar backgrounds or methods sections to those already employed in published articles could stem from the new paper using the same dataset or the same experimental or empirical environment. In addition, Vermuelen (2012) pointed out that scholars developing widely recognized concepts and theories might need to publish them in multiple journals to spread them across discipline boundaries. Last, Callahan (2018) highlighted that self-citations instead of self-plagiarism might torpedo double-blind peer review processes as this could unveil easily the identity of the original authors.

Nevertheless, "there is no doubt that duplicate publication of original data and/or failing to reference previous work constitute research misconduct" (Chrousous et al, 2012: 231). As a case in point, Martin (2013) discussed the case of an article submitted to *Research Policy* that had more than 1700 words in common with an already published article from the same authors. In addition, neither the theoretical framework nor the implications of the newly submitted paper differed from the already published paper. The editors therefore decided to reject this paper (Martin, 2013). Horbach and Halffman (2019) investigated self-plagiarism among Dutch scientists. They showed that about 13% of the investigated economics articles and nearly 5% of the investigated psychology articles overlapped contextually at least 10% with papers published previously by at least one of the co-authors.

The fact that nearly every type of publication can contain some sort of plagiarism explains why countless research articles and editorials have already dealt with it (e.g. Arce et al., 2008; Enders, & Hoover, 2006; Pupovac et al., 2008; Shahabuddin, 2009; Stitzel et al., 2018; Weber-Wulff, 2014). Fortunately, there exist already solutions to this problem: plagiarism detection softwares like *UniCheck* (2020) or *TurnItIn* (2020). In fact, many top-tier social science journals like the *American Economic Review* (2020), *The American Sociologist* (2020) and the *Academy of Management Journal* (2020) employ plagiarism detection software to check all submissions. The enhancing developments of artificial intelligence in those software applications already allows checking not only textual similarities but also content similarities and even assessing similarities in different languages (Ali et al., 2011; Luparenko, 2014; Lykkesfeldt, 2016).

1.2.1.2 Plagiarism and motivation for this thesis

Plagiarism played a vital role in the topic search for this PhD thesis. Prof. Dr. Christian Hopp lectured the seminar *The Economy of Sciences* in the winter semester 2017/18. I had the honor to support him. Scholarly careers, university incentives and academic misconduct represented the major topics of this seminar. Prof. Dr. Hopp covered these aspects thoroughly in the introduction session and the students should dig deeper into these topics for writing their seminar papers. The students had to submit their papers until the end of the semester and, as we often observe, some of them deregistered from the course before submitting a paper. Therefore, we received only a handful of seminar papers. When I read those papers, one of them addressing the scientific publication process stood out in the amount of used literature and the quality of the inputs. However, this paper used inconsistent reference styles and did not include all references mentioned in the text also in the reference section. I forwarded my remarks to Prof. Dr. Hopp. Keeping my comments in mind, he spotted several textual similarities between the seminar paper and Hirschauer (2004), a paper that was part of the seminar's mandatory literature list. As we possessed reason for suspecting plagiarism, I went through the seminar paper again and checked every paragraph. The result was shocking: Most paragraphs in the seminar paper originally stemmed from published articles and reviews written in English. The student had translated these paragraphs one by one into German without citing or any other sort of referencing. We consequently reported the plagiarism case to the competent authority of the RWTH Aachen University. Considering that the seminar topic was closely related to academic misconduct, this incident was obviously very ironic.

This plagiarism case influenced my topic selection heavily. If even students who were explicitly taught about cases of and dangers arising from academic misconduct behave unethically in the very same seminar, why should not scientists, who face various incentives and pressures "to publish or perish" (Teute, 2001: 102), use similar shortcuts to increase their productivity and reputation (Ioannidis, 2012). As a case in point, Stürmer et al. (2017) conducted a survey among junior faculty psychologists. They showed that more than 80% of their respondents believed that competition for publication spots, tenure positions and funding are causes of academic malpractices. Moreover, even scientists focusing on unveiling and overcoming academic misconduct have been convicted of taking such shortcuts themselves in recent years:

On the one hand, there exists the case of John P. A. Ioannidis. He is one of the most influential scholars in meta-research (the academic field of researching science itself) and one of the strongest critics of the publish or perish incentives. Ioannidis (co-)authored two studies assessing the mortality of COVID-19 (Dirnagl, 2020). Ioannidis (2020) included a meta-analysis of studies investigating the prevalence of antibodies. His goal was to assess the infection fatality rate (IFR) (Ioannidis, 2020). Ioannidis (2020) concludes that the fertility rate of COVID-19 is low compared to other infectious diseases. The reviewers did not agree with this assessment as they assessed this study as showing data that "are useful and add to the emerging picture on IFR, however substantial conclusions cannot be drawn" (Hallett, 2020). The second study assessed the prevalence of antibodies among residents of Santa Clara County (Bendavid et al., 2020) and concluded originally that based on the widespread existence of antibodies, the IFR of COVID-19 would be 0.17%. Furthermore, the study stated that "[t]he authors have declared no competing interest" (Bendavid et al., 2020). However, several experts criticized the study shortly after its pre-print release for using non-licensed antibody tests and several other methodological issues (Ting, 2020). Only few weeks later, a whistleblower reported that the study was financed by David Neeleman, the founder of JetBlue, questioning the independence of the results (Offord, 2020). According to Ting (2020) the authors revised the study highlighting an IFR of 0.33%, a substantially higher number. Yet the article available at medrvix was still the unrevised version in September 2020 (Bendavid et al., 2020). Morover, the professional websites of the first author and the last author did not include the paper as working paper (Bendavid, 2020; Ioannidis, 2020).

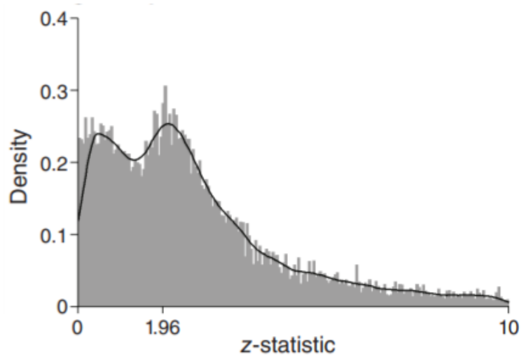
On the other hand, addressing specifically the social sciences, there exists the case of Schimke and Ambrose (2011). This editorial in *Management and Organization Review* outlined the major challenges arising from academic misconduct and gave suggestions for authors, reviewers and journal editors on how to avoid and overcome them. Yet Schimke (2009) published earlier an editorial in the *Academy of Management Review* addressing (un)ethical behavior among members of the *Academy of Management.* Schimke (2009: 586) started with "Our profession has no formal audit function. More precisely, our research and publishing activities are not monitored by a formal audit process." Schimke and Ambrose (2011: 398) wrote "Our profession has no formal, regular, auditing process. In particular, our research and publishing activities are not monitored by a formal audit process." While those two excerpts are nearly identical, Schimke and Ambrose (2011) did not cite Schimke (2009). In fact, Schimke and Ambrose (2011) did not even include Schimke (2009) in their reference list. In November 2013, a reader reached out to the editors of *Management and Organization Review* stating that students in his seminar on ethical publishing found several similarities between the two articles. Subsequently, the editors reached out to the authors and agreed on retracting the editorial (Tsui et al., 2014). Understandably, Schimke felt deeply humiliated but acknowledged the irony of retracting a paper aiming at educating scholars on ethical behavior (Tsui et al., 2014).

These two ironic incidents further highlighted the issues arising from academic misconduct. If even those who research and publish on topics related to the integrity of science use questionable research practices, how scientists will act who care much less about ethics in research?

1.2.1.3 Academic malpractices during research execution

The cases of Diederik Stapel and Ulrich Lichtenthaler highlighted prime examples of academic misconduct occurring during research execution. On the one hand, Stapel's faking of experimental data showed what can go wrong in the data collection. On the other hand, Lichtenthaler's stars indicating significance for non-significant coefficients questioned the integrity of data analysis and results reporting. However, the red line is not clearly defined for many other instances of questionable research practices. Simmons et al. (2011: 1359) argued that this especially applies to many data related subjects like "Should more data be collected? Should some observations be excluded? Which conditions should be combined and which ones compared? Which control variables should be considered? Should specific measures be combined or transformed or both?" Scholars usually cannot clarify all of these decisions in advance because some of them are very case specific (e.g. the response rate, the number and strength of outliers, the validity scores of instruments) (Simmons et al., 2011). Consequently, scholars can use the leeway of these decisions to shape the results in a preferential way by reporting only those specifications that lead to significant results (Banks et al., 2016). As a case in point, Brodeur et al. (2016) discovered the so-called *inflation bias*. Based on the already discussed *publication bias*, Brodeur et al. (2016) suspected that smaller p-values increase publication chances. The authors tested this assumption by analyzing test statistics from 641 articles published in three top-tier economic journals. Their results showed that the actual published p-values did not correspond to the postulated constant increase in publication chances. Instead, they observed two humps with few articles containing p-values around 12% and a high number of articles containing p-values that are just below 5% (Figure 1.2). Brodeur et al. (2016) argued that the humps point out that economists played around with model specifications until they reached significant results.[5] In other words, if economists' initial results indicate p-values just slightly below the 5% level, some refrain from running the models with other specifications (e.g. different treatments of outliers, N/A answers, …). In turn, if economists' initial p-values lie just above the 5% level, some run the models again with different specifications until they reach "publishable" p-values below 5% (Brodeur et al., 2016).

Figure 1.2: Distribution of z-values among economic papers



Note: Figure from Brodeur et al. (2016: 9)

---

[5] The 5%-threshold, the most common significance threshold in the social sciences, goes back to Fisher (1922).

In the worst case, different model specifications can result in two papers addressing the same phenomenon and using the same dataset presenting very different findings and implications (Simmons et al., 2011). Silberzahn and Uhlmann (2013) and Silberzahn et al. (2014) represent two examples of such contradicting publications. Silberzahn, & Uhlmann (2013) showed by using hierarchical linear models that Germans with noble-sounding surnames like “Kaiser” worked more often as executive board members. Renowned statistician Prof. Uri Simonsohn read this journal article and was very suspicious of the finding. He asked the authors to send him the original data and conducted further investigations. He concluded that the highlighted effect resulted from different name frequencies and teamed up with the original authors to employ name matching to derive the actual causal effects (Silberzahn et al., 2014). Silberzahn et al. (2014) reported that the effect estimated in the HLM regressions disappeared in the matching models.

Unfortunately, self-corrections as the one in Silberzahn et al. (2014) occur very rarely in the social sciences (Rohrer et al., 2020). This derives from the fact that retractions and similar incidents reduce scientists’ future citation rates and hence harm their career prospects (Azoulay et al., 2017). As researchers are unlikely to retest or reassess their own previous findings (Rohrer et al., 2020), many journals have opened calls for replication studies to incentivize scholars to validate their own work and the work of others (e.g. Brüggen, 2020; Burman et al., 2010; Clapp-Smith et al., 2018; Cousineau, 2014; Experimental Psychology, 2020; Simons et al., 2014). Essay 3 is a response to the call of Clapp-Smith et al. (2018) for replicating leadership studies conveying important implications. The paper provides new insights on the effects of high school leadership and contains a hands-on step-by-step guide on how to conduct a thorough and robust replication of quantitative empirical articles. To conduct the replication, we were able to use the same dataset as the original study because fortunately the American Institutes of Research still curates the dataset.

However, having access to a study’s original data is often the exception than the norm in the social sciences: Wicherts et al. (2006) asked 141 corresponding authors of papers published in top psychology journals for their data but only received datasets for 38 articles. Krawczyk and Reuben (2012) contacted 200 economists requesting datasets belonging to articles stating that the authors would make the data available upon request. In the end, they only received datasets from 88 authors. Dr. Jermain Kaminski analyzed the 200 most cited (according to Google Scholar) open innovation papers in 2018 searching for their data. None of those papers made the data available via the publication source. To understand the underlying reasons of these non-disclosures, Essay 4 analyzes data sharing among innovation researchers and its determinants. We specifically addressed innovation management researchers as they regularly highlight the benefits of opening science and innovation processes (Bogers et al., 2017; Randhawa, Wilden and Hohberger, 2016; von Hippel, 2017; West et al., 2014) and should therefore be prone to open science practices like data sharing (Beck et al., 2020).

1.2.3 Overview of academic practices and the credibility crisis

Figure 1.3 summarizes section 1.2. It distinguishes the credibility crisis into two aspects: scholars' credibility and research process credibility. Moreover, we divide the research process into research design and research execution phase. Figure 1.3 then lists the above discussed most prevalent academic (mal)practices in red and highlights possible solutions for overcoming them in green. The figure also includes relevant literature references for the academic (mal)practices. Last, Figure 1.3 highlights the literaric context each essay of this thesis is embedded in. Hereby, Figure 1.3 shows how all four essays contribute to overcoming the credibility crisis in the social sciences by providing guidance for ethical and responsible behavior in applied empirical research. The following chapter outlines the exact contributions of each essay in detail.

Figure 1.3: Academic practices and the credibility crisis in the social sciences

Note: Text in red represents problematic academic practices. Text in green highlights solutions to ensure and strengthen ethical and responsible behavior in applied empirical research.

1.3 Summary of the essays

All four essays address forms of academic (mal)practices and offer various solutions. The sequence of the essays accords to their first mentioning in this thesis. Consequently, Chapter 2 contains Essay 1 addressing ghost and honorary authorship in the social sciences. Essay 2 in Chapter 3 shows how transaction costs influence the size of author teams. Chapter 4 with Essay 3 represents a step-by-step guide and hands-on example of replication studies. Last, Essay 4 in Chapter 5 highlights proliferating and limiting factors for data sharing practices among innovation scholars.

1.3.1 Chapter 2

Essay 1 in Chapter 2 begins with showing that nowadays more research articles than ever possess multiple authors, enhancing better quality and higher productivity. However, the rise of multi-author papers also pathed the way for the emergence of authorship malpractices. Because incorrect authorship assignments alter academic citation and publication counts, they constitute a serious threat to the integrity of science. Existing research from the life sciences indicates that ghost and honorary authorship are the most frequent authorship malpractices. As a response, many life science journals have started to require contribution disclosures upon paper submission. However, such a practice does not exist in the social sciences. In fact, they even lack an assessment of whether they also face ghost and honorary authorship. Based on a large scaled survey, the study shows that both, ghost and honorary authorship occur in the social sciences with honorary authorship spreading being substantially more common than ghost authorship. Yet while more than every third paper contains wrongly assigned authorship credits, most of the participating social scientists assign authorship correctly in three hypothetical scenarios. This difference between the correct hypothetical assignments and the incorrect actual assignments indicates that underlying forces like publication incentives and hierarchical pressure induces scholars to assign author credits improperly despite knowing better. Based on these findings we call for journals to implement contribution disclosures, for research institutions and publishers to implement authorship-whistleblowing platforms and for employers to move beyond authorship-based citation and publication rankings in hiring and tenure processes.

1.3.2 Chapter 3

Essay 2 in Chapter 3 also starts with discussing the unprecedented rise in the size of author teams across all academic disciplines. Yet the speed of increase in co-authors varies considerably across research fields, geographical regions, job positions and experience even within the social sciences. Unfortunately, existing studies failed to provide conceptional frameworks theoretically explaining the differences. Moreover, they focused solely on one aspect each. Essay 2 overcomes this issue by drawing from transaction-cost theory. We generate a solid conceptional model explaining the factors influencing team authorship. We test the model in a multivariate analysis employing data from a large-scaled worldwide survey. Our results show that psychologists as well as information technologists and operations researchers work on average in larger author teams while sociologists and political scientists work in smaller author teams. In addition, we find that Eastern European scholars work in smaller author teams and that postdocs tend to have more single-authored publications. Based on our results we call upon those in charge of search and tenure procedures to keep the applicants' different research fields and geographical backgrounds in mind and to move beyond merely counting publications and citations. Furthermore, we highlight the importance of academic conferences as they allow scholars to establish fruitful networks for future collaborations.

1.3.2 Chapter 4

Essay 3 in Chapter 4 replicates and extends Kuhn and Weinberger's (2005) "Leadership Skills and Wages". The original article found that those white males who were club presidents and team captains in high school earned significantly more eleven years later. As the empirical relationship between leadership positions and subsequent earnings includes those characteristics that predate high school and those that are developed because of leadership activity participation in high school, the original study cannot differentiate between leadership skills developed earlier and those developed in high school. We employ propensity score matching on leadership exposure in high school to control for potential endogenous observable selection and provide estimates from instrumental variable regressions to assess the robustness of the original effects to other omitted causes. To investigate the generalizability of the original findings, we also extend the sample by including females and non-white males. Last, we investigate how an extension of the initial (eleven years) time horizon to almost 50 years affects the coefficient estimates. We can corroborate the original effect that those who occupied leadership positions as captains and presidents earn more eleven years after high school and report higher income 50 years after high school. We fail, however, to find effects for those who occupied only a role as captain or president solely. Moreover, the findings do not generalize to the samples of females and non-white males. Our findings provide important insights into later-life benefits of early leadership exposure and have implications for those designing leadership training programs and those taking on (or refraining from) leadership positions in early life.

1.3.2 Chapter 5

Essay 4 in Chapter 5 bases on the fact that many researchers, practitioners, and policy makers attest to the benefits of open data. Yet many scholars, even those in the area of innovation management and open innovation research, still prefer not to disclose their empirical data. Drawing from the resource-based view we generate potential factors that motivate or hinder data sharing. We test these stipulations using a survey among management researchers. The results show that less than a third of those researchers have made at least one of their datasets publicly available. While respondents are aware of the communal benefits of revealing research data to other scientists, their perceived costs and risk of open data are consequential in inhibiting data sharing. Despite refraining from sharing their own data, paradoxically most respondents would like to see journal policies that foster data sharing. Our results therefore call upon editors to introduce open data policies for their journals and highlight an overdue need to reshape the academic institutions and incentive systems so that not only article publications are rewarded, but also data publications.

1.4 Publication history

The four essays in this thesis are theory driven, empirical quantitative research articles. The doctoral candidate wrote large parts of the essays and substantially contributed to all aspects of the relevant research projects ranging from defining theoretical frameworks to statistical data analysis. We submitted all articles to conferences and/or journals and the articles were included in conference proceedings and/or sustained at least the first submission round. At the time of submission of this doctoral thesis, all essays were either under review or accepted for publication. The essays represent the most up-to-date versions of the research articles at the thesis submission date.

1.4.1 Essay 1 – "Ghost and honorary authorship in the social sciences"

Keywords
Ethics in science; Meta-research; Scientific authorship; Ghost authorship; Honorary authorship; Academic incentives

Under review at *PLoS ONE*:

- Pruschak, G., & Hopp, C. 2020. And the credit goes to … - Ghost and honorary authorship among social scientists. Under review at *PLoS ONE.*

Presentations:

- American Accounting Association Annual Meeting 2018. Washington DC, USA. August 7 2018.
- Academy of Management 2019. Chicago, IL, USA. August 13 2019.
- Jahrestagung des Bildungsökonomischen Ausschusses 2020. Vienna, AUSTRIA. February 28 2020.

1.4.2 Essay 2 – "Team authorship in the social sciences"

Keywords

Science Sociology; Meta-research; Transaction costs; Scientific authorship; Academic incentives; Research collaborations

Under review at *Research Policy*:

- Pruschak, G. 2020. Transaction costs and the size of author teams in the social sciences. Under review at *Research Policy*.

Presentations:

- Academy of Management 2020. Vancouver, BC, CANADA. August 10 2020.

1.4.3 Essay 3 – "Replicating the effect of high school leadership on later life earnings"

Keywords

Leadership; Earnings; Replication; Causality; Endogeneity

Published in *The Leadership Quarterly*:

- Hopp, C., & Pruschak, G. 2020. Is there such a thing as leadership skill? – A replication and extension of the relationship between high school leadership positions and later-life earnings. *The Leadership Quarterly* Forthcoming.

1.4.4 Essay 4 – "Open data practices in innovation management research"

Keywords

Open data; Research data; Resource-based view; Innovation; Replication

Under review at *Industry & Innovation*:

- Barczak, G., Hopp, C., Kaminski, J., Piller, F., & Pruschak, G. 2020. How Open is Innovation Research? – An Empirical Analysis of Data Sharing among Innovation Scholars. Under review at *Industry & Innovation*.

Presentations:

- 16th Annual Open and User Innovation Conference 2018. New York City, NY, USA. August 7 2018.
- 1st Open Innovation in Science Research Workshop 2019. Vienna, AUSTRIA. May 3 2019.

2 Essay 1 – Ghost and honorary authorship in the social sciences

2.1 Introduction

"Publish or Perish" epitomizes the academic reward system across scientific disciplines nowadays better than ever before (Kendall, & Campanario, 2016; McGrail et al., 2006; Miller et al., 2011; Rawat, & Meena, 2014; Sestak et al., 2018).[6] Advances in communication technology as well as the continuing growth of English as the dominant scientific language have constantly increased the competition for publication spots in top tier journals (Di Bitetti, & Ferreras, 2017; Tardy, 2004; van Raan, 2001). To withstand the pressure researchers have increasingly teamed up with colleagues to increase their productivity (Glaenzel, & Schubert, 2001; Lee, & Bozeman, 2015). This has led to a strong increase in the share of published papers written by more than one author in almost every discipline (Hsiehschien et al., 2015; Manton, & English, 2007; Ossbenblok et al., 2014). While scientific collaborations increase creativity, profoundness and replicability (Bergman, & Danheiser, 2016; Busch, & Hattery, 1956; Ductor, 2015), the concept of co-authorship also leaves room for pitfalls. Martinson et al. (2005) found that 10% of their questioned researchers have assigned authorship inadequately at least once in their career. Overall, scientists across disciplines perceive that problems related toauthorship happen ten times more likely than data fabrication or falsification (Marusic et al., 2011). Ghost authorship and honorary authorship represent two of the most infamous authorship malpractices because they falsify citation and publication counts, the "gold standard" (Altbach, 2015: 6) of academic productivity. In order to prevent adulterations of academic rankings, certain journals have started to demand statements from each co-author in which they disclose their specific contributions to the paper (McDonald et al., 2010). Especially high ranked journals in the nature and life sciences have already implemented requirements for so-called contribution disclosures upon paper submission (Sauermann, & Haeussler, 2017). However, an extensive literature and online search as well as reaching out to scholars from various social sciences indicate that social science journals do not require such statements. Two possible scenarios emerge from this: Either ghost authorship and honorary authorship do not occur in the social sciences, or social

[6] I wish to thank Prof. Christian Hopp for the collaboration and his inputs on this paper.

science journals have ignored the threat arising from these authorship issues. To answer the question on which of the two scenarios depicts the reality better, we need clear and comprehensive facts on the prevalence, distribution and motivational factors of ghost and honorary authorship in the social sciences.

2.1.1 Defining authorship

There exists neither a uniform definition nor a standard list of authorship criteria applicable to all scientific disciplines (Claxton, 2005). In fact, authorship guidelines vary even within the same discipline from journal to journal (Bates et al., 2004). Therefore, a universally valid definition of authorship can only be broad and imprecise. Such a generalized proposition would call researchers authors if they added "substantial contributions" (Osborne, & Holland, 2009: 4) to the publication. The perception of the term "substantial contribution" distinguishes the different definitions. The most common more precise definition of authorship criteria originates from the International Committee of Medical Journal Editors (ICMJE) (Wager, 2012). The Committee on Publication Ethics (COPE), a multidisciplinary advisory body on moral issues to which several large academic institutions, journals and societies subscribe[7], advises to use these standards (Albert, & Wager, 2004; Ioannidis et al., 2018; Wager, 2012). According to the ICMJE "Authorship credit should be based only on 1) substantial contributions to conception and design, or acquisition of data, or analysis and interpretation of data; 2) drafting the article or revising it critically for important intellectual content; and 3) final approval of the version to be published" (Davidoff, 2000: 230). Every researcher fulfilling all three requirements is not only eligible for but, in fact, must receive authorship (Davidoff, 2000).

2.1.2 Ghost authorship

The phenomeon of a person meeting all authorship criteria but not being awarded authorship is called ghost authorship (Kennedy et al., 2014). Clearly, ghost authorship shares similarities with plagiarism as both concepts incorporate the attribution of the work of one person to someone else (Liddell, 2003). The difference between those two malpractices lies within the attitude and knowledge of the creator of the original material. In plagiarism, the authentic writers do not have to be aware of the research paper "stealing" their ideas. Ghost authors, though, contribute consciously towards the research paper and (in-)voluntarily accept the decision of not

[7] Subscribers include the Academy of Management, the American Chemical Society and the Royal Society. A full list of all subscribers is available at https://publicationethics.org/members/publishers

receiving authorship (Bosch, & Ross, 2012). As nearly all academic rankings refer to citation and publication counts, putting large efforts into conducting research and writing a paper but in the end decline to receive author credits for this work does not seem reasonable. However, existing literature points towards three possible explanations of such behaviors. First, pressure from co-authors, whose citation and publication counts usually benefit from a lower number of contributors, can lead to researchers declining authorship. This situation compares to the observation of Harbring and Irlenbusch (2011) where individuals gained more in tournaments by tricking others into losses. Such behaviors can easily occur if the ghostwriter is a subordinate of one of the co-authors like it is the case in faculty-student collaborations (Oberlander, & Spencer, 2006). Second, scientists can voluntarily decide to reject author credits due to controversial findings that they might perceive as doubtful or weak and might hinder their future career (Bennet, & Taylor, 2003; Klein, & Moser-Veillon, 1999). Third, researchers might not aspire authorship to disguise potential conflicts of interests. For example, freelancers sponsored by pharmaceutical companies regularly approach life scientists asking them to write research papers based upon a provided, often biased bundle of articles and study results. The published articles should then foster the official approval and/or boost the sales of the drug. As the freelancers do not receive authorship (and thus are ghost authors), the monetary commitment of the firm stays secret (Brennan, 1994; Moffatt, & Elliott, 2007).

2.1.3 Honorary authorship

Honorary authorship refers to researchers who do not or only moderately contribute to a paper but receive authorship (Greenland, & Fontanarosa, 2012). Gift authorship and guest authorship represent alternative terms for this malpractice (da Silva, & Dobranszki, 2016). Honorary authorship per se is very beneficial to those receiving it because they can add publications to their CVs without providing the effort and time usually required to conduct research and write articles. As with ghost authorship, where individuals can exert pressure to force a contributor to withdraw from the authors list, here pressure from an individual to receive authorship can explain the occurrence of honorary authorship (Feeser, & Simon, 2008). Examples are senior scientists who demand authorship on the grounds of employing the original authors or providing financial funding (Drenth, 1998; Moffatt, 2011). This especially applies to faculty-student collaborations (Oberlander, & Spencer, 2006). Nevertheless, certain circumstances exist in which the original authors even voluntarily include honorary authors. The Matthew effect describes the phenomenon of including well-known researchers into author lists to increase the acceptance chances (Merton, 1968). This especially occurs in single-blind review processes (Blank, 1991). Furthermore, famous co-authors also often boost citations, which is beneficial to all authors (Valderas et al., 2007). To receive those benefits authors might ask coryphaei to co-author their papers although they did not participate in the research or writing process (da Silva, & Dobranszki, 2016). Third, the general increase in co-authored papers enables researchers to trade co-authorship. Hereby, a scientist adds only a small contribution towards a research project but still receives author credits. In return the original author also receives co-authorship when the honorary author from before publishes the next paper. This behavior occurs sometimes within chairs or research groups where, for example, reciprocal proofreading might induce honorary authorship (Padmanabhan, 2015).

2.1.4 Research Question

Existing research especially from the life sciences already discussed authorship malpractices and points towards the existence of these phenomena. For example, supervisor-subordinate collaborations seem to be more susceptible to authorship malpractices than collaborations amongst colleagues on eye-level (Bartle et al., 2000; Costa, & Gatz, 1992; Sandler, & Russel, 2005; Tryon et al., 2007)). Consequently, journal policies like enforcing contribution statements (Sauermann, & Haeussler, 2017) are unlikely to overcome misguided incentives within the academic publishing system, when supervisors can strong-arm their subordinates. We therefore need to better understand whether hierarchical power, institutions and norms actually induce authorship malpractices and how subordinates might succumb to supervisor influences. Consequently, we study authorship assignments in a field where journals do not require contributions statements, where the number of authors on average is tractable, and where there is (to the best of our knowledge) limited prior work on honorary and ghost authorship: the social sciences. The field specificity is of utmost importance as the results from the life sciences might not provide good approximations for the social sciences due to different incentives and competition patterns. The average acceptance rates of papers submitted to journals provide proof for this argument. In the social sciences, journals publish only around 20% of all submissions. In nature and life sciences, journals publish approximately 60% of all submission (Pfeffer, 1993). Therefore, we do not develop research hypotheses based on existing findings from non-comparable fields but instead conduct exploratory research on ghost and honorary authorship in the social sciences. More specifically we investigate the circumstances and factors that relate to authorship malpractices by comparing scholars' actual authorship assignments to their assignments in three hypothetical scenarios.

2.2 Materials and Methods

2.2.1 Survey Design

This paper bases on empirical findings derived from a survey consisting of three parts. In the first section, respondents answered questions about their demographic and job characteristics. The second part of the survey asked about the distribution of authorship and contributor acknowledgements in the latest published paper that names the respondents as authors. The third section covered three vignettes about research projects and asked respondents to assign authorship to the researchers.

2.2.1.1 Actual prevalence of authorship issues

To assess the existence of ghost and honorary authors, we asked respondents for the number of authors and number of other people (excluding peer reviewers) contributing to their last published paper. In the style of Mowatt et al. (2002), they specified for each of the authors and contributors (or for the top five in case of more than five authors or contributors) whether that person participated in creating the research design, the search for literature, analyzing the literature, collecting and/or preparing data, describing the results, writing the paper, reviewing and remarking the written paper and approving the final version of the paper. Only a person covering the last task, the antepenultimate and/or penultimate task and at least one of the other tasks should receive authorship according to the ICMJE authorship criteria (Davidoff, 2000). Based on these we identified ghost authors as contributors fulfilling the requirements and honorary authors as authors not fulfilling the requirements. Furthermore, respondents also needed to indicate on a scale from 0 (disagree) to 100 (agree) whether they agree with that for their last published paper all researchers making significant contributions also received authorship. We employed this question as an indicator for respondents' perceptions of ghost authors. The last question in the second part of the survey asked respondents to indicate on a scale from 0 (disagree) to 100 (agree) if researchers only received authorship if they participated actively in the creation process. Hereby, survey recipients stated their perceived degree of honorary authors.

2.2.1.2 Hypothetical assignment of ghost and honorary authorship

The assignment of ghost and honorary authorship in the vignettes should highlight the actual knowledge of authorship criteria among social scientists. We therefore developed three scenarios capturing different emergences of research papers. Ideas for these vignettes derived from Bartle et al. (2000), Costa, & Gatz (1992), Sandler, & Russel (2005) and Tryon et al. (2007) although the research focus and/or context differs from these studies. Before showing the survey participants the vignettes, Qualtrics randomly split the respondents into two groups. While both groups received the same first vignette to detect potential sample bias, each of the two groups then received a different version of the second and third vignette. After reading each of the vignettes, respondents decided for every person included in the scenario whether to award authorship to them.

The first vignette described a postdoc-professor collaboration on eye-level in which a student assistant helped in the data collection process.[8] According to the ICMJE authorship criteria the professor and the postdoc should receive authorship. The student assistant is not eligible for author credits.

In the second vignette respondents in group A assessed a postdoc/professor collaboration while respondents in group B assessed a professor/professor collaboration. Both versions again inlcuded a student assistant helping with the data collection. Hereby, the division of the workload shifts heavily towards the postdoc/Professor 1 while the second researcher just reaches the minimum threshold of the ICMJE authorship criteria by participating in the conception, reviewing and final approval. Still no author credits should go to the student assistant.

The third vignette focused on gender differences. Hereby, a professor published a paper based on a dissertation. In group A the PhD student was male and in group B the PhD student was female. In this case, only the professor fulfills the ICMJE authorship criteria because the PhD student does not engage in the paper writing and therefore does not meet the requirement of proofreading the paper before the journal submission. However, many large research societies including the Academy of Management (2018), the American Sociological Association (2018) and the American Psychological Association (1983) require PhD students to become

[8] The exact vignettes are available in the appendix.

even first author when published articles base on their dissertation. Therefore, also the PhD student should receive authorship.

2.2.2 Distribution and Sample

We designed the survey in spring 2018 using the online survey tool Qualtrics. In May 2018, we asked friendly scholars for feedback. After the incorporation of this feedback, a test-run took place by sending the survey link to 275 scholars who presented at least one paper at the European Accounting Association Annual Congress 2018. Data and feedback from the test-run showed that no further need for adaptions of the questionnaire existed. Therefore, the data gathered in the pilot phase are included in the analysis.

To ensure a large distribution of the questionnaire among researchers from various social sciences, we selected corresponding authors from published articles in well reknown expert journals as well as from papers presented at conferences organized by large field-specific research societies between January 2010 and June 2018. [9] In total, we gathered 126,480 unique email addresses after the deletion of duplicates. A random selection of half of these addresses led to an initial sample of 63,240. These scholars received an email containing a brief explanation of the purpose of the study and a Qualtrics URL-link to the questionnaire in late August and early September 2018. The link to the survey was the same for everyone to ensure the anonymity of respondents. Still, ballot stuffing was not possible because taking the survey was limited to once per IP-Address. After sending out the survey, 15,573 emails bounced automatically back due to the email addresses being no longer in use. The contacted sample therefore consists of 47,697 valid recipients.

The distribution of the survey resulted in 2,817 respondents. This is equivalent with a response rate of 5.91%, which compares to other recently conducted online surveys among scientists investigating academic misconduct (Hopp, & Hoover, 2017; Liao et al., 2018). Out of all respondents, 2,223 completed the survey. However, the samples used for the following analyses differ because some conference attendees have yet not published in journals or some respondents selected the N/A option at one or more questions. Furthermore, we exclude one

[9] A full list of the journals and conferences is available upon request from the corresponding author.

respondent from the analysis because this person stated to have been working in academia for 100 years although being only 67 years old.

2.2.3 Statistical Analysis

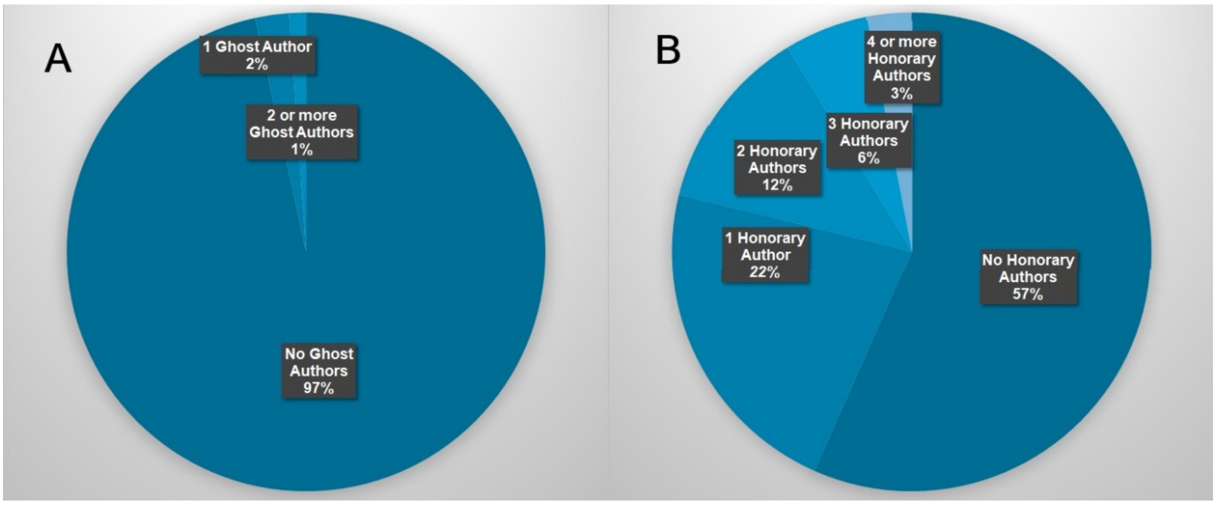
We conducted the whole data analysis using Stata 13. As our data sample largely exceeds the requirement of 100 data points for the application of the central limit theorem, we can refer to the dependent variables as normally distributed. We therefore employ logistic regressions if the dependent variable is dichotomous and OLS regressions if the dependent variable is interval-scaled.

2.3 Results

2.3.1 Actual Prevalence of Authorship Issues

Out of 1,878 papers with full data on ghost authorship, one ghost author participated in the creation of 43 (2.29%) papers and two or more ghost authors participated in the creation of 21 (1.12%) papers (Figure 2.1A). Honorary authorship occurs much more frequently with 418 papers (22.22%) containing one honorary author, 234 papers (12.44%) containing two honorary authors, 107 papers (5.69%) containing three honorary authors and 57 (3.03%) containing four or more honorary authors (Figure 2.1B). Hence, one in about thirty papers suffers from ghost authorship and nearly every second paper suffers from honorary authorship.[10]

Figure 2.1: Share of ghost (A) and honorary (B) authorship.

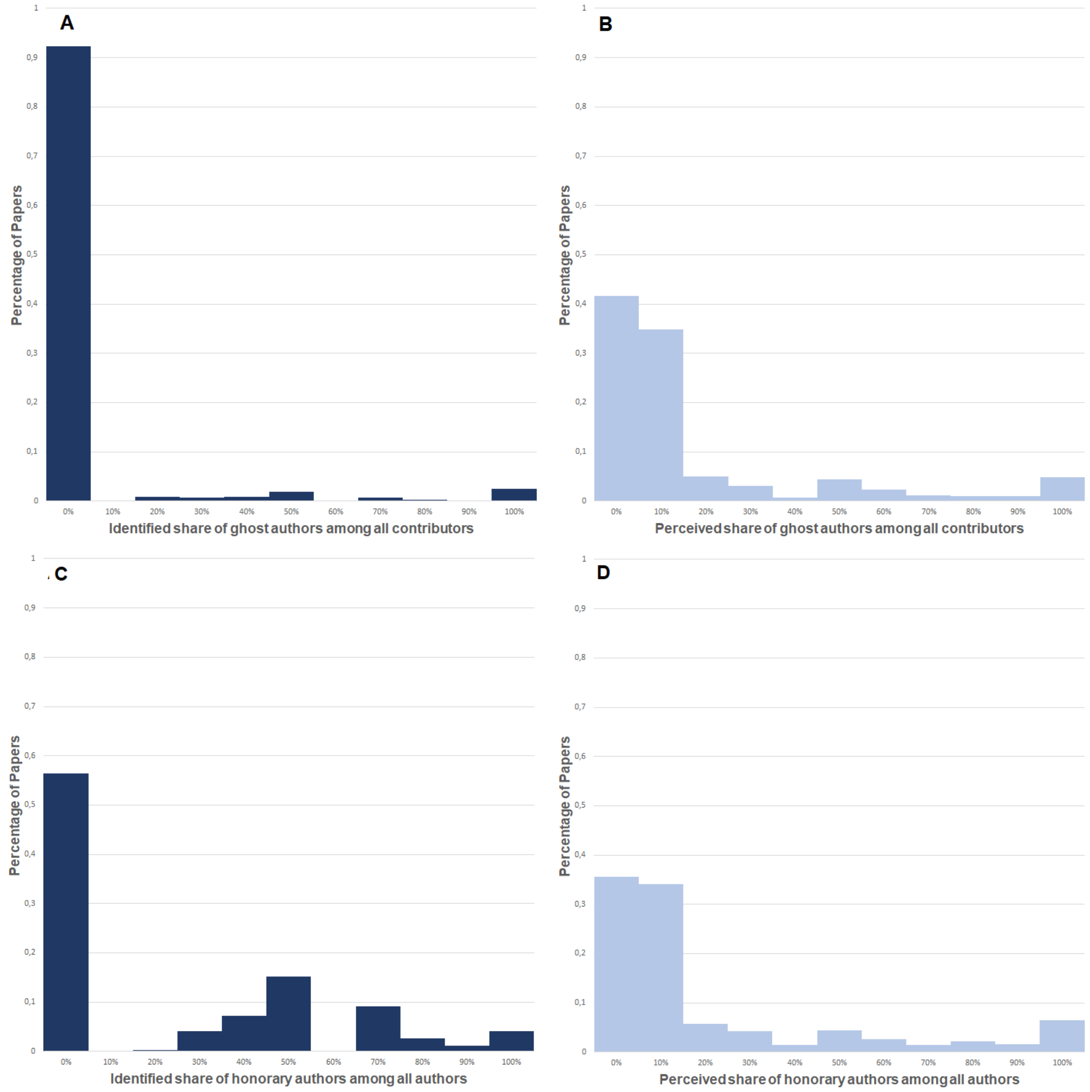


Note: N. of obs. are 1,878 for (A) and 1,881 for (B).

[10] Descriptive statistics are available in Table 2.4 and Table 2.5 in the appendix.

We compare the share of identified ghost authors to respondents' assessments of the share of ghost authors in their last published paper. The histograms in Figure 2.2 show that the perceived share of ghost authors (Figure 2.2B) exceeds the identified share of ghost authors (Figure 2.2A). In other words, our respondents perceive ghost authors more often than they actually exist.[11] This difference is significant according to a two-sided t-test (t=-8.0874; df=812; p=0.0000).

Figure 2.2: Histograms of the identified share of ghost (A) and perceive share of ghost authors (B) as well as the identified share of honorary (C) and perceived share of honorary authors (D)



Note: N. of obs. are 813 for (A) and (B) and 1,842 for (C) and (D).

---

[11] The number of observations is only 813 because we assess the share of ghost authors among all non-author contributors.

Figure 2.2 also compares the share of identified honorary authors in each paper to respondents' assessments of the share of honorary authors in their last published paper. The difference between the identified (Figure 2.2C) and perceived (Figure 2.2D) occurrences is not as stark as before. Nevertheless, using a two-sided t-test we find that the perceived share of honorary authors is significantly lower than the identified share of honorary authors ($t=7.5946$; $df=1841$; $p=0.0000$).

To identify the reasons for the mismatches between scholars' perception of the prevalence of authorship issues and their actual occurrence, we investigate their antecedents and correlates. Table 2.1 depicts the regression results investigating the effects of the exploratory variables. Model 1 and 2 contain logistic regressions on a dummy variable indicating whether researchers' last papers include at least one ghost author (1) or honorary author (2). Models 3 and 4 depict linear regressions with the number of ghost and honorary authors in each paper as the dependent variables. We find that Anglophone and Continental European scholars exhibit less (often) ghost and honorary authors than scholars from other world regions. Furthermore, PhD Students indicate increased occurrences of both authorship issues in their last published paper. Yet professors indicate less honorary authors. Scholars who published more papers in the past three years encounter moderately significantly more (often) honorary authors. Concerning field specific differences, we find on the one hand that economics and finance researchers as well as political scientists are less likely to have ghost and honorary authors in their last published paper than researchers from other fields like business, computer or operations research. On the other hand, Model4 highlights that Psychologists include more honorary authors in their published papers.

Models 5 and 6 depict respondents' assessments of the share of ghost and honorary authors in the last published paper as the dependent variables. The results show that scholars generally identify ghost and honorary authors correctly; a higher share of ghost and honorary authors is associated with a higher perceived share. Nevertheless, the coefficients of 0.142 for ghost authors and 0.187 for honorary authors indicate that there exists a mismatch between the perceived and the actual inclusion of honorary and ghost authors

Table 2.1: Regression Results of Actual and Perceived Prevalence of Ghost and Honorary Authorship

	Model 1 Ghost Authorship	**Model 2** Honorary Authorship	**Model 3** # of Ghost Authors	**Model 4** # of Honorary Authors	**Model 5** Perceived Ghost Authors	**Model 6** Perceived Honorary Authors
Share of Ghost Authors					0.143**	
					(0.050)	
Share of Honorary Authors						0.187***
						(0.023)
Female	-0.324	0.175†	-0.020	0.079	0.024	0.009
	(0.302)	(0.106)	(0.017)	(0.053)	(0.020)	(0.015)
Anglophone	-1.619***	-0.964***	-0.118***	-0.520***	-0.095**	-0.056†
	(0.456)	(0.206)	(0.033)	(0.102)	(0.036)	(0.029)
Continental Europe	-1.227**	-0.786***	-0.114***	-0.474***	-0.088*	-0.082**
	(0.419)	(0.203)	(0.032)	(0.101)	(0.036)	(0.028)
Developing Countries	0.031	0.009	-0.000	0.004	-0.001	-0.001
	(0.020)	(0.008)	(0.001)	(0.004)	(0.002)	(0.001)
Age	0.259	0.131	-0.008	-0.100	-0.010	-0.060†
	(0.405)	(0.231)	(0.036)	(0.114)	(0.038)	(0.032)
PhD Student	1.268**	0.513*	0.110***	0.278**	0.008	0.061*
	(0.440)	(0.204)	(0.032)	(0.101)	(0.035)	(0.028)
Professor	-0.201	-0.287*	-0.004	-0.171**	-0.015	-0.032†
	(0.343)	(0.128)	(0.020)	(0.063)	(0.024)	(0.018)
Editor	0.245	0.067	0.012	0.083	0.002	0.008
	(0.307)	(0.121)	(0.019)	(0.060)	(0.022)	(0.017)
Years in Academia	-0.009	-0.011	0.001	-0.007	0.001	-0.001
	(0.021)	(0.008)	(0.001)	(0.004)	(0.002)	(0.001)
Published Papers	-0.023	0.074†	-0.007	0.047*	0.009	0.004
	(0.108)	(0.042)	(0.007)	(0.021)	(0.008)	(0.006)
Written Reviews	0.063	-0.010	0.004	-0.018	-0.017**	-0.004
	(0.096)	(0.036)	(0.006)	(0.018)	(0.007)	(0.005)
Business	-0.397	0.216	-0.037	0.063	0.028	-0.010
	(0.399)	(0.166)	(0.026)	(0.083)	(0.030)	(0.023)
Economics and Finance	-0.515	-0.430*	-0.055†	-0.251**	-0.012	-0.047†
	(0.500)	(0.199)	(0.031)	(0.097)	(0.036)	(0.027)
Computer and Statistics	0.056	0.319†	0.000	0.134	-0.034	-0.037
	(0.407)	(0.188)	(0.030)	(0.094)	(0.036)	(0.026)
Political Sciences	-0.967	-0.625**	-0.067*	-0.316**	0.019	-0.086**
	(0.670)	(0.222)	(0.034)	(0.106)	(0.041)	(0.030)
Psychology	-0.410	0.321	-0.021	0.321**	-0.029	0.012
	(0.676)	(0.240)	(0.039)	(0.121)	(0.044)	(0.034)
Sociology	-0.334	-0.241	-0.026	-0.096	0.039	-0.023
	(0.605)	(0.226)	(0.036)	(0.111)	(0.042)	(0.031)
Chi-Square	55.30	145.69				
P > Chi-Square	0.00	0.00				
Pseudo R-squared	0.01	0.06				
F-Value			9.51	3.52	2.83	8.92
P > F-Value			0.00	0.00	0.00	0.00
R-squared			0.03	0.08	0.06	0.08
Observations	1854	1857	1854	1857	804	1818

† $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Note: Models 1 and 2 present marginal effects derived from logistic regressions with standard errors in parentheses. Models 3, 4, 5 and 6 present coefficients derived from linear regressions with standard errors in parentheses.

2.3.2 Hypothetical Assignments of Authorship

We show above that ghost and honorary authorship exist in the social sciences. However, it is unclear whether the occurrences of these authorship issues derive from a lack of knowledge of authorship criteria or from conscious practices that are embedded in the reasons outlined in the beginning. To answer this question, we split the same respondents who were asked about the prevalence of ghost and honorary authorship above, into two groups and presented them three vignettes. Table 2.2 depicts respondents' authorship assignments for the three hypothetical scenarios.

Table 2.2: Authorship Assignments in the Vignettes Split by Treatment Group

	Group 1	Group 2
Total Assessments Vignette 1	973	984
Professor Vignette 1	935	937
Postdoc Vignette 1	955	968
Student Assistant Vignette 1	284	292
Total Assessments Vignette 2	975	986
Postdoc/Professor Vignette 2	950	963
Professor Vignette 2	607	700
Student Assistant Vignette 2	82	87
Total Assessments Vignette 3	964	987
Professor Vignette 3	756	732
PhD Student Vignette 3	919	945

Note: The numbers correspond to the number of respondents who assigned authorship to the respective figure in the respective vignette.

The first vignette lists the same scenario for both groups: A professor and a postdoc collaboratively write a paper and a student assistant supports the data collection. Unsurprisingly, the answers of both groups are nearly identical. As shown in Table 2.2, almost all respondents (917 in the first group and 923 in the second group) award authorship correctly to the professor and the postdoc. Still, a substantial number of scholars (284 in the first group and 292 in the second group) propose a case of honorary authorship by incorrectly awarding author credits to the student assistant. Nevertheless, Models 1, 2 and 3 in Table 2.3 show that the assessments do not differ significantly between the groups when we control for demographic and job-related factors as the coefficient of *Group* is insignificant. This is important for the subsequent analyses when each group judges slightly different scenarios.

Table 2.3: Regression Results of Hypothetical Authorship Assignments in the Vignettes

	Vignette 1			Vignette 2			Vignette 3	
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**	**Model 8**
	Prof.	Postd.	SA	Prof./ Postd.	Prof.	SA	Prof.	PhD
Group	-0.268	0.025	0.012	-0.006	0.429**	0.082	0.080	-0.248*
	(0.233)	(0.362)	(0.103)	(0.302)	(0.101)	(0.167)	(0.223)	(0.111)
Female	-0.014	-0.411	0.010	-0.133	-0.060	-0.193	-0.072	-0.009
	(0.247)	(0.390)	(0.113)	(0.321)	(0.110)	(0.189)	(0.239)	(0.120)
Anglophone	1.081*	2.603**	0.044	0.709	-0.195	-0.017	0.788†	-0.333
	(0.482)	(0.623)	(0.225)	(0.499)	(0.232)	(0.352)	(0.411)	(0.249)
Continental	0.062	1.705**	0.299	1.052*	-0.772***	0.080	0.637	-0.127
Europe	(0.437)	(0.469)	(0.220)	(0.515)	(0.227)	(0.341)	(0.401)	(0.247)
Developing	-0.350	1.051*	0.631**	0.145	-0.503*	0.513	-0.290	-0.125
Countries	(0.460)	(0.509)	(0.243)	(0.540)	(0.255)	(0.363)	(0.414)	(0.279)
Age	-0.064***	-0.027	0.011	0.008	0.001	0.006	-0.016	0.013
	(0.015)	(0.030)	(0.009)	(0.025)	(0.008)	(0.014)	(0.017)	(0.010)
PhD Student	-0.320	-1.036†	0.117	-0.713	0.260	-0.267	-0.376	0.135
	(0.405)	(0.609)	(0.199)	(0.482)	(0.196)	(0.359)	(0.375)	(0.212)
Professor	0.136	0.197	-0.181	-0.118	-0.194	-0.198	0.026	-0.127
	(0.294)	(0.499)	(0.135)	(0.412)	(0.134)	(0.214)	(0.293)	(0.148)
Editor	-0.945***	-0.597	0.070	-0.337	0.094	0.370†	-0.037	-0.119
	(0.265)	(0.414)	(0.129)	(0.366)	(0.130)	(0.196)	(0.282)	(0.141)
Years in	0.044**	0.004	0.001	-0.012	0.002	0.002	0.012	-0.011
Academia	(0.016)	(0.031)	(0.009)	(0.026)	(0.009)	(0.014)	(0.018)	(0.010)
Published	-0.009	-0.188	0.083*	-0.140	0.136**	0.255***	-0.164†	0.120*
Papers	(0.101)	(0.130)	(0.042)	(0.119)	(0.045)	(0.064)	(0.088)	(0.050)
Written	0.319**	0.036	-0.037	0.236†	0.004	-0.169**	0.211*	0.079†
Reviews	(0.109)	(0.133)	(0.039)	(0.128)	(0.038)	(0.065)	(0.096)	(0.043)
Business	0.339	0.503	-0.804***	-0.325	-0.058	-0.673*	-0.056	0.292
	(0.340)	(0.638)	(0.171)	(0.420)	(0.172)	(0.281)	(0.313)	(0.190)
Economics	1.461*	0.683	-0.808***	0.599	-0.089	-0.067	0.429	-0.059
Finance	(0.582)	(0.861)	(0.210)	(0.685)	(0.202)	(0.309)	(0.445)	(0.218)
Computer and	0.813†	-0.830	0.345†	0.732	0.642**	0.595*	0.403	0.815**
Statistics	(0.436)	(0.564)	(0.183)	(0.618)	(0.208)	(0.266)	(0.394)	(0.236)
Political	-0.274	-0.696	-0.514*	-0.393	-0.959***	-0.871*	0.101	-0.889***
Sciences	(0.399)	(0.675)	(0.218)	(0.552)	(0.213)	(0.425)	(0.448)	(0.221)
Psychology	0.791	0.762	-0.637*	Perfect	1.044**	-1.315*	Perfect	1.383**
	(0.668)	(1.127)	(0.260)	Predictor	(0.318)	(0.560)	Predictor	(0.386)
Sociology	0.076	0.894	-0.225	0.107	-0.506*	-0.088	0.049	-0.521*
	(0.435)	(1.120)	(0.221)	(0.687)	(0.221)	(0.352)	(0.469)	(0.234)
Chi-Square	81.01	47.40	120.42	20.69	169.26	86.48	29.63	129.74
P > Chi-	0.000	0.000	0.000	0.241	0.000	0.000	0.029	0.000
Pseudo R-	0.117	0.142	0.051	0.048	0.069	0.076	0.042	0.061
Observations	1931	1931	1931	1935	1935	1935	1926	1926

† $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Note: Coefficients correspond to marginal effects derived from logistic regressions with standard errors in parentheses.

The assignments to the primary researcher do not differ substantially between the groups in the second vignettes. However, respondents in both groups award authorship less often to the secondary researcher. Looking at the exact numbers we find that only 607 researchers in Group 1, who assessed a postdoc-professor collaboration, assigned authorship to the second researcher (a professor). Meanwhile, 700 researchers in Group 2, who assessed a professor-professor collaboration, assigned authorship to the second researcher. This difference stays highly significant even when we control for various demographic and job-related factors (Model 5 in Table 2.3). Continental European social scientists as well as social scientists from developing countries are more likely to withhold authorship from the second researchers. In turn, scientists that are more prolific more often correctly assign authorship to the second researcher. The same applies to computer, operations and statistics researchers as well as psychologists. In fact, all psychologists in our study correctly assigned authorship to the primary researcher. Political scientists and sociologists though are more likely to withhold authorship from the second researcher just merely meeting the ICMJE authorship criteria.

The third vignette differs from the first two as it does not refer to a collaboration on a specific research project but instead covers a dissertation that provides the basis for a research paper written by the supervisor. According to Table 2.2, 919 (945) respondents in Group 1 (Group 2) gave author credits to the PhD student and 756 (732) respondents gave author credits to the professor. Overall, 231 survey participants in Group 1 and 272 survey participants in Group 2 award authorship in a way that includes ghost authorship. Model 7 in Table 2.3 points out that respondents in Group 2 award statistically significant less often authorship to the professor than respondents in Group 1 although the only difference between the two groups lies in the gender of the PhD student with Group 1 reading about a male PhD student and Group 2 reading about a female PhD student. We again investigate different assessments among the exploratory variables. We find that Anglophone researchers moderately significantly more often award authorship to the PhD student. In addition, more prolific researchers are moderately significantly less likely to give author credits to the PhD student and are in turn more likely to give author credits to the professor. Scholars who wrote more reviews in the past three years assign authorship more often to the PhD student and moderately significantly more often correctly to the professor. Computer, operations and statistics researchers are highly significantly more likely to give author credits to the professor. Yet political scientists and sociologists exhibit highly significantly greater chances of suggesting ghost authorship by not awarding authorship to the professor. Like before, all psychologists assign authorship to the PhD student. Moreover, they also give highly significantly more often author credits to the professor.

To detect potential hidden influencing factors, we create interaction terms with the product of *Group* and all exploratory variables for Models 4 through 8 in Table 2.3 and run again logistic regressions. The results are available in Table 2.6 in the appendix. Most notably, we find that scholars in Developing Countries are less likely to assign authorship to a PhD student if she is female. Additional fine-tuning of the analysis by calculating individual regressions by regions reveals that respondents from Western Europe are significantly more likely to award authorship to the PhD student if she is female.

2.3.3 Robustness of Results

We conduct several robustness and endogeneity checks. First, we run all models with robust standard errors. This changes no significance levels. Second, we apply firthlogit, a special form of a logistic regression that considers rare events, to correct for the relatively low amount of 64 ghost authorship observations (Williams, 2018). This adaption does not change any significance levels. Third, the average team size differs across research fields (Wren et al., 2007). Therefore, we run the same regressions with the percentages of honorary and ghost authors in the supplementary material. The only difference is that psychologists do not face more honorary authors anymore. Fourth, we test the approximation of six authors and contributors for respondents stating that more than five authors and/or contributors participated in their last published paper. Hereby, we run all analyses including only respondents of papers with at maximum five authors and contributors. The significance levels and implications do not differ from the findings above. Last, model (5) might contain sample selection bias since honorary authorship increases the number of authors and lowers the number of contributors because a contributor receives authorship. The inclusion of papers with only at least one contributor in the analysis of the relationship between perception and occurrence of ghost authors may therefore suffer from endogeneity. To overcome this issue, we employ a Heckman two-step regression by applying the number of honorary authors as the selection variable. This returns an insignificant inverse Mills ratio. Therefore, our results do not suffer from sample selection bias (Heckman, 2001).

2.4 Discussion

Ghost and honorary authorship occur in the social sciences with the honorary authorship occurring much more frequently than ghost authorship. The different degrees of prevalence might have its root cause in the academic incentive system. Individual researchers benefit from receiving honorary authorship for an additional publication without spending the time and effort required to fulfill the authorship criteria. Yet, they squander the time and effort spent on a research project for which they do not receive authorship despite fulfilling the authorship criteria if they are ghost authors. In addition, social scientists' perception of the occurrence of authorship malpractices provide grounds to suspect that not everyone might be aware of what constitutes authorship. Our findings from the vignettes support this stipulation as a substantial number of respondents assign ghost and/or honorary authorships. For example, fairness considerations might lead to some researchers assessing authorship not based on objective authorship criteria but instead might become sensitive towards hierarchical pressure. This might explain why respondents assign ghost authorship more often in the Postdoc-Professor case than in the Professor-Professor case.

In the third vignette, we show that social scientists assign authorship to PhD students if published papers base upon dissertations. This accords to the authorship guidelines of their academic societies: Articles based on dissertations should always include the PhD students as authors. Contrarily to this assessment we find that actual papers authored by PhD students include more (often) ghost and honorary authors. This difference between the high share of correct authorship assessments in the vignettes and the higher likeliness of PhD students to face authorship issues, hints towards the existence of questionable practices: Senior scientists incorrectly withholding authorship from their junior colleagues despite knowing that they actually should receive authorship. Furthermore, the gender differences in the responses to the third vignette might also indicate that social scientists themselves are aware of such power abusing practices. As the survey was conducted in the aftermath of the Weinstein scandal, respondents might have been more aware towards the abuse of power towards females and hence have more carefully reviewed the case of a female PhD student than of a male PhD student (Sigurdsson, 2018).

Hierarchical pressure might also explain regional differences in the prevalence of ghost and honorary authorship. We find that Anglophone and Central European scholars exhibit those authorship malpractices less often. This implies that researchers in the baseline category consisting mainly of Asian researchers more often include ghost and honorary authors in their papers. The cultural background of these scientists coming from countries with generally stronger obedience towards supervisors and the generally higher degree of respect towards experienced persons might explain this phenomenon as sometimes department or faculty heads receive authorship even without having read the paper (Salita, 2010).

Awarding authorship to department or faculty heads might also explain why researchers publishing more papers more often exhibit honorary authors in their papers and assign honorary authorship more often in the vignettes. For example, if supervisors put pressure on their subordinates to name them as authors despite not fulfilling authorship criteria, their name appears on more published papers although they are, in fact, honorary authors. Ioannidis et al. (2018) raised this issue by surveying "scientists who publish a paper every five days" which equals 72 or more published papers per year. More than 70% of their respondents state that they did not conduct at least one of the three required tasks of the ICMJE authorship criteria in at least every fourth paper.

Besides these discussed effects stemming from the academic incentive system and hierarchical pressure, we also show that the prevalence of ghost and honorary authorship varies across research fields. Different authorship orders might explain this. Economic, finance and political science represent disciplines that usually rank authors alphabetically (Laband, 2002; Lake, 2010). Consequently, awarding someone else authorship increases the likeliness of falling under "et al." in the citations. The actual contributors seek to avoid too many authors on their papers because falling under "et al." would diminish their author credits. This induces an incentive against including honorary authors. At the same time scholars from fields with strong preferences for ordering authors by contributions, for example psychologists, sometimes even include the individual in charge for financing the research project (e.g. the department head) as last-author even if that individual did not participate in the research process itself (Yu-Wei, 2019). If someone receives authorship just for providing the resources sometimes without even reading the paper before submission this constitutes a clear act of honorary authorship.

2.4.1 Implications for academic practice

The existence of ghost and honorary authorship in the social sciences highlights the necessity of introducing contribution disclosures also for these academic disciplines. Contribution disclosures allow insights on the workload distributions among author teams. Hence, they might help to reduce honorary authorship by giving editors, neutral individuals outside of the author teams, the chance to assess whether all authors also fulfill the authorship criteria. In turn, contribution disclosures might also reduce the share of senior scientists that receive authorship despite not contributing substantially to manuscripts as they cannot completely free ride on their subordinates' research projects but instead would need to write contribution disclosures. Of course, the introduction of contribution disclosures requires editors (and maybe reviewers) to be aware of authorship criteria and malpractices. To increase their knowledge, we recommend academic societies, research institutions and publishers to provide tutorials and workshops on authorship as our results from the vignettes indicate that not all scholars are aware of what tasks (do not) qualify for authorship.

Our results highlight that authorship assignments might suffer from abuse of power as PhD students more often have honorary authors in their papers. However, we also show that many scholars are aware of this issue. We therefore call upon academic societies, research institutions and publishers to implement whistleblowing platforms that allow anonymous reporting of authorship malpractices. In turn, this creates the need for journal policies to allow author changes also after submission, and, in some cases, even after publication.

Last, we show that there exist discrepancies between authorship and effort in some cases. Authorship represents the primary indicator used in citation and publication counts. We thus call upon research institutions and other employers to not solely rely on these estimates in hiring and tenure processes. Instead, applicants' capabilities and characteristics should also be screened similar to hiring processes in the private sector. This would ensure more comprehensive insights leading to better hiring and tenure decisions. At the same time, this would also reduce the importance of receiving authorship and, hence, reduce the incentives to become an honorary author.

2.4.2 Limitations and Future Research

The strongest limitation of this study lies within the fact that no general applicable authorship criteria exist. Therefore, we applied the ICMJE authorship criteria that derive from the life sciences. Although many universities and societies apply them, the ICMJE authorship criteria are not suitable for all social science research fields. For example, authorship criteria might differ in less data related disciplines like theoretical sociology or political sciences. Another restriction of this study may derive from the relatively low amount of identified ghost authors. While the application of firthlogit indicates that the existing findings are robust and valid, the small number of observations might hide further effects. The unequal distribution of researchers among geographical regions and research fields might also reduce the studies' expressiveness for less covered groups like Caribbean scholars or law researchers. We also did not ask respondents for qualitative reasoning of their answers in order to keep the questionnaire as anonymous as possible. However, even this quantitative survey contains sensitive questions that even with assuring the respondents anonymity might induce understatements of actual wrongdoings. This represents a common problem of research on questionable research practices and academic misconduct (Persoskie, & Nelson, 2013). However, since these understatements actually indicate less frequent occurrences and assignments of ghost and honorary authorship our findings present at least a lower boundary on the existence of and the attitude towards these authorship issues. Generally, most of our limitations hinder the study only from gaining more expressiveness. Therefore, the presented results are robust and provide meaningful implications.

Several areas of future research derive from this study. First, comparing research field specific authorship criteria could explain why scholars from different disciplines vary in their authorship assessments. Second, examining the authorship assignments of extremely prolific social science scholars in-depth could clear up any doubts on whether they really exhibit higher productivity levels or just receive honorary authorship more often. Third, a qualitative study asking researchers that experienced ghost and/or honorary authorship could return more detailed reasonings. Last, the application of item-sum-techniques could increase respondents' perceived anonymity. This would allow for more accurate data that might indicate more accurate assessments of the actual occurrences of ghost and honorary authorship.

2.5 Conclusion

Based on findings from the life sciences we ask the question whether honorary and ghost authorship exist in the social sciences. Our results show that these authorship malpractices occur also quite frequently in the social sciences. By investigating correlates of these malpractices, we identify that hierarchical pressure might represent a driving force as social scientists are generally aware of authorship criteria but do not always accord with them. Several possible solutions exist: A widespread introduction of authorship disclosure statements as a pre-requirement for publishing would enable editors and reviewers to check authors' contributions. The implementation of whistleblowing platforms would allow scholars to anoumisly report abusive authorship behaviors. Last, we call for shifting the focus of hiring and tenure procedures away from merely counting citation and publication towards focusing, assessing and testing the actual skill sets of the candidates.

2.6 Appendix

2.6.1 Texts of the vignettes employed in the study

2.6.1.1 First Vignette

A Postdoc has a research idea and proposes this to the professor. They together set up the research design. Afterwards the postdoc engages in the search for literature and summarizes the findings. Based on this the professor sets up a survey which is conducted and documented by a student assistant. The data is afterwards statistically analyzed by the professor who passes the results on to the postdoc. The postdoc then writes a journal paper which is reviewed by the professor before the submission.

2.6.1.2 Second Vignette

A Postdoc/Professor 1 has a research idea and creates the research design. This research design is then presented to the professor/Professor 2. The professor/Professor 2 revises the research design and suggests that the survey could be conducted in one of his larger classes instead of using an internet survey tool. The postdoc/Professor 1 engages in the search for literature and summarizes the findings. Based on this the postdoc/Professor 1 sets up a survey and conducts it in the professor's/Professor 2's lecture together with the help of a student assistant. The postdoc/Professor 1 then inputs the data into a statistical program and analyzes it. The results are presented to the professor/Professor 2 who gives comments about using two additional statistical tools to test for robustness and sample bias. The Postdoc/Professor 1 includes these suggestions and writes a journal paper which is proof-read by the professor/Professor 2 before the submission.

2.6.1.3 Third Vignette

A professor approaches a PhD student with an idea for his/her dissertation. The PhD student likes the idea and sets up a research design. He/She presents this to the professor, who suggests a different data collection method. The PhD student analyzes the literature and shows the professor his/her findings. The professor recommends the inclusion of five additional papers. The PhD student includes these papers and conducts the data collection in accordance with the professor's suggestions. The PhD student analyzes the collected data using a statistical software tool. He/She presents the empirical findings to the professor, who suggests to use two additional statistical tests in order to ensure significant and robust findings. The PhD student implements this suggestion and writes up his/her thesis. Before the final submission of the thesis the professor reviews the thesis and gives small comments about writing style and grammar use. After the successful graduation of the PhD student the professor creates a scientific paper based on the empirical results of the thesis and submits it to a journal.

2.6.2 Descriptive statistics

Table 2.4: Descriptive Statistics for Dichotomous Variables

	Number	Share
All Respondents	2222	100,00%
Female	738	33,21%
Anglophone (British Isles, North America, Australia & NZ)	866	38,97%
Continental Europe	900	40,50%
Developing Countries (Latin America, Africa, Southeast Asia)	300	13,50%
PhD Students	204	9,18%
Professors	1156	52,03%
Editors	541	24,35%
Business Researchers	789	35,51%
Economics and Finance Researchers	291	13,10%
Computer, Operations and Statistics Researcher	352	15,84%
Political Scientists	214	9,63%
Psychologists	136	6,12%
Sociologists	184	8,28%

Table 2.5: Descriptive Statistics for Integer Variables

	N	Mean	SD	Minimum	Maximum
Age	2215	46.52	13.03	21	93
Academic Working Years	2222	17.46	12.49	0	63
Papers Published	2222	6.40	7.74	0	60
Reviews Written	2222	6.84	9.37	0	60
# Authors in Last Paper	2052	2.83	1.25	1	6
# Contributors in Last Paper	2052	1.16	1.64	0	6

2.6.3 Additional regression

Table 2.6: Regression Results of Hypothetical Authorship Assignments in the Vignettes Including Interaction Terms

	Vignette 2			Vignette 3	
	Model 1	Model 2	Model 3	Model 4	Model 5
	Prof./ Postd.	Prof.	SA	Prof.	PhD
Group	15.084	-0.602	-2.632*	-0.034	-0.195
	(1262.236)	(0.828)	(1.308)	(1.642)	(0.914)
Female	-0.350	-0.203	-0.306	-0.179	-0.136
	(0.463)	(0.159)	(0.293)	(0.340)	(0.181)
Anglophone	1.545*	0.051	-0.046	1.196*	-0.399
	(0.603)	(0.316)	(0.470)	(0.512)	(0.370)
Continental Europe	1.630**	-0.792*	-0.475	0.782	-0.020
	(0.597)	(0.309)	(0.469)	(0.483)	(0.369)
Developing Countries	1.290†	-0.470	0.328	0.645	-0.386
	(0.755)	(0.352)	(0.493)	(0.577)	(0.413)
Age	0.008	-0.007	-0.035	-0.023	0.026†
	(0.036)	(0.012)	(0.022)	(0.024)	(0.015)
PhD Student	-0.463	0.436	0.046	-0.697	0.158
	(0.703)	(0.279)	(0.465)	(0.499)	(0.307)
Professor	-0.381	-0.111	-0.151	0.150	-0.223
	(0.599)	(0.185)	(0.320)	(0.419)	(0.219)
Editor	-0.619	0.052	0.702*	-0.041	-0.300
	(0.497)	(0.181)	(0.283)	(0.393)	(0.205)
Years in Academia	0.008	-0.008	0.031	0.014	-0.027†
	(0.038)	(0.013)	(0.022)	(0.026)	(0.015)
Published Papers	0.013	0.143*	0.176†	-0.154	0.158*
	(0.179)	(0.063)	(0.097)	(0.123)	(0.076)
Written Reviews	0.055	-0.015	-0.143	0.145	0.071
	(0.161)	(0.053)	(0.095)	(0.128)	(0.063)
Business	-0.119	0.009	-0.764†	0.028	0.029
	(0.586)	(0.236)	(0.443)	(0.428)	(0.290)
Economics and	0.297	-0.337	0.473	0.200	-0.219
Finance	(0.867)	(0.276)	(0.433)	(0.577)	(0.333)
Computer and	0.758	0.823**	0.812*	0.406	0.648†
Statistics	(0.867)	(0.290)	(0.398)	(0.546)	(0.362)
Political Sciences	-0.598	-1.342***	-1.066	0.240	-1.265***
	(0.711)	(0.303)	(0.679)	(0.622)	(0.329)
Psychology	Perfect	1.299**	-0.713	Perfect	1.548*
	Predictor	(0.456)	(0.692)	Predictor	(0.649)
Sociology	14.191	-0.804*	0.000	0.730	-0.818*
	(1301.751)	(0.314)	(0.550)	(0.803)	(0.354)
Female	0.542	0.256	0.208	0.295	0.210
X Group	(0.657)	(0.223)	(0.389)	(0.487)	(0.244)
Anglophone	-15.996	-0.473	0.020	-1.278	0.094
X Group	(1262.235)	(0.473)	(0.719)	(0.942)	(0.502)
Continental Europe	-15.408	0.023	1.050	-0.619	-0.205
X Group	(1262.235)	(0.463)	(0.701)	(0.928)	(0.498)
Developing Countries	-16.358	-0.022	0.341	-2.067*	0.439
X Group	(1262.235)	(0.522)	(0.743)	(0.966)	(0.563)
Age	0.004	0.015	0.069*	0.015	-0.023
X Group	(0.051)	(0.017)	(0.028)	(0.034)	(0.019)
PhD Student	-0.295	-0.394	-0.950	0.815	-0.053
X Group	(0.980)	(0.394)	(0.794)	(0.789)	(0.428)
Professor	0.574	-0.196	-0.067	-0.158	0.157
X Group	(0.839)	(0.272)	(0.440)	(0.595)	(0.299)
Editor	0.609	0.125	-0.525	0.028	0.315
X Group	(0.761)	(0.265)	(0.398)	(0.576)	(0.285)

Years in Academia X Group	-0.036 (0.053)	0.018 (0.018)	-0.048† (0.029)	-0.004 (0.036)	0.028 (0.020)
Published Papers X Group	-0.285 (0.239)	0.008 (0.091)	0.168 (0.132)	-0.065 (0.180)	-0.056 (0.101)
Written Revies X Group	0.373 (0.259)	0.043 (0.077)	-0.054 (0.132)	0.153 (0.193)	0.015 (0.086)
Business X Group	0.139 (0.857)	-0.157 (0.349)	0.117 (0.578)	0.357 (0.638)	0.481 (0.386)
Economics and Finance X Group	1.032 (1.439)	0.509 (0.413)	-1.196† (0.645)	0.855 (0.918)	0.299 (0.444)
Computer and Statistics X Group	0.358 (1.251)	-0.394 (0.420)	-0.402 (0.542)	0.434 (0.797)	0.294 (0.479)
Political Sciences X Group	1.018 (1.159)	0.713 (0.435)	0.252 (0.877)	0.183 (0.908)	0.726 (0.449)
Psychology X Group	Perfect Predictor	-0.507 (0.639)	-1.572 (1.269)	Perfect Predictor	-0.226 (0.810)
Sociology X Group	-14.342 (1301.751)	0.535 (0.453)	-0.197 (0.723)	-0.749 (1.012)	0.509 (0.474)
Chi-Square	40.93	203.39	112.75	42.91	141.08
P > Chi-Square	0.162	0.000	0.000	0.116	0.000
Pseudo R-squared	0.095	0.082	0.099	0.062	0.067
Observations	1877	1935	1935	1868	1926

† $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Note: Coefficients correspond to marginal effects derived from logistic regressions with standard errors in parentheses. *X Group* indicates interaction terms created as the product of the respective variable and *Group*.

3 Essay 2 - Team authorship in the social sciences

3.1 Introduction

An article published in the *Journal of the American Society for Information Science and Technology* in 2013 showed that the average number of publications of Norwegian scientists between 2005 and 2008 varied greatly from research field to research field (Piro et al., 2013). The authors found that the average natural scientist published more than six articles in this time span while the average humanities researcher published less than four papers. Yet these numbers changed dramatically when the authors introduced so-called fractionalized counts by dividing the total publications by the number of co-authors. This led to the average humanities researcher receiving a fractionalized count of more than three papers while the average natural scientist receiving a fractionalized count of 1.34 papers.

To investigate these changes that solely derived from the author team sizes varying across research fields, a follow-up study analyzed Flemish author teams using publication-based data (Ossenblok et al., 2014). As a first result, this study found that the average number of authors per journal article had increased strongly in the period from 2005 to 2010 compared to the period from 2000 to 2005, a common trend across academic disciplines (O'Brien, 2011). Various reasons exist for this. For example, multi-authored papers get more often cited (Valderqas et al., 2007). Lariviere et al. (2015) investigated this phenomenon by including more than 28 million papers published between 1900 and 2011. They found that the scientific impact measured by citations of multi-authored papers exceeded those of single authored papers substantially. The fact that multi-authored papers were usually more extensive, go more into details and cover more topics might explain this phenomenon (Yitzhaki, 1994). In addition to citation effects, scholars collaborating with other scholars publish also more papers than those not collaborating or those collaborating only with fewer other scholars (Lee, & Bozeman, 2005). However, Abramo et al. (2009) pointed out that the existence and strength of this relationship varies across research fields.

To investigate such field differences more closely, Ossenblok et al. (2014) specifically focused on the social sciences and humanities when assessing the size of author teams. However, they still found differences in author team sizes even within the social sciences and humanities with author teams in the social sciences being on average larger than author teams in the humanities. For example, the study showed that psychology articles contained on average

4.44 authors while the average law article only had 3.34 authors. The authors attributed their findings to different collaboration patterns as outlined in Lariviere, Gingras and Archambault (2006). This study addressed differences in research collaborations by separately investigating research field, language and country correlates. Yet the authors did not include all three factors in a multivariate model. As a case in point, Piro et al. (2013: 317) highlighted that "the reduced productivity in technology may be due to the large number of younger and less advanced researchers (PhD students) and the low share of professors and associate professors." Although this quote addressed productivity differences and not differences in author team sizes, we might very well consider that differences in geographical and job position distributions might also bias the research field effects on author team sizes highlighted in Lariviere et al. (2006) and Ossenblok et al. (2014).

In addition to the lack of multivariate assessments of author team sizes in the social sciences, all so far mentioned studies did not incorporate their findings into an underlying theory. This highlights the need for a profound follow-up study investigating author team sizes in the social sciences based on a solid theoretical framework. We therefore build upon a well-known theory and employ a multivariate investigation of the effects of research fields, geographic regions and job positions on the size of author teams in the social sciences. The research question therefore goes as follows:

Research question: *What are the research field, geographic location and job position effects on the number of authors of a social scientific paper?*

To answer the research question, we first derive hypotheses about potential author team size correlates from the existing literature. Hereby, we base our arguments on one of the most well-known theories in the social sciences: *Transaction costs.* Afterwards, we conduct an empirical study surveying social scientists from various disciplines. Based on the results of the statistical analyses, we discuss our findings putting them into context with the existing research and elaborate on the implications and future research opportunities. Last, this paper concludes with practical recommendations when handling author teams for social scientists, research institutions, journal editors and academic societies.

3.2 Theoretical background and research hypotheses

"Trade can make everyone better off" (Mankiw, 2015: 9). This principle dates way back. Already the first modern economists (e.g. Smith (1776); Ricardo (1817)) discussed the positive effects of trading. However, (market) trading is often also associated with costs (e.g. fuel for driving to the market, opportunity-costs of standing at the market and selling products, …). Coase (1937) identified this phenomenon in his groundbreaking paper *The Nature of the Firm*. Building on this work, Oliver E. Williamson (1979) defined the costs connected with a market trade as *market transaction costs*. In turn, there also exist *managerial transaction costs*. These costs are not associated with procuring a product from a market but instead producing this product in an organization (Williamson, 1979). Coase (1937) already argued that firms' very right to exist derives from the total *market transaction costs* exceeding the total *managerial transaction costs*.

With respect to research collaborations and author teams, we might find something very similar: Researchers can decide whether they want to conduct and publish a research project on their own (similar to a market transaction) or team up with colleagues to create collaboratively a publication (similar to a transaction within an organization) (Mitchell, 1961). In accordance to the elaborations above on *transaction costs* (Coase, 1937; Williamson, 1979), scholars face incentives to team up with other researchers if their overall costs for publishing individually exceed their overall costs for publishing collaboratively.

When talking about costs, it is important to keep in mind that we refer to the economic cost concept (Pitman, 2009). Consistently, we do not only need to account for the explicit costs but also for the implicit costs of a research project (Foster et al., 2007). Examples for explicit costs are money spent on experiments, buying datasets as well as research assistants' salaries and, specifically for collaborative research projects, communication resources and travel costs for meetings (Kummings, & Kiesler, 2007). Implicit costs include amongst others the time and efforts scholars spent on getting theoretical input, on collecting data, on writing the paper and on dealing with reviews. Further examples are losing publication opportunities in case a similar paper gets published earlier and, specifically for collaborative research projects, reduced merits from publications due to fractionalized publication and citation counts (Beck et al., 2020).

If scholars' toils for getting theoretical input, collecting data, writing the paper, dealing with reviews, etc., exceed their costs of collaborating with other researchers, *transaction costs* predict that they would rather look for research collaborations (Williamson, 1979). Consequently, scholars will be more likely to publish in teams if their costs of teaming up with other researchers are small and if their own costs of procuring data and materials are high. This goes in line with the historical development of multi-authorship.

Back in the 1940s, most papers only had one author. As a case in point, Smith (1958) was the first scholar to investigate the phenomenon of co-authorship in a social scientific research discipline. Looking at papers presented at the APA meetings between 1946 and 1957, he found that the share of single authored research papers dropped substantially during this period with more than 75% of all publications possessing a single author in 1946 compared to only 50% of all publications possessing a single author in 1957. Nowadays, 89% of publications in the most impactful psychological journals are multi-authored (Piocuda et al., 2015). The lightspeed discoveries and advances in information and communication technology partially explain this sharp increase in the share of multi-authored papers as telephones, computers and especially the internet have eased correspondences substantially (Hiltz, 1982; Walsh, & Maloney, 2003). In other words, the reduction in communication costs led to an increase in scientific collaborations. This endorses the application of *transaction costs* to research collaborations.

According to Williamson (1979), transaction costs (especially *market transaction costs*) increase with *uncertainty* and *specificity* and decreases with the *frequency* of the transaction. Thus, an uncertain transaction for a very specific commodity that occurs only once takes more often place inside organizations than on markets (Williamson, 1979). In the following, we assess the *uncertainty*, *specificity* and *frequency* aspects of research projects to assess their effects on researchers teaming up to collaboratively publish a paper.

3.2.1 Uncertainty

Uncertainty has always played a substantial role in scholarly careers. Today, the "tightening of systems of measuring, evaluating and managing researchers' performance as well as the fostering of competitive dynamics are central aspects of the changes in the organization of research work" (Fochler, & Sigl, 2018: 350). At the same time academic competition has increased fiercely due to the space-age advances in communication technology and the nearly worldwide adaption of English as the primary scientific language (van Raan, 2001; Tardy, 2004; Di Bitetti, & Ferreras, 2017). Furthermore, Kim (2001) argued that the transition from paper journals to e-journals has increased the number of submissions substantially. Consequently, to avoid overwhelming amounts of publications, journal acceptance rates have dropped considerably, increasing *uncertainty* for scholars (Wardle, 2012). Moreover, Aarsen et al. (2008) showed that the highest impact journals often hold the lowest acceptance rates. Lower acceptance rates reduce scholars' chances of publishing and increase the risks of spending all the efforts, resources and time on a research paper for nothing (Moizer, 2009). Therefore, acceptance rates represent a good proxy for scholars' perceived *uncertainty* as they measure the chances of publication (and in turn academic) success.

Acceptance rates are not the same across all disciplines. For example, nature and life science journals often have acceptance rates as high as 60% while social science journals' acceptance rates are sometimes even as low as 20% (Pfeffer, 1993). The same also applies to the different research fields within the social sciences. Sugimoto et al. (2013) conducted a cross-disciplinary study on acceptance rates. According to them, business journals possess on average an acceptance rate of 30.92%, computer science and operations research journals possess on average an acceptance rate of 32.27% and psychology journals possess on average an acceptance rate of 35.46%. These numbers go in line with results from discipline specific data: Krueger et al. (2012) reported an average acceptance rate of 29.7% for management journals while the American Psychological Association (2018) stated that in 2017 the overall acceptance rate for all journals published by the APA was 30%. Considering that the APA primarily publishes high impact journals which correlates with lower acceptance rates (Aarsen et al., 2008), the higher acceptance rate of 35.46% in Sugimoto et al. (2013) seems plausible. Unfortunately, there exists no other study to validate the results for computer science and operations research. In addition, Sugimoto et al. (2013) leaves out several important social sciences disciplines like sociology or political sciences. We therefore refer to other studies in the following.

Table 3.1: Overview of acceptance rates across disciplines

Discipline	Acceptance Rate 1	Acceptance Rate 2
Psychology	35.46% (Sugimoto et al., 2013)	29.7% (APA Journals, 2017)
Computer Science and Operations Research	32.27% (Sugimoto et al., 2013)	
Business	30.92% (Sugimoto et al., 2013)	29.7% (Krueger et al., 2012)
Economics and Finance	26.20% (Haensly et al., 2009)	26.20% (Cherkashin et al., 2009)
Sociology	20.32% (*American Sociological Review, 2019)*	11.35% (*Sociological Theory, 2019)*
Political Sciences	10-14% (Esarey, 2016)	11% (Mc Kee et al., 2020)

Note: Percentage numbers represent the acceptance rates. Regular brackets indicate the sources for the acceptance rates. Italic brackets indicate the journal and the year for which the acceptance rates apply.

For economic and finance journals, Haensly et al. (2009) reported an average acceptance rate of 26.20%. This perfectly aligns with Cherkashin et al. (2009) who reported an average acceptance rate of 26.20% for the *Journal of International Economics*. The acceptance rate of the *American Sociological Review* was 20.32% while *Sociological Theory* had an acceptance rate of 11.35% in 2019 (American Sociological Association, 2020). Esarey (2016) collected data from various political science journals. He showed that acceptance rates at political science journals are very low ranging between 10 to 14% (Esarey, 2016). As a case in point, *Political Research Quarterly* had in 2019 an acceptance rate of 11% (McKee et al., 2020). Table 3.1 summarizes this small literature review on acceptance rates.

Based on Table 3.1, we apply these acceptance rates to identify research fields associated with higher/lower degree of uncertainty. More specifically, we expect scholars to face higher *uncertainty* when the acceptance rates are lower. In accordance with *transaction costs*, higher *uncertainty* should increase the likeliness of research collaboration as organizations trump the market mechanism in situations with high *uncertainty* (Langlois, 1998). Hence, those fields associated with the highest acceptance rates should favor single and/or smaller author teams because there exists lower *uncertainty* while those fields associated with the lowest acceptance rates should favor author teams. The following set of hypotheses summarizes this concept.

Hypothesis 1a: *Psychologists work on average in smaller author teams.*

Hypothesis 1b: *Computer science and operations researchers work on average in smaller author teams.*

Hypothesis 1c: *Sociologists work on average in larger author teams.*

Hypothesis 1d: *Political scientists work on average in larger author teams.*

3.2.2 Specificity

Specificity addresses whether the assets are idiosyncratic and can take on various forms (Williamson, 1979). The same applies to research papers as they can also vary in nearly countless characteristics. One example would be language. While English represents the dominant language in today's science, there still exist many journals publishing research in other languages (Tardy, 2004). However, journals publishing in another language than English often have lower citation rates and thus lower impact factors (Di Bitetti, & Ferreras, 2017).

Addressing specifically research collaborations, Lariviere et al. (2006) investigated language-based differences between francophone and anglophone scholars. They showed that anglophone scholars are more likely to collaborate with other scholars because their pool of potential co-authors speaking English is substantially greater than the pool of potential co-authors speaking French. While this finding is primary applicable to anglophone researchers, also natives in languages other than English might possess an advantage if their language is spoken by many other people. Consequently, Chinese, Hindi and Spanish natives might also profit from facing less language specificity (Ethnologue, 2020).[12]

In *transaction costs* economics, languages are part of the so-called *location specificity* that increasingly has gained popularity due to the growing trends of outsourcing and gig economy (Aubert et al., 2004 Heeks, 2017; Henten, & Windekilde, 2015; Selmier, & Oh, 2012). The primary characteristic of *location specificity* is the geographical location (Anand, & Delios, 1997). This also applies to research collaborations. As a case in point, academic collaborations between scientists from different research institutions increase with the introduction of low-cost carrier non-stop flights between the locations of the research institutions (Catalini et al., 2016). Hence, we might find larger author teams in North America, Western Europe and Asia, the three leading regions for low-cost airlines (Wall, & Carey, 2017).

Furthermore, cultural norms also influence the *location specificity* (Wood, & Parr, 2005). Consequently, they might also explain why research processes and collaborations differ between countries or continents (Hochreiter, & Waldhauser, 2014; Smith et al., 1989). This could especially play a role when looking at the size of author teams for Asian researchers as their obedience towards and respect for their supervisors exceed Western standards (Salita, 2010). In fact, Asian researchers tend to give authorship to their supervisor even if they did not contribute to the research project (Yukawa et al., 2014). Based on these considerations on *location specificity* we formulate the following research hypotheses:

Hypothesis 2a: *Asian researchers work on average in larger author teams.*

Hypothesis 2b: *Anglophone researchers work on average in larger author teams.*

Hypothesis 2c: *Western European researchers work on average in larger author teams.*

3.2.3 Frequency

According to Williamson (1979), *frequency* represents the last characteristic when assessing *transaction costs*. He states that if the same or quite similar transactions occur more often, the *transaction costs* per transaction are lower than if the transaction happens only once. Applying this to research collaborations, this would imply that those scholars, who know the "publishing game" (Broad, 1981: 1137) well, those scholars who have already collaboratively published in

[12] We set the cut-off at Spanish as there are globally about 538 million Spanish speakers. The next language, French, is only spoken by 277 million people (Ethnologue, 2020).

the past and those scholars who are in frequent exchanges with their colleagues, publish more often in teams. In other words, *transaction costs* tell us that more experienced faculty members should have more often co-authors than junior faculty members.

Existing research highlights that more relationships with other scholars induce larger networks and in turn lead to more collaboration (Gersik et al., 2000). Using data from interdisciplinary research projects, Cummings and Kiesler (2008) showed that the more collaborations individuals had in past projects, the more likely they were to also collaborate in future projects. This provides proof for our *frequency* assessment. Consequently, we expect scholars who have already spent more time in academia to be more often part of larger author teams.

In addition to time spent in academia, also job position can act as a proxy for working and collaborative publishing experience. As a case in point, (full) professors publish more articles than other researchers (Long, 1978). Amongst other things, this derives from professors mentoring several doctoral students at the same time (Long, 1978). Consequently, they receive more often co-authorship than researchers who do not mentor doctoral students. As a result, professors should have a higher *frequency* of publishing multi-authored papers. Taking additionally into account that postdocs collaborate on average on about 40% of their publications with their former PhD-supervisor (Borrego et al., 2008), we find ground for suspecting also a positive relationship between professors and multi-authored publications.

Besides looking at working years in academia and job positions separately, we also need to reflect on them together. Common sense, as discussed above, would say that professors represent those individuals who have worked for a substantial amount of years in academia. Yet there exist also exceptions to this rule. As a case in point, Akerlind (2005: 30) classified several Australian researchers as "Postdoctoral research positions as a career in their own right". These individuals were scholars who had either failed or were not interested in reaching a tenured position. However, other researchers have good reasons to avoid academic relations that might prove harmful (e.g. by lowering their productivity) (Gersik et al., 2000). This might make it more difficult for non-tenured but experienced scholars to collaborate with other researchers as many researchers view not achieving tenure as a big career blow (Schwenk, 1993). Consequently, experienced professors might work together with more co-authors than experienced non-tenured faculty members. The following hypotheses capture the job position and experience related stipulations.

Hypothesis 3a: *Experienced scholars work on average in larger author teams.*

Hypothesis 3b: *Professors work on average in larger author teams.*

Hypothesis 3c: *Experienced professors work on average in larger author teams than experienced non-tenured faculty members.*

3.2 Research methods

3.2.1 Survey design and distribution

We used the online survey administration tool Qualtrics to set up the questionnaire. The first section of the survey covered demographic and job-related characteristics to enable the identification of scholars' research fields, geographical regions, job positions and experience. Subsequently, the questionnaire asked respondents to state the number of authors of their last published paper. In line with Ossenblok et al. (2014) we capped the maximum of authors by including the option "more than five authors" to allow respondents with large author teams to answer the question easily and at the same time reduce the impact of outliers. This should not shift our results dramatically as only about 3% of the social scientific publications contain more than five authors (Ossenblok et al., 2014). Nevertheless, we specifically address this assumption in the robustness checks. The questionnaire contained additional questions in the second section focusing on the contribution of the individual authors. Furthermore, there existed also a third section. Yet the last two sets of questions are not relevant for this research project.

We generated the survey in spring 2018. After its completion we sent it to friendly colleagues and asked them for feedback. We received only a minor amount of (mainly grammatical) improvement suggestions and incorporated all of them. In July 2018, we sent e-mails including a description and a link to the survey for testing purpose to 275 scholars who presented at least one paper at the 2018 European Accounting Association Annual Congress. The responses showed that the survey worked smoothly and did not need any further update. Therefore, we include these responses in the analysis.

After the conclusion of the successful survey testing, we distributed the questionnaire by sending e-mails containing a description and hyperlink to 63,240 unique e-mail addresses in late summer 2018. The e-mail addresses comprised corresponding authors of papers presented at meetings of large academic societies between January 2010 and June 2018 as well as papers published in impactful field-relevant journals between January 2010 and June 2018.[13] We received 15,573 bounce back e-mails and processed all of them manually. Following the update of our contact address inventory, we sent a reminder e-mail ten days after the first e-mail. We again manually processed bounce backs. This allowed us to conclude that 48,015 researchers received either the first or the reminder e-mail.

Overall, our survey counts 2,817 respondents implying a response rate of 5.87%. Hereby, the response rate lies in the same range as other research employing large scaled studies addressing ethics in social scientific research (Hopp, & Hoover, 2017; Liao et al., 2018). Our final sample differs from the 2,817 respondents as we only include respondents who have already published a paper and who did not skip questions employed in the study or chose at least one N/A option for such a question. Hence, we only include 2,046 respondents in our analysis.

3.2.2 Variables and Statistical Methods

We employ two dependent variables stemming from the same question. First, *Number of Authors* indicates the response to the question "How many authors does this [last published] paper possess (including you)?" As this question only contained the answer option ">5" for papers with six or more authors, we approximate this option with the value 6. Second, *Multi-Authored Paper* represents a dummy variable that takes the value 1 if *Number of Authors* exceeds one and 0 otherwise.

The explanatory variables capture *uncertainty* (as respondents' research fields), *specificity* (as respondents' location) and *frequency* (as respondents' *Academic Working Years* and job positions). For *uncertainty* we specify *Psychology* (1 if respondents conduct primarily psychological research; 0 otherwise), *IT and OR* (1 if respondents primarily conduct research on computer science or operations research; 0 otherwise), *Business* (1 if respondents primarily conduct accounting, business, management or marketing research; 0 otherwise), *Economics* (1 if respondents conduct primarily economic or financial research; 0 otherwise), *Sociology* (1 if respondents primarily conduct sociological research; 0 otherwise) and *Political Sciences* (1 if respondents primarily conduct research on politics or international affairs; 0 otherwise) with all other social scientific disciplines and general social scientists assembling the baseline category.

[13] A full list of societies and journals is available upon request from the corresponding author.

The second set of independent dummy variables identifies the *specificity*. Hereby, we differentiate between *Asia* (1 if respondents live in Asia excluding Russia; 0 otherwise), *Anglophone* (1 if respondents live in the US, Canada, United Kingdom, Ireland, Australia or New Zealand; 0 otherwise), *Continental Western Europe* (1 if respondents live in the continental European Union, Iceland, Switzerland or Norway; 0 otherwise) and *Latin America and Africa* (1 if respondents live in Latin America or Africa; 0 otherwise) with scholars from all other regions (mainly Eastern Europe) assembling the baseline category. Third, we employ *frequency* as *Academic Working Years* consisting of the responses to the question "Please indicate your working years as scientific staff member (excluding Bachelor and Master studies):". Furthermore, we assess the following job positions: *Postdoc* (1 if respondents hold a postdoc position; 0 otherwise), *Junior Professor* (1 if respondents hold a not-tenured professorship; 0 otherwise) and *Professor* (1 if respondents hold a tenured professorship; 0 otherwise) with the remaining respondents (mainly PhD students) assembling the baseline category.

Besides the explanatory variables, we also include several controls. Several studies have shown that gender might influence academic productivity and collaboration (e.g. Borsuk et al., 2009; Leahey, 2006; Reed et al., 2011; West et al., 2013). We therefore include *Female* (1 if respondent is female; 0 otherwise). Furthermore, whether papers are single- or multi-authored might also relate to productivity (Abramo et al., 2009; Lee, & Bozeman, 2005). To control for such effects we include *Published Papers* (0 = 1-2; 1 = 3-5; 2 = 6-9; 3 = 10-14; 4 = 15-19; 5 > 19 papers published in the last three years leading up to the survey) and *Written Reviews* (0 = 0; 1 = 1-2; 2 = 3-5; 3 = 4 = 10-14; 5 = > 14 journal reviews written in last year leading up to the survey). Last, we also control for *Editor* (1 if respondents are editors of at least one peer-reviewed journal; 0 otherwise) to capture the effect that scholars heavily engaged in their academic community know more colleagues and therefore might possess more opportunities for collaborating on research projects (Long, & McGinnis 1982).

Due to the nature of our dependent variables, we employ two different types of regressions in the statistical analysis. On the one hand, we use logit models to address the question of what differentiates single-authored from multi-authored papers by including *Multi-Authored Paper* as dependent variable. On the other hand, we apply standard OLS regressions to investigate the effects on *Number of Authors*.

3.3 Results

3.3.1 Descriptive statistics

Table 3.2 shows the descriptive statistics for all dichotomous variables. Overall, we find that 87.10% out of our sample of 2,046 papers possess more than one author. Regarding research field affiliation, nearly one third of the respondents identify themselves as *Business* researchers. *Economics* and *IT and OR* represent the second strongest disciplines with each constituting 15.59% of our sample. *Political Sciences*, *Sociology* and *Psychology* each comprise more than 100 observations although they only have a single digit share among all research fields. Furthermore, more than two thirds of our respondents live in *Anglophone* or *Continental Western European* countries. Scholars from *Asia* or *Latin America and Africa* form about 10% of our sample each. The majority of our respondents are *Professor*s. Additionally, about every fifth respondent possesses a *Junior Professor*ship. *Postdoc*s only constitute about 8% of our sample while the remaining share of nearly 20% consists mainly of PhD students. Concerning the controls, females represent about one third of our sample and approximately every fourth respondent possesses an editorial role at a peer-reviewed journal.

Table 3.2: Descriptive statistics for dichotomous variables

Total Respondents:	**2,046**	**100,00%**
Multi-Authored Paper	1,782	87.10%
Psychology	128	6.26%
IT and OR	319	15.59%
Economics	319	15.59%
Business	672	32.84%
Sociology	168	8.21%
Political Sciences	198	9.68%
Asia	215	10.51%
Anglophone	793	38.76%
Continental Western Europe	715	34.95%
Latin America and Africa	203	9.92%
Postdoc	170	8.31%
Junior Professor	414	20.23%
Professor	1,109	54.20%
Female	668	32.65%
Editor	530	25.90%

Table 3.3 lists the descriptive statistics for the integer variable *Number of Authors* and the two ordinal variables *Papers Published* and *Written Reviews.* We find that most papers in the social sciences either have two or three authors with each category containing more than 30% of the papers. About 15% of the papers possess four authors while papers with five or even

six or more authors exist considerably less often. For productivity, we find that many scholars published between three to five papers in the three years leading up to the survey. A substantial number of scholars seem to be less productive as they only published one or two papers in the same time horizon. Approximately a similar share of scholars seems to be more productive as they published between six to nine papers. Interestingly, our sample includes 90 very prolific scholars who have published 20 or more papers in the three years leading up to the survey which transform into at least 6 papers published each year. Table 3.3 shows that scholars write much more reviews than they publish papers. About 30% of the scholars have written between three to five reviews in the year leading up to the survey. In addition, more scholars have written fifteen or more reviews in the last year than published fifteen or more papers in the past three years.

Table 3.3: Descriptive statistics for categorical and restricted integer variables

	1	2	3	4	5	> 5	Total
Number of Authors	264 (12.90%)	627 (30.65%)	636 (31.09%)	318 (15.54%)	110 (5.38%)	91 (4.45%)	2,046
	1-2	3-5	6-9	10-14	15-19	> 19	Total
Published Papers	464 (22.68%)	793 (38.76%)	424 (20.72%)	191 (9.34%)	84 (4.11%)	90 (4.40%)	2,046
	0	1-2	3-5	6-9	10-14	> 14	Total
Written Reviews	169 (8.26%)	475 (23.22%)	610 (29.81%)	348 (17.01%)	227 (11.09%)	217 (10.61%)	2,046

Academic Working Years represents the only non-restricted integer variable. On average, the respondents have been in academia for 17.79 years with a standard deviation of 12.34 years. Our study includes 23 scholars who have not worked scientifically for more than one year. At the same time, our sample also contains an 86 years old scholar who has been in academia for 63 years.

We do not include a correlation table as many of our dummy variables are exclusive (e.g. researchers cannot possess two primary research disciplines). Instead we discuss the Variance Inflation Factors for every model in the robustness section to address potential problems that could arise from multicollinearity.

3.3.2 Regression results

Table 4 presents the results from logistic regressions with *Multi-Authored Paper* as the dependent variable. Model 1 allows assessing Hypotheses 1a, 1b, 1c and 1d by including the research fields and the control variables. We do not find support for any of our hypotheses. In fact, we find the exact reverse effects. *Multi-Authored Paper* exist highly significantly more often in

Psychology and *IT and OR*. Moreover, *Sociologists* and *Political Scientists* publish significantly less often in teams. In addition to our hypotheses, we find that *Business* researchers also collaborate significantly more often with colleagues. Yet we do not find any effects for *Economics*.

Model 2 in Table 4 includes location and the controls. We find support for Hypotheses 2a, 2b and 2c as scholars living in *Asia* as well as in *Anglophone* or *Continental Western European* countries more often collaborate with co-authors. This also applies to scholars from *Latin America and Africa* implying that single-authorship spreads especially across Eastern Europe.

Model 3 investigates possible *frequency* effects stemming from job position and experience and allows us to assess Hypotheses 3a and 3b. As the only significant effect in the explanatory variables presents that *Postdoc*s highly significantly less often work in author teams, we cannot confirm Hypotheses 3a and 3b. Model 4 contains the same variables as Model 3 but additionally includes an interaction term between *Professor* and *Academic Working Years* to specifically address Hypothesis 3c. However, the inclusion of the interaction term does not alter any significance levels and its coefficient is also not significant. We therefore do not find support for Hypotheses 3c.

The last two models in Table 4 include all variables with Model 6 additionally including the interaction term between *Professor* and *Academic Working Years*. Overall, there appears only a single adulteration in the interpretation of the results: *Sociology* does not longer correspond to a lower chance of team authorship.

Looking at the controls of Model 6, we see that scholars with more *Published Papers* in the last three years have highly significantly more often co-authors. Yet papers written by *Editor*s are marginally significantly less often multi-authored.

We also apply z-tests using the coefficients from Model 6 to test whether the strength of the significant effects differ. However, we do not find any significant differences between the positive effects of *Psychology, IT and OR* and *Business* as well as the positive effects of *Asia, Anglophone, Continental Western Europe* and *Latin America and Africa*.

Table 3.4: Effects of research field, geographical region, job position and experience on multi-authored papers

	Model 1 Multi-Authored Paper	**Model 2** Multi-Authored Paper	**Model 3** Multi-Authored Paper	**Model 4** Multi-Authored Paper	**Model 5** Multi-Authored Paper	**Model 6** Multi-Authored Paper
Psychology	1.112**				1.057*	1.059*
	(0.429)				(0.433)	(0.433)
IT and OR	0.863**				0.827**	0.831**
	(0.274)				(0.277)	(0.278)
Economics	0.249				0.228	0.232
	(0.241)				(0.244)	(0.245)
Business	1.120***				1.088***	1.094***
	(0.239)				(0.247)	(0.247)
Sociology	-0.479†				-0.317	-0.314
	(0.252)				(0.260)	(0.261)
Political Sciences	-0.755**				-0.742**	-0.736**
	(0.237)				(0.242)	(0.243)
Asia		1.236***			0.842**	0.840**
		(0.295)			(0.310)	(0.310)
Anglophone		1.261***			0.819**	0.818**
		(0.236)			(0.256)	(0.256)
Continental Western Europe		1.027***			0.830***	0.828***
		(0.232)			(0.248)	(0.248)
Latin America and Africa		1.299***			1.004**	0.996**
		(0.302)			(0.318)	(0.319)
Postdoc			-0.700**	-0.691**	-0.478†	-0.470†
			(0.249)	(0.250)	(0.263)	(0.264)
Junior Professor			-0.220	-0.215	-0.313	-0.309
			(0.215)	(0.216)	(0.227)	(0.228)
Professor			0.060	0.142	-0.099	-0.023
			(0.219)	(0.319)	(0.230)	(0.335)
Academic Working Years			-0.009	-0.005	-0.006	-0.003
			(0.007)	(0.011)	(0.007)	(0.012)
Academic Working Years X Professor				-0.005		-0.005
				(0.014)		(0.015)
Female	0.076	0.075	0.048	0.047	0.068	0.068
	(0.148)	(0.144)	(0.144)	(0.144)	(0.151)	(0.151)
Published Papers	0.292***	0.301***	0.282***	0.281***	0.301***	0.302***
	(0.073)	(0.071)	(0.070)	(0.070)	(0.074)	(0.074)
Written Reviews	0.100†	0.113†	0.133*	0.131*	0.096	0.094
	(0.059)	(0.059)	(0.059)	(0.059)	(0.064)	(0.064)
Editor	-0.373*	-0.323*	-0.400*	-0.402*	-0.325†	-0.326†
	(0.166)	(0.163)	(0.164)	(0.164)	(0.171)	(0.171)
Chi2	146.13	68.71	50.80	50.93	164.70	164.79
p > Chi2	0.000	0.000	0.000	0.000	0.000	0.000
Pseudo R2	0.093	0.044	0.032	0.032	0.105	0.105
Observations	2046	2046	2046	2046	2046	2046

† $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Note: Coefficients correspond to the marginal effects for the independent variables calculated at the mean levels of the remaining variables derived from logistic regressions with standard errors in parentheses.

Table 3.5 shows the results from OLS regressions with the *Number of Authors* as the dependent variable. For the sake of simplicity and readability, we only cover Model 6 containing all variables and the interaction term in the text. Like before, papers published in *Psychology* and in *IT and OR* possess highly significant more authors and papers published in *Political Sciences* possess highly significantly less authors, thus again we find evidence contradicting Hypotheses 1a, 1b and 1d. In addition, researchers in *Sociology* have highly significantly less co-authors. Contrary to the findings above, *Business* scholars do not seem to differ in author team sizes from the remaining and not specifically included social sciences while author team sizes in *Economics* are significantly smaller. In turn, the effects of *Asia, Anglophone, Continental Western Europe* and *Latin America and Africa* mirror the findings from above with scientists living in those regions collaborating highly significantly with more co-authors. Regarding *frequency* effects on the *Number of Authors*, we do not find that any differences exist for any job position (including *Postdoc*s). In the controls, we find, similar to Table 4 that scholars with more *Published Papers* in the last three years collaborate with more co-authors. Yet there exists no significant effect anymore for *Editor*s.

We also compare the strength of the significant effects pointing into the same direction by applying z-tests on the coefficients in Model 6. We find that *Psychology* papers contain highly significantly more authors than *IT and OR* papers, thus making it the field with highest number of co-authors. Researchers in *Sociology* collaborate with significantly less co-authors than researchers in *Economics.* Researchers in *Political Sciences* collaborate highly significantly with less co-authors than researchers in *Sociology* and researchers in *Economics*, thus making it the field with the lowest number of co-authors. Concerning regional effects, scholars living in *Anglophone* countries share their papers with marginally significantly more co-authors than scholars living in *Asia* or in *Latin America and Africa.*

Table 3.5: Effects of research field, geographical region, job position and experience on the number of authors

	Model 1 Number of Authors	**Model 2** Number of Authors	**Model 3** Number of Authors	**Model 4** Number of Authors	**Model 5** Number of Authors	**Model 6** Number of Authors
Psychology	0.830***				0.772***	0.776***
	(0.130)				(0.130)	(0.130)
IT and OR	0.294**				0.299**	0.304**
	(0.101)				(0.101)	(0.101)
Economics	-0.205*				-0.214*	-0.210*
	(0.102)				(0.101)	(0.101)
Business	0.137				0.089	0.095
	(0.090)				(0.091)	(0.091)
Sociology	-0.505***				-0.445***	-0.442***
	(0.119)				(0.120)	(0.120)
Political Sciences	-0.700***				-0.719***	-0.712***
	(0.114)				(0.114)	(0.114)
Asia		0.608***			0.390**	0.387**
		(0.140)			(0.136)	(0.136)
Anglophone		0.737***			0.568***	0.566***
		(0.121)			(0.120)	(0.120)
Continental Western		0.623***			0.491***	0.488***
Europe		(0.122)			(0.119)	(0.119)
Latin America and Africa		0.577***			0.389**	0.381**
		(0.142)			(0.139)	(0.139)
Postdoc			-0.308**	-0.296*	-0.172	-0.162
			(0.116)	(0.117)	(0.112)	(0.113)
Junior Professor			-0.124	-0.117	-0.079	-0.074
			(0.091)	(0.091)	(0.088)	(0.088)
Professor			-0.183*	-0.078	-0.168†	-0.078
			(0.089)	(0.128)	(0.086)	(0.123)
Academic Working Years			-0.006*	-0.001	-0.007*	-0.003
			(0.003)	(0.005)	(0.003)	(0.005)
Academic Working Years				-0.007		-0.006
X Professor				(0.006)		(0.006)
Female	0.101†	0.141*	0.112†	0.111†	0.075	0.075
	(0.057)	(0.058)	(0.059)	(0.059)	(0.057)	(0.057)
Published Papers	0.120***	0.158***	0.158***	0.158***	0.137***	0.137***
	(0.024)	(0.025)	(0.025)	(0.025)	(0.024)	(0.024)
Written Reviews	-0.018	-0.030	0.005	0.002	-0.018	-0.021
	(0.022)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
Editor	-0.047	-0.013	-0.007	-0.007	0.012	0.012
	(0.064)	(0.066)	(0.067)	(0.067)	(0.064)	(0.064)
F	24.93	10.65	8.00	7.26	16.37	15.56
p > F	0.000	0.000	0.000	0.000	0.000	0.000
R2	0.109	0.040	0.030	0.031	0.127	0.127
Observations	2046	2046	2046	2046	2046	2046

† $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Note: Coefficients derived from OLS regressions with standard errors in parentheses.

3.3.3 Robustness of results

Assessing the robustness of the results in Table 3.5, we must consider that we approximated the statement that a paper possesses six or more authors by setting *Number of Authors* to six. To test whether we would receive the same results if we exclude those cases where we do not know the exact number of authors, Table 3.6 depicts the same OLS regressions but with the dependent variable *Number of Authors < 6* that only possesses non-missing values for papers with one to five authors. Two differences in significance levels arise from this. *Economics* paper possess no longer significantly less authors. Instead, *Business* papers contain significantly more authors. Hence, we must be careful with the implications for those two fields as the findings are somewhat ambivalent.

Besides investigating our assumption of capping the *Number of Authors* at six we conduct a series of additional robustness checks. First, we run all models with robust standard errors. This does not alter any implication. Second, we calculate variance inflation factors (VIFs) to identify potential problems arising from multicollinearity. Only the VIFs for *Anglophone, Written Reviews, Professor, Experience* and the interaction term exceed the "conservative threshold of 5" (Alauddin, & Nghiemb, 2010: 351). However, only the interaction term (VIF=10.75 as highest value) exceeds the threshold of 10 with all other VIFs lying below six (Alauddin and Nghiemb, 2010; O'Brien, 2007). As the interaction term is the product of *Professor* times *Experience*, we should expect it to correlate highly with those variables. O'Brien (2007) argues that in such cases any VIF below 40 can be sufficient. Third, we run all models excluding the control variables. This does not alter any implication. Last, we run separate models for each research field, geographical area and job position. Again, our results do not change.

Table 3.6: Effects of research field, geographical region, job position and experience on the number of authors for papers with less than six authors

	Model 1 Number of Authors < 6	**Model 2** Number of Authors < 6	**Model 3** Number of Authors < 6	**Model 4** Number of Authors < 6	**Model 5** Number of Authors < 6	**Model 6** Number of Authors < 6
Psychology	0.712***				0.667***	0.670***
	(0.118)				(0.117)	(0.117)
IT and OR	0.312***				0.309***	0.314***
	(0.089)				(0.089)	(0.089)
Economics	-0.123				-0.135	-0.131
	(0.089)				(0.088)	(0.088)
Business	0.214**				0.170*	0.175*
	(0.079)				(0.080)	(0.080)
Sociology	-0.377***				-0.308**	-0.305**
	(0.104)				(0.104)	(0.104)
Political Sciences	-0.668***				-0.682***	-0.676***
	(0.100)				(0.100)	(0.100)
Asia		0.628***			0.421***	0.418***
		(0.123)			(0.119)	(0.119)
Anglophone		0.734***			0.571***	0.568***
		(0.107)			(0.105)	(0.105)
Continental Western Europe		0.622***			0.503***	0.500***
		(0.107)			(0.104)	(0.104)
Latin America and Africa		0.654***			0.468***	0.461***
		(0.124)			(0.121)	(0.121)
Postdoc			-0.277**	-0.265**	-0.157	-0.147
			(0.102)	(0.103)	(0.098)	(0.098)
Junior Professor			-0.083	-0.077	-0.072	-0.067
			(0.080)	(0.080)	(0.077)	(0.077)
Professor			-0.106	-0.003	-0.128†	-0.038
			(0.079)	(0.113)	(0.075)	(0.107)
Academic Working Years			-0.007**	-0.002	-0.007**	-0.003
			(0.002)	(0.004)	(0.002)	(0.004)
Academic Working Years X Professor				-0.007		-0.006
				(0.005)		(0.005)
Female	0.036	0.065	0.031	0.031	0.014	0.014
	(0.050)	(0.052)	(0.052)	(0.052)	(0.050)	(0.050)
Published Papers	0.082***	0.108***	0.108***	0.107***	0.098***	0.098***
	(0.021)	(0.022)	(0.022)	(0.022)	(0.021)	(0.021)
Written Reviews	-0.007	-0.011	0.016	0.013	-0.007	-0.009
	(0.019)	(0.020)	(0.020)	(0.021)	(0.020)	(0.020)
Editor	-0.066	-0.031	-0.028	-0.028	-0.009	-0.009
	(0.056)	(0.058)	(0.059)	(0.059)	(0.056)	(0.056)
F	24.73	9.46	5.89	5.42	16.92	16.10
p > F	0.000	0.000	0.000	0.000	0.000	0.000
R2	0.113	0.037	0.024	0.024	0.136	0.137
Observations	1955	1955	1955	1955	1955	1955

† $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Note: Coefficients derived from OLS regressions with standard errors in parentheses.

3.4 Discussion

3.4.1 Summary of Results

In general, we find support for Hypotheses 2a, 2b and 2c (*location specificity)*. Yet we find the exact inverse effect than stipulated in Hypotheses 1a, 1b, 1c and 1d (*uncertainty)*. Additionally, we cannot confirm Hypotheses 3a, 3b and 3c *(frequency)*. Several implications arise from these results.

We show that the share of multi-authored papers has substantially increased compared to previous research (Ossenblok et al., 2014). Our findings indicate that research fields in the social sciences differ in the ratio between single-authored and multi-authored papers as well as in the average number of authors per paper. This does not represent a novel result as Ossenblok et al. (2014) as well as Henriksen (2016) already highlighted the existence of such differences. Yet we find that these research field differences uphold even when we include additional demographic and job-related characteristics. However, we cannot confirm that researchers facing increased *uncertainty* (as measured in journal acceptance rates) more often collaborate with co-authors. Instead, we find that psychologists, who have comparatively high acceptance, more often publish multi-authored papers than researchers from many other social scientific disciplines. We even show that papers in psychology possess the highest number of authors per paper in the social sciences. Moving beyond *transaction costs,* the following considerations might explain this result. Many psychological research projects include expensive experiments (Martin, 2008). As those experiments require substantial amounts of funding, there exists a tendency in psychology to also include the person financing the project (e.g. the faculty dean) as the last author (Duch et al., 2012; Tichy, 1997). This obviously induces a higher likeliness for psychological papers to have more authors. Moreover, conducting profound experimental studies usually requires solid knowledge to build a valid theoretical basis and experience in the actual conduction of experiments (Lych, 1966). Consequently, experts in psychological theory have to team up with experts conducting the experiments to write good research papers (Martin, 2008).

Second, information technologists and operations researchers publish more multi-authored papers and collaborate with more co-authors. This contradicts Hypothesis 1b. Yet Serenko et al. (2010) discussed that especially knowledge management and information technology are fast developing and changing fields. As such, scholars in these fields face high *uncertainty* despite their higher acceptance rates compared to other social science disciplines because of the quickly changing environment. This might explain the positive effect of *IT and OR* on multi-authored papers and the number of authors per paper.

Third, economists seem to mirror the average social scientists when addressing the share of single-authored papers. Yet we show that they work on average in smaller author teams. This might derive from the special author order economists employ: While most social sciences sort authors by contribution with the individual contributing most receiving the first author position, most economic journal list authors strictly alphabetically (Laband, 2002). Hence, including other authors raises the risk for the main contributor(s) to fall under "et al." in the references. As falling under "et al." often reduces publication and citation credits, scholars publishing in disciplines that order authors alphabetically face a strong incentive to include as few authors as possible (Laband, 2002). This gives a valid reason why economists publish in smaller teams compared to psychologists, information technologists and operations researchers as well as business researchers and other general social scientists.

Fourth, we point out that business researchers publish more multi-authored papers than other social scientists. Yet these papers might not possess more authors. However, the robustness checks show that business researchers collaborate with more co-authors when we exclude papers with six or more authors. This implies that large author teams with more than five authors are not common in business research. Nevertheless, papers in business research possess more often two to five authors than papers published in other social science disciplines (excluding psychology as well as information technology and operations research).

Fifth, our results show that sociologists do not differ robustly from the remaining not specifically included social sciences in their ratio of multi-authored to single-authored papers. This result again contradicts our *uncertainty*-based hypothesis. However, our finding is in line with Thanuskodi (2010) as well as Ossenblok et al. (2014) who show, without controlling for demographic, job-related and other factors, that the share of multi-authored papers in sociology corresponds to the average share of multi-authored papers across all social sciences.

Last, we find that political scientists publish much less multi-authored papers and possess on average much less co-authors than any other group of social scientists. We originally expected the exact opposite result based on *transaction costs*. However, Henriksen (2016: 464) stated that "Political Science have similarities with research fields in the humanities." In turn, publications in the humanities possess a much lower share of multi-authored papers (Piro et al., 2013). This derives from the circumstance that the dominating research approaches in the humanities, theoretical research and literature analysis, require a more comprehensive view of the whole topic and therefore allow often only few collaborations (Katz, & Martin, 1997; Tymoczko, 2001). Besides this explanation already covered in the existing literature, the application of alphabetical author orders by many political science journals might further induce tendencies to single-authored publications for the same reasons as discussed above with economics (Lake, 2010).

The second set of our hypotheses addresses the geographical regions. We find that researchers from all other regions publish more multi-authored papers than researchers from Eastern Europe. Furthermore, papers published by researchers from all other regions contain on average more authors than papers published by researchers from Eastern Europe. Several explanations based on *location specificity* exist for these findings. First, travelling between Eastern European cities requires usually more time and efforts due to the lack of high-speed railways, less well-developed highway networks and a low market share of low-cost carriers (Milan 1997; Stelder, 2016; Wall, & Carey, 2017). Second, especially elderly Eastern European citizens often lack proficient English skills due to them growing up in Russian dominated spheres (Kryuchkova, 2001). Yet the average level of English skills in Eastern Europe is not worse than in Latin America, where more co-authorship exists (EF Education GmbH, 2019). However, except for Brazil which is quite a large country by itself, most Latin American possess Spanish as mother tongue. Hence, there exists no language barrier for collaborating with colleagues from countries nearby for Latin American scholars while scholars from Eastern Europe face language barriers very well as many of the smaller Eastern European countries have their own languages (Marshakova-Shaikevich, 2006). Third, collaboration might not always pay off for Eastern European scholars as their papers do not receive a citation boost by having more co-authors on it (Glänzel, 2001). Unfortunately, Glänzel (2001) did not explain potential reasons for this phenomenon. Therefore, Nagy (2016) investigated Eastern European research collaborations more closely. She found that Eastern European scholars tended to collaborate more with scholars from Western Europe and North America than with scholars from Eastern Europe. As such collaborations suffer from long travel distances and time zone differences, this might explain why Eastern European scholars less often publish multi-authored papers and work together with fewer co-authors.

Looking at the specific number of authors per paper, we find that Anglophone scholars might have on average more co-authors than scholars from other regions. With English as the dominant language in nowadays science and a great network of (low-cost) flights in North America and high-speed train connections in Great Britain, this finding is not surprising (Graham, & Melo, 2011; Wall, & Carey, 2017). Nevertheless, we need to ask why Asian scholars, who have access to low-cost flights and in certain regions to high-speed trains too, might collaborate with less co-authors although their cultural norms induce them to include even their supervisors as co-authors (Yukawa et al., 2014). Salita (2010) highlighted that sometimes even not the researchers conducting the experiments and writing the paper receive authorship but just the chairholder and/or the department head (Salita, 2010). Such a practice shrinks author team sizes and hence provides a plausible explanation why Asian publications possess on average less authors than Anglophone publications.

The last three hypotheses address the *frequency* by including experience and job positions. However, we do not find any (consistent) effects for professors nor experience. As for job position, our baseline category was PhD students. As they often publish together with their supervisors (who usually are professors), this explains why the share of multi-authored papers does not differ much between those two groups (Long, 1978). However, our results indicate that postdocs publish more single-authored papers. We can think of two possible explanations for this finding. On the one hand, scholars who fail in achieving tenure might take on postdoc positions afterwards (Akerlind, 2005). These researchers might find it harder to find other researchers who are willing to collaborate with them as Gersik et al. (2000) showed that scholars try to avoid harmful academic relations. On the other hand, junior researchers looking for a tenure-track position also take up a substantial share of the postdoc positions (Akerlind, 2005). These scholars might need to polish their CVs and publication counts in order to find a decent job and therefore might tend to publish more papers on their own as this gives them more credit. As a case in point, Rotgheb and Burger (2009) discuss that single-authored publications are essential in the academic job market.

Last, we also discuss the results from the controls. While gender as well as review writing do not relate to the size of the author team, editors might more often publish single authored papers. This could derive from the fact that many editors write and publish editorials as well as cover letters on their own (Rousseau, 2009). Furthermore, we show that scholars who published more papers in the last three years publish more often multi-authored papers and work in larger author teams. This goes in line with the findings of Lee and Bozeman (2005) who showed that academic collaboration measured by co-authorship correlates with productivity. Going back to *transaction costs,* these findings confirm the effect of *frequency* on research collaborations: Those who published more in the past had higher chances to collaborate with others and therefore chose also to collaborate with others on the research project in question.

3.4.2 Impact on Academia

This study possesses wide-ranging implications on the academic community as it represents the first interdisciplinary and international multivariate analysis of author team sizes in the social sciences. First, based on our findings we call for the application of advanced techniques when assessing academic productivity instead of the currently widespread practices of simply counting citations and/or (first-authorship) publications (Carpenter et al., 2014). Especially when assessing job or funding applications from different research fields and different geographical regions, one should keep in mind that scholars' differences in productivity might not only derive from individual capabilities but also include discipline specific norms as well as infrastructural and cultural influences.

Second, we show that interdisciplinary research as well as research conducted across geographical regions might suffer from different perceptions of author team sizes and authorship even within the social sciences. Hence, we would advise scientists engaging in such projects to set authorship parameter as early as possible to avoid possible conflicts in later stages.

Third, our results indicate that there exists a link between author team size and productivity. Based on this we recommend scholars to build academic networks as early as possible to foster relationships that prove not only helpful in the job searching phase but even during regular academic life. Hereby, we also highlight the importance of academic conferences as they provide great opportunities to network and find potential co-authors. Therefore, research institutions should support scholars of all ages in attending such meetings by giving them time off and paying for travel costs. In addition, we believe that these physical meetings should also persist in future and should not be fully substituted by the emerging trend of online academic conferences (Thatcher, 2006).

Last, we highlight that the share of single-authored papers continues to drop. This results from more and more scientists experiencing the benefits of research collaborations (Piocuda et al., 2015). Consequenlty, we call upon universities and other research institutions to forfeit single-authorship requirements in search and tenure procedures. In turn, they should view multi-authored papers as a signal that shows that scholars work well in teams, have an established network, know their strength and their weaknesses and team up with others who level those.

3.4.3 Limitations and Future Research

Although we try to address as many aspects influencing the generalizability and robustness of this study as possible, one can never overcome all disputable issues. Hence, as every study also this study needs to be viewed within its setting. We rely on self-reporting data from an online survey. On the one hand, this enables us to include more demographic and job-related variables than if we use data from journal websites or databases like the Web of Science. On the other hand, several authors discuss that socially desirable responding might bias results of surveys asking sensitive questions (Krumpal, 2013). The awarding of authorship as well as the number of co-authors represent such sensitive questions as they provide the basis for publications and citation counts, the top two measures in the current academic incentive system (Altbach, 2015). In addition, the respondents self-selected them into the sample by voluntarily responding to the survey. This might imply that our sample might overrepresent scholars interested in authorship issues and other questionable research practices. Nevertheless, socially desirable responding and self-selection only bias our results in the direction of social scientists' perceived optimum. So even if our study did not perfectly assess the actual size of author teams, we would at least assess social scientists' desirable size of author teams.

Katz and Martin (1995: 1) showed that "co-authorship is no more than a partial indicator of collaboration." This is the reason why we cannot directly interfere from our findings of the size of author teams on actual research collaboration processes. Research collaboration goes beyond co-authorship as also reviewers, conference attendees as well as mentors and assistants might provide fruitful contributions to a research project (Katz, & Martin, 1995). This might also explain partially why especially the research field findings deviate from the hypotheses. Collaboration patterns differ across disciplines (Zheng et al., 2016). The same applies also to authorship patterns (Pruschak, & Hopp, 2019). Consequently, contributing something to a research project might qualify an individual for authorship in one research field but not in another. Thus, our authorship-based measure might also cover discipline specific authorship assignment effects. However, for assessing academic productivity, mainly authorship counts. Hence, our findings provide a good starting ground for assessing effects on and correlates of research collaborations.

Another limitation of our study derives from capping the variable *Number of Authors* at a maximum of six. While more than 95% of our respondents stated that their last published paper had not more than five authors, our study does not assess the mechanics and correlates for large author teams. With Open Science collaborations on the rise also in the social sciences, this might become a fruitful area for future research (Dutton, & Jeffreys, 2010).

Future research could also aim at replicating our findings using data from the Web of Sciences, Google Scholar and/or institutional websites. By including multiple papers per scholars, one could employ fixed-effects regressions that would at least partially substitute the control variables like editorial position(s) or number of reviews written. This would allow a substantial increase in coverage inducing a much higher sample size. Especially for including more scholars from second and third world countries, this could prove worthy.

Of course, one might also conduct a similar but translated survey among researchers from specific regions. It would be interesting to see whether the inclusion of more Latin American but also Chinese, Indian and Japanese researchers would confirm our results. To overcome issues arising from socially desirable responding, future research could include item-sum techniques that allows respondents to completely anonymously respond to sensitive questions (Trappmann et al., 2013). Additionally, forthcoming surveys might not only cover co-authorship but also other forms of research like contributors, reviewers and other feedback givers.

Last, we believe that investigating potential connections between differences in academic incentives and the size of author teams might reveal some underlying reasons for our findings that go beyond *transaction costs*. This is of special importance as assessment tools for scholars' productivity vary across research institutions, countries and disciplines (Carayol, 2004; Ossenblok et al., 2012). For example, if some research fields employ whole counts as publication and/or citation counts, papers published in those fields might contain more co-authors than in research fields employing fractionalized counts. Based on this, future research projects might then address the issue of which incentives work best and give meaningful implications to those in charge of hiring and tenure processes.

3.4.4 Conclusion

The aim of this research project was to clarify whether findings from existing literature on the correlates of the size of author teams withstand the application of transaction costs and multivariate analyses. Using data from a large scaled worldwide and interdisciplinary survey among social scientists, we show that research field differences persist. For geographical region, we find that Eastern European scholars publish in smaller author teams and for job positions, we find that postdocs tend to have more single-authored publications. With Eastern European scholars facing language barriers and infrastructural challenges and the job search of postdocs incentivizing them to publish single-authored papers, we derive fruitful improvement suggestions from our findings. First, those in charge of search and tenure procedures as well as funding grants need to be aware that productivity measures vary across research fields and geographical regions especially when comparing applicants with different backgrounds. Second, scholars in interdisciplinary research projects should agree on authorship as early as possible to avoid conflicts. Third, we highlight that academic productivity correlates with larger author teams. We therefore call upon research institutions to stop requiring a certain number of single-authored publications in application and tenure procedures because multi-authored publications work as a signal for higher productivity, teamwork ability and good networking skills. To foster such academic networks and to boost productivity, participation in academic conferences is essential. Consequently, such (physical) scientific meetings enhance academia substantially and should not be completely replaced by online gatherings.

4 Essay 3 – Replicating the effect of high school leadership on later life earnings

4.1 Introduction

In a 2005 article, Kuhn and Weinberger (2005: 396) asked a short yet impactful question: “Is there such a thing as leadership skill?”.[14] Their research studies the value of an unobserved bundle of leadership skills that correlates with the granting of leadership opportunities in high school. To answer the research question empirically, Kuhn and Weinberger (2005) conducted a variety of tests to determine whether leadership is “a distinct skill that is related to the management of people” (Kuhn, & Weinberger, 2005: 397). Using several behavioral measures of leadership positions granted to individuals during high school, they found that those individuals who were club presidents and team captains earned significantly more eleven years later. Given the findings, the study has been quite influential. It showed that leadership skills (including those that might predate high school and/or those that might be developed through leadership positions in high school) are financially rewarded in subsequent working life, and, importantly, are distinct from analytical and cognitive skills. Not surprisingly, Kuhn and Weinberger (2005) is among the top 100 published articles in the Journal of Labor Economics, with 405 citations (September 2020), according to Google Scholar. This amounts to an average of around 25 citations per year.

The study of processes that develop leaders is among the core research fields in the leadership literature (Day et al., 2014). Many studies have shown that for hiring decisions employers focus more on soft skills, including presentation skills, social communication, and teamwork, than on grades achieved in high school or college (Andrews, & Higson, 2008; Palmer et al., 2001; Robles, 2012; Weinberger, 2014a). In addition, more and more companies are sending their employees to leadership training seminars and workshops (Kaiser, & DeVries, 2000; Manolis et al., 2009; Breuer, & Kampkötter, 2013). Having occupied leadership roles during high school provides an advantage in the admission processes of top colleges in the United States (Morse, 2000). This implies that the possession of leadership skills might also provide advantages in the labor market.

[14] I wish to thank Christian Hopp for the collaboration and his inputs on this paper.

It is important to note that Kuhn and Weinberger (2005) only noted positive correlational evidence between being granted leadership opportunities in high school and later working life earnings but refrained from making causal inferences. This raises the question as to when and for whom these leadership skills emerge. While they point out "at least some component of leadership skill is fostered by occupying leadership positions during high school," they also acknowledge that their analysis cannot differentiate between leadership skills developed earlier and those developed in high school (Kuhn, & Weinberger, 2005: 430). The relationship between high school leadership positions and subsequent earnings captures the value of both, those characteristics that predate high school and those that are developed because of leadership activity participation in high school. Under the right circumstances, correlation can mean causation in non-experimental studies (Niles, 1922). Yet one needs to overcome a major problem in empirically analyzing non-experimental data: endogeneity, where "the effect of x on y cannot be interpreted because it includes omitted causes" (Antonakis et al., 2010: 1087).

Endogeneity problems arise when variables included in the estimation specification (e.g. the leadership position in high school) are not exogenous but, for a variety of reasons, correlate with the error term of the estimated model (the residuals of the subsequent earnings regression). To make inferences on the relationship between earnings and the leadership skills that are developed through leadership positions in high school, it is important to ensure that the estimation is consistent and that the sample coefficient converges to the true population mean (Antonakis et al., 2010). The treatment group and the control group need to be interchangeable so that each can provide the counterfactual, what would have happened to the other group (Lechner, 2002; Caliendo, & Kopeinig, 2008). This is complicated in the presence of endogeneity (Hamilton, & Nickerson, 2003; Bascle, 2008). Omitted selection – in this case, a non-modelled or non-included variable related to leadership positions in high school and subsequent income – invites the risk of comparing a non-equivalent control group to the treatment group. Similarly, the treatment and control groups might be different because of other omitted causes, such as omitted variables, simultaneity, common-method variance, or measurement error (Antonakis et al., 2010; Antonakis et al., 2014).

We therefore replicated and extended the research in Kuhn and Weinberger (2005). Our work uses quasi-experimental methods (propensity score matching and instrumental variable regression) that statistically randomize the granting of leadership opportunities, and, after ensuring that a certain set of assumptions are met, allow for making causal claims (Rouse, 2012;

Andersen, & Lu, 2016). Whereas prior work using similar models focused on the effects of leadership experiences in high school on the probability to attend college (Rouse, 2012), we follow Kuhn and Weinberger (2005) in employing later life wages as the dependent variable.

In our view, using models to distinguish between the component of leadership skill that is developed because of the granting of leadership opportunities in high school and the component that predates the granting of the leadership increases the explanatory power of the original study's findings. Against this background, we also reflect on the representativeness of the underlying original study to justify practical and policy recommendations. When making recommendations based on the correlational evidence in Kuhn and Weinberger (2005), it is important to bear in mind that the study focusses on white males only. Notwithstanding the implications the study has for white male leadership, we believe it to be of equal importance to establish the link between leadership roles and earnings among females, especially in light of prior work on the role of female leadership styles in explaining the gender wage gap (Cohen, & Huffman, 2007) and the likely role incongruity that female leaders face (Eagly, & Karau, 2002). Additionally, given the sole focus on white males, it also is important to extend the findings to non-white individuals to test for the generalizability of the leadership role-earnings effect across ethnicity (Ospina, & Foldy, 2009). Especially because Weinberger (2014b) showed that the earnings effects of high school activities for black males differ from the earnings effects found for white males, we believe that an extension to this sample provides interesting insights on the effect of leadership skills on income across race.

Furthermore, Kuhn and Weinberger (2005) base their analysis on income data up to eleven years after high school graduation. Expanding the time horizon could help to distinguish between the signaling value of leadership skills upon job market entry and long-term career and income influences. Therefore, the interpretation of Kuhn and Weinberger's (2005) results would benefit from including income data from the 2011-2012 Project Study covering later life stages.

We do not aim at disconfirming the association initially discovered by Kuhn and Weinberger (2005). Instead we want to increase the understanding of the conceptual relationship between what makes individuals become leaders in early life, and how in turn early life leadership positions affect subsequent occupational income.

4.2 Summary and contribution of the original article

Kuhn and Weinberger's (2005) article "Leadership Skills and Wages" pointed out the importance of leadership skills. The article empirically showed a positive correlative relationship between leadership positions in high school and individual income in later working life, as well as the probability of obtaining a managerial position (Kuhn, & Weinberger, 2005).

Kuhn and Weinberger (2005) did not include a specific research hypothesis, but asked a simple yet impactful research question: "Is there such a thing as leadership skill?" (Kuhn, & Weinberger, 2005: 396). The study empirically analyzes whether the holding of a leadership position, as president of a club or captain of a team, in high school, is related to increased remuneration in later life. Practical recommendations on interventions depend on whether the relationship is causal: if leadership experience pays, and positively affects career advancement, then the development of leadership skills should be fostered as early as high school to groom potential future managers.

In their study, Kuhn and Weinberger (2005) expected the groups of club presidents and team captains to possess a higher share of leadership skills than the group of those who were not leaders during high school. The behavioral measure of leadership skills includes both characteristics that predate high school and those that are developed through leadership in high school.

While existing research showed that active participation in athletic teams leads, on average, to higher earnings in later working life (Barron et al., 2000), the study by Kuhn and Weinberger (2005) included both athletic teams and non-sport-related high school clubs and their leaders. The study focused on white males only, to investigate leadership consequences "without the confounding effects of race or gender discrimination or of the changing roles and expectations of women during this time period" (Kuhn, & Weinberger, 2005: 399).

Kuhn and Weinberger (2005) used data from three longitudinal studies: Project TALENT from 1960, the National Longitudinal Study of the Class of 1972 (NLS72), and the High School and Beyond survey (HSB 1982). All of the studies interviewed high school students and followed the respondents in up to three additional waves that took place up to 11 years after the first interview. Overall, their sample sizes were 24,041 for data gathered from Project TALENT, 3,083 for data gathered from the NLS72 and 2,383 for data gathered from the HSB 1982. While Project TALENT provided information about both hourly and annual earnings, the

HSB1982 only provided information about annual earnings, and the NLS72 only provided information about hourly earnings. Therefore, the original paper used two different dependent variables. Analyses conducted with data from Project TALENT used the log of hourly and the log of annual earnings as the dependent variable. Analyses conducted with data from the HSB 1982 and the NLS72 used the log of annual earnings or the log of hourly earnings as the dependent variable, respectively. Kuhn and Weinberger (2005) employed a dummy variable indicating whether individuals had held a leadership role as a team captain and/or club president during the past three years leading up to the interview date as a measure for leadership positions in high school for the Project TALENT data. Analyses based on data from the HSB 1982 and the NLS72 included a dummy variable indicating whether individuals were team captains and/or club presidents in the last year leading up to the interview date.

In their main empirical analyses, they controlled for school (or state) fixed effects, a small number of control variables including math scores, parents' education, and the respondents' education after high school. Subsequently, the authors expanded the analysis using self-assessed psychological characteristics and weight and height measures. The psychological characteristics included "sociability," "tidiness" (the degree to which respondents are neat and organized), "vigor" (the degree to which respondents are physically energetic), "self-confidence", and "mature personality" (the degree to which respondents are hardworking and reliable).

The OLS regressions from the original article showed that being engaged in a leadership position during high school positively relates to 4% to 33% higher subsequent earnings in later working life, depending on the dataset used and the time horizon applied. The wage premium was significantly higher for those white men who had been both team captains and club presidents than for those who were either team captains or presidents only. After including cognitive skills, physical attractiveness, and psychological characteristics available in the Project TALENT data, Kuhn and Weinberger (2005) still found significant differences with respect to an effect of high school leadership positions on earnings. The authors also investigated whether team captains and/or club presidents were more likely to hold a managerial position at the time of the last Project TALENT follow-up interview, 11 years after the initial interview. The analysis showed that a significantly higher percentage of interviewees who had been in a leadership role in high school were in a managerial position at their subsequent job compared to those who had not been in a leadership role in high school.

Furthermore, Kuhn and Weinberger (2005) also separated the effect of the exposure to leadership opportunities in high school from students' personal characteristics. Hereby, they included the number of leadership positions that were available per high school as an independent variable. Their results indicated that at least a portion of the leadership-earnings relationship derived from being granted leadership opportunities in high school. The original article also showed that this relationship becomes stronger for students who assess themselves as good leaders. This essentially suggests that leadership opportunities are relevant especially for those "who had previously demonstrated a propensity to leadership" (Kuhn, & Weinberger, 2005: 429).

4.3 High-school leadership effect and causality

In summary, Kuhn and Weinberger (2005: 395) showed "that leadership skills may be fostered by exposure to high school leadership opportunities". Their analysis augmented traditional human capital models, by showing that there exist correlates to earnings that go beyond human capital and work experience, that were ubiquitously studied. Yet it is important to note that the original study only stipulated correlative evidence, and that the leadership skills measure employed comprises both a component of leadership skills that predate high school and a component that is developed because of leadership in high school. In this respect, the original study did not aim to assess causality. Hence, it did not include tests for observable and unobservable confounders that lead to selection into leadership opportunities nor did it aim to delineate the extent to which leadership experience causes the development of leadership skills that subsequently lead to higher earnings.

In the following, we therefore revisit the question whether leadership positions in high school have an effect on later life earnings of individuals, conditional on the leadership skills that predate the attainment of a leadership position in high school.

We are specifically interested in contrasting the earnings of individuals who were granted leadership opportunities to the counterfactual, the earnings for those who were not granted such opportunities. In non-experimental data, however, the groups are unlikely to be interchangeable due to observable and unobservable causes that determine whether certain individuals are granted leadership opportunities (Antonakis et al., 2010). Leadership opportunities are potentially endogenous. Observable omitted selection and other unobservable omitted causes invite the risk of comparing non-equivalent control groups to the treatment group. Consequently, it is difficult to distinguish the component of leadership skills that is developed because leadership opportunities were granted in high school from the component that predates the granting of leadership opportunities.

4.3.1 Endogeneity as a result of observable omitted selection

Observable omitted selection in model explanatory variables has been widely discussed as invalidating coefficient estimates and hypothesis tests (Antonakis et al., 2010). Antonakis, et al. (2010: 1091) emphasize that observable omitted selection may cause inconsistent estimates when assessing the causal effect of high school leadership experience on earnings when "Comparing entities that are grouped nominally where selection to group is endogenous (e.g., comparing men and women leaders on leadership effectiveness where the selection process to leadership is not equivalent)". In the presence of a correlation between the error term of the earnings regression and the leadership dummy regressors, the estimates from an OLS regression might not reflect the true causal effect of leadership experience on earnings; they are, statistically speaking, biased. These differences in estimates might derive from different characteristics among those who were granted leadership opportunities and those who were not (Hamilton, & Nickerson, 2003; Shaver, 1998). Consequently, the OLS estimates do not provide evidence on the extent to which leadership experience translates into leadership skill development, as the measure of leadership experience might still comprise a component of individual characteristics that determines leadership selection.

In the data studied in Kuhn and Weinberger (2005), there might well be factors that determine whether or not individuals find themselves in leadership positions, and these factors might predate high school. Murphy and Johnson (2011) listed characteristics that are beneficial for becoming a leader. A person with superior skills in a specific field of study or sport would be more likely to be elected president or captain of the respective school club or team. Furthermore, Kuhn and Weinberger (2005: 398) report that the effect of the leadership opportunities that they measured captures "some sort of social skill". Sociability (which is a self-assessed measure indicating whether people like to be around other people) influences whether individuals are granted leadership opportunities in high school, with less-social individuals being less likely to be granted leadership opportunities (Heckman et al., 2006).

The factors that determine the granting of leadership positions might simultaneously also affect the performance of individuals in later working life. Math skills obviously have a positive impact on how individuals perform after high school. Students who were "less social or are bookworms" might not become leaders yet might still achieve higher outcomes subsequently through studying (Rouse, 2012: 115). Because observable omitted selection into the granting of leadership opportunities might subsequently predict income, the measurement of the leadership opportunities correlates with the error term of the income regression.[15] Hence, the influences of math and social skills on leadership selection are consequential for the interpretation: the OLS estimates will not reveal the causal effect of leadership experience on the development of leadership skills. Leadership selection based on math skills might overstate the effect of being granted leadership opportunities on the development of leadership skills, while leadership selection based on social skills might understate the effect of being granted leadership opportunities on the development of leadership skills. Therefore, it is important to recognize and explicitly model the omitted selection that influences whether or not high school students are granted leadership opportunities to estimate the true causal effect of leadership positions on earnings. To overcome this type of selection, we explicitly investigate and model the type of observable omitted selection in the Kuhn and Weinberger (2005) leadership data and empirically make the treatment and control groups comparable using propensity score matching.

4.3.2 Endogeneity as a result of other (unobservable) omitted causes

The treatment and control groups might also differ because of other unobservable omitted causes like omitted variables, simultaneity, common-method variance, or measurement error (Antonakis et al., 2010; Antonakis et al., 2014). Again, these types of unobservable distortions make it difficult to compare the outcome for the treated group with the control group: the other group is unlikely to be counterfactual due to inherent differences in unobservable characteristics (and/or measurement). For example, in the Kuhn and Weinberger (2005) study, personal characteristics are measured at the same time as questions about leadership opportunities are elicited, inviting the risk of simultaneity bias. Similarly, many years after the original data collection, leadership research has identified important dispositional and behavioral variables such as the need for cognitive closure that affect leadership effectiveness (Pierro et al., 2005). Hence,

[15] Related work has shown the interrelatedness of leadership effectiveness and the antecedents for being granted those leadership opportunities. Exhibiting personal maturity and tenacity (Yukl, & Van Fleet, 1992) as well as being sociable and self-confident (Northouse, 1997) have an impact on leadership effectiveness (Judge et al., 2002) and are antecedents to the granting of leadership opportunities (Shamir, 2007; Shondrick et al., 2010).

important variables that affect earnings in real life and the effectiveness of leadership opportunities were not elicited; an omission that could invalidate causal interpretations of correlational findings (Antonakis et al., 2010). Thus, the coefficients that Kuhn and Weinberger (2005) derived could include effects from other omitted causes, resulting in a correlation between the leadership opportunities measured and the error term of the earnings regression. This prevents Kuhn and Weinberger (2005) from interpreting the regression effects found as causal, as the presumed correlational relationship might not converge to the true population parameter (Hamilton, & Nickerson, 2003; Antonakis, et al., 2014). We therefore additionally employ instrumental variable estimations to overcome the potential existence of unobserved causes that simultaneously affect whether or not individuals were granted leadership opportunities and their subsequent earnings.

We begin by discussing how observable omitted selection may influence the leadership positions estimates on earnings and subsequently address how we plan to deal with it. We then proceed to explain how we deal with other unobservable omitted causes, in particular instrumental variable techniques.

4.4 Data and methods

The replication of an existing empirical study can generally follow three different approaches (Schmidt, 2009). First, a direct replication uses original data, or, if original data is not available or accessible, data collected identically. In the latter case, the data collection is based upon the same questionnaire and sample group. The analysis of the data then applies the same theoretical model and statistical techniques to the same hypotheses as the original study. Second, a follow-up study (sometimes called an extension) starts with a direct replication and then extends the original analyses, applying new models or new data to the same underlying theoretical question. Third, a conceptual replication addresses the same research question, but uses a completely different approach and design; data, model and method differ from the original study (Schmidt, 2009).

We conduct a follow-up study, a direct replication and extension, of Kuhn and Weinberger (2005) by including additional data and implementing new statistical techniques. We structure this follow-up study as follows. We begin by exactly replicating the findings from Kuhn and Weinberger (2005) using the original dataset. We then describe the proposed application of additional quasi-experimental statistical techniques to the original data for the follow-up study. We address endogeneity concerns arising from omitted selection influencing both the granting of leadership opportunities in high school and later life earnings using propensity score matching (Rouse, 2012; Li, 2013). We assess the susceptibility of these findings to the omission of confounding variables and simultaneity using sensitivity bounds. In addition, we apply instrumental variable regressions to deal with other omitted causes. The Appendix contains detailed equations and in-depth explanations of the statistical methods employed. We also extend the original sample, which uses white males only, to white females and non-white students, and apply the quasi-experimental methods to these sample extensions. Last, we expand the original time horizon using data from the 2011-2012 Project Study.

4.4.1 Data for replication

We focus our replication and extension on the data originating from the Project TALENT, for which also a Pilot Study from 2011 and 2012 (with 4,879 of the original respondents) is available to test for long-term effects beyond the initial 11-year time window (Stone et al., 2014). Project TALENT was the first and most comprehensive study that surveyed high school students in the US. Initial data collection began in 1960; follow-up interviews took place until 11 years after graduation.[16] The cohort is remarkable in many respects. First, it included more than 400,000 students from almost 1,400 schools across the US. Second, the participants were around 15 years of age, reached their top-level positions around the year 2000, and left the work force around 2015. This made it possible to re-approach them for subsequent surveying regarding the outcomes of work-related experience and stress on medical and health outcomes, a project that was pilot-tested in 2011 and 2012, and is currently underway at the University of Michigan's Survey Research Center (Stone, et al. 2014). Third, the sample is reflective of many trends and effects found in subsequent leadership studies. Arguably, the sample respondents are representative of all US high school students in 1960, and the implications therefore offer foundational insight for subsequent empirical work. Project TALENT data has been investigated across many different disciplines, including economics, psychology and mathematics (Kenny et al., 1979; Austin, & Hanisch, 1990; Wise, 1985).

4.4.2 Direct replication

In a first effort, we directly replicate the findings from Kuhn and Weinberger (2005), to provide

[16] The data that underlies this study stems from the Project Talent. The American Institutes for Research curates this dataset. Interested researchers can obtain the data by contacting the American Institutes for Research via ProjectTalent50@air.org

a solid understanding of the scope of the original study and of the relevant data. We therefore replicate the original results (see Kuhn, & Weinberger, 2005: 405; Table 2) using OLS models with the log of hourly earnings as our dependent variable. We also employ school fixed effects to control for potentially omitted factors associated with the high school each responded attended. Following Kuhn and Weinberger (2005), we restrict our sample to white males who earn between $1 and $50 per hour.

4.4.3 Propensity-score matching

To address endogeneity concerns arising from observable omitted selection, we need to ask two questions: What factors influence whether an individual is granted leadership opportunities? And what is the impact of these leadership opportunities (conditional on the influencing factors) on subsequent earnings?

We employ propensity score matching as a non-experimental way of isolating the effects of a treatment variable on an outcome based on observable confounders. Propensity score matching controls for a set of available variables that are likely to influence the propensity of being granted leadership opportunities and later-life earnings at the same time. Controlling for all potential variables is, of course, infeasible as some of them might not be observable, but we try to include as many variables as possible.

The matched control group is found by estimating the individual conditional predicted probabilities (a single variable summarizing how likely each individual is to be granted leadership opportunities conditional on observable variables). Individuals in the treatment group are matched with individuals who did not have a leadership position, but have similar predicted probabilities to form the control group. These matched individuals are considered statistically equal, and thus, each provides the counterfactual outcome for each other. By focusing on this predicted probability, we derive the counterfactual based on several antecedents to leadership opportunities granted simultaneously.

Our work follows prior related empirical literature in the leadership and human resource domain to overcome endogeneity concerns stemming from observable omitted selection through matching (Jez, 2014; Dale, & Krueger, 2002). Rouse (2012), especially, shares similarities with our approach. She addressed confounding variables (both observable variables using matching techniques but also unobservable variables using other approaches) and studied the effect of leadership positions in high school on subsequent educational attainment. As in experimental studies, such as Andersen and Lu (2016), we would like to get as close as possible to a randomization of leadership opportunities in our propensity score matching procedure within the bounds of the observed variables available. Our procedure follows Rouse (2012) in spirit, but we employ hourly earnings eleven years after high school instead of academic achievements.

We estimate a probit model with an individual's probability of being granted leadership opportunities high school as the dependent variable. This so-called propensity score is equal to the conditional probability of receiving the treatment (being granted a leadership opportunity in high school) (Rosenbaum, & Rubin, 1983).

After calculating the propensity score, we match individual subjects based on their respective propensity score, essentially re-weighting the sample. As a consequence, subjects with the same probability of being granted leadership opportunities are treated as statistical twins. After matching, given a set of observable variables in the dataset, subjects are identical in all observable aspects except for receiving the treatment. The basic idea behind making causal inferences based on the propensity score matching procedure is that if two subjects have the same probability of receiving treatment (the same propensity score), yet are in different groups (leaders or non-leaders), we are, statistically speaking, comparing two individuals who might have been exogenously assigned to the leadership and non-leadership groups. Hence, after matching, we can directly infer the differences between the matched groups. This allows us to estimate the net effect of the treatment on outcomes, much like in experiments that randomize the treatment assignment (Andersen, & Lu, 2016).

4.4.3.1 Empirical matching procedure

In carrying out our empirical investigation, we follow Li (2013). First, we establish the (theorized) notion of endogeneity based on observable confounders in leadership opportunities. We report the estimates of the regressions (probit and OLS, respectively) using high school leadership positions and earnings eleven years later as the dependent variable. We augment the personal data variables included in Kuhn and Weinberger (2005) and include additionally the number of *Dates* per week and the financial situation of the family of the respondent in our matching equation. Addressing personal characteristics, we further employ standardized grade-relative measures for all available test scores (Table 11 in the Appendix contains a list of all variables employed). To proxy for recognitions of achievement, we also include different awards that students won in the three years leading up to the interview date. Last, we include personal interests in our matching equation to identify those who (at the interview date) were striving to become leaders in later working life. Because we cannot control for all possible factors influencing the leadership selection, we will assess the susceptibility of our results to omitted variables and other omitted causes subsequently.

Nevertheless, if at least some of these variables are related to both leadership opportunities in high school and earnings eleven years later, we have established grounds to suspect an endogenous relationship between predictor variables, observable confounders, and the granting of leadership opportunities. This might cause differences between the true causal effect and the expected value of the OLS estimates provided in Kuhn and Weinberger (2005). Being able to overcome observable omitted selection can therefore help to make causal inferences regarding the effect of high school leadership experience on earnings.

We calculate the propensity score following Becker and Ichino (2002) and carry out the matching procedure using psmatch2 in Stata 13 (Leuven, & Sianesi, 2018). The most crucial part here is to assess the balance of the sample after matching. This involves testing whether differences in the mean values for predictor variables persist after matching, or whether these can be successfully removed through matching. If matching removes statistical differences, endogeneity concerns based on observable selection are alleviated, and the procedure has created statistical twins. Kuhn and Weinberger (2005) include all leadership position variables into the same regression, so their interpretation is always against the omitted "no leadership" group.

Our PSM analysis differs slightly here in terms of comparison group and sample composition: We individually estimated the effects of each leadership category. In doing so, we compared those who were presidents and captains to all other respondents. In this way, our results regarding president and captain would have a stronger interpretation regarding the relationship between leadership and earnings in later-life than the original findings (we report a direct OLS comparison in Table 2B in the online appendix). We did, however, compare those were presidents or captains against those who were not granted leadership opportunities making the results directly comparable with the corresponding OLS findings.

We therefore employ different matching models to infer leadership coefficient estimates vis-à-vis the respective control group. After balancing, we estimate the samples' average treatment effects on the treated. These effects provide evidence as to what would have happened had the leader not led (conditional on the included observable variables). Would earnings have been higher, lower, or unaffected?

4.4.3.2 Sensitivity analysis for omitted variables in the PSM estimation

The PSM adjusts for omitted selection based on observable variables that correlate with a suspected endogenous regressor, but is, of course, limited to those variables available in the Project TALENT dataset. The results would still not unveil the actual causal leadership effect if there were selection on other unobservable variables or other omitted causes, which correlate with the disturbance in the outcome regression. We tackle this problem using two additional measures: calculating PSM sensitivity bounds and using instrumental variables.

Matching estimators are susceptible to the omission of variables that predict both the selection into treatment and the subsequent earnings (Guo, & Fraser, 2014; Dehejia, & Wahba, 2002). In response, we estimate Rosenbaum bounds to test for potential unobservable confounding variables (Rosenbaum, 2002). Rosenbaum bounds help to determine how strongly an unobservable confounding variable must influence the leadership variable to undermine the conclusions about the causal impact of leadership positions on subsequent earnings derived from the PSM estimates (DiPrete, & Gangl, 2004). Hence, Rosenbaum bounds check how sensitive the treatment effect estimates are to the inclusion of new variables that would affect both the probability of receiving the treatment and the outcome variable. Rosenbaum bounds do not rely on the search for variables that satisfy the assumptions for instrumental variables. Rather, they work on the assumption that there are unobservable confounding variables, and they can thus be applied to derive the sensitivity of the PSM estimates (Peel, & Makepeace, 2012). The bounds indicate how the confidence interval for the estimated treatment effects would change if unobservable confounding variables exist. For example, for the positive leadership position coefficient under investigation, the confidence intervals would widen (and include zero) if there are unobserved variables that can cause the odds ratio of being granted leadership opportunities in high school to differ between the treatment and comparison groups by the calculated value of the test statistic. Only variables that greatly increase the odds of treatment and outcome simultaneously are considered problematic. The Rosenbaum bounds convey worst-case information regarding the uncertainty inherent in matching estimators. They show how influential a potential confounding variable must be, in order to completely undermine the causal conclusions drawn from the matching analysis (DiPrete, & Gangl, 2004; Peel, & Makepeace, 2012). We report the sensitivity of each significant treatment effect using Rosenbaum bounds.

4.4.4 Overcoming endogeneity as a result of other (unobservable) omitted causes through instrumental variables

To overcome the problem of endogeneity arising from unobserved causes, we follow Antonakis et al. (2010: 1100), who suggest that "the coefficient of x could be interpreted causally if an exogenous source of variance, say z, were found that strongly predicts x and is related to y via x only, and unrelated to e (the combined term)." We therefore exploit instrumental variable techniques, where instruments are strongly related to the suspected endogenous leadership variable but only indirectly connected with the earnings measure via the instrumented variables. We isolate the variance in the error term of the outcome equation that the control variables and the instruments predict in the suspected endogenous leadership variable to recover the causal effect of leadership opportunities on earnings (Schaffer et al., 2013). Employing instrumental variable techniques is an additional strategy for the consistent estimation of causal effects between leadership opportunities in high school and later-life earnings. Importantly, because the uncertainty embedded in the instrumental variable estimation is different from the uncertainty underlying the preceding matching approach both approaches increase the information about the causal effect of the variable of interest on the outcome (Antonakis et al., 2010; DiPrete, & Gangl, 2004). In the corresponding first stage, the endogenous regressor (leadership variable) is regressed on the instruments and covariates (Bascle, 2008). The first stage therefore isolates the variation in the endogenous variable that is not correlated with the error term of the outcome regression. The corresponding instrumented (fitted) value of the endogenous variable is used in the second stage in lieu of the endogenous regressor (Bascle, 2008).

More specifically, we rely on information about parental behavior as instrumental variables. According to social learning theory, socialization occurs via observational (vicarious) learning from role models (Bandura, 1977). As such, the social context in which children find themselves in has a tremendous impact on how they form their own self-image (Bandura, 1977).

Especially in early adolescence, children develop self-concepts about who they are and who they would like to be. In this respect, parents can act as reflective modelling agents (Wiese, & Freund, 2011). Hence, children will engage in a process of cognitive evaluations about behaviors they have seen (or heard of) by their parents or may react positively to suggestions made by their parents about certain types of behavior. We would therefore suggest that if parents were members of clubs or teams, children would engage in similar behaviors and also join clubs or teams, which in turn, might increase their chances of being granted leadership opportunities in similar teams or clubs. As a case in point, Seppänen (1982) showed that parents (sports) club membership links to the type of sports club their children join.

We therefore employ the information provided about whether each parent was a member of a club or a team during the child's adolescence as instrumental variables. We test for each effect (President and Captain, Captain only, and President only) separately. In order to conduct this analysis, parental team/club membership must not influence children's subsequent earning directly when conditioning on the endogenous regressor children team/club membership. As such, theoretically the team or club membership of the parents should only affect the selection of children into activities through parental role modelling but does not allow an inference where team or club membership of the parents predicts earnings of the children some 11 years later.

We think that prior research supports this notion. Eime et al. (2013) showed that only a very limited number of types of child sport memberships (e.g. golf or tennis) were associated with higher socioeconomic status of the parents while many other types of sport memberships (e.g. gym or martial arts) are unrelated to parental socioeconomic status. Also, extracurricular involvement (such as club membership) among adolescents is strongly associated with perceived parental support which in turn is uncorrelated with socio-economic status (Anderson et al., 2003; Marcen et al., 2013).

Our analysis does not specifically differentiate between different types of club or team memberships but rather subsumes very many possible types of club membership (comprising religious groups, charity work, school boards, unions,) and team membership (sports, arts, crafts, music) for the parents that span very many socioeconomic groups. We capture all different types of leisure-time sport team or community club membership of the parents (for example, the dummy variables *Mother Team* possesses the value of *1* if the mother is member in any type of team). Hence, we do not suspect that there are effects between parents' club and team membership and children's subsequent wages eleven years later.

Furthermore, we employ the control variables *Wealthy* and *Comfortable* in our main analyses. These variables capture whether respondents' families currently can afford a comfortable or wealthy lifestyle. These variables derive from the question "Which of the following best describes your family's finances?" and the answer options ranged from "A – Barely able to make a living" to "F – Extremely Wealthy".

Of course, we also provide empirical evidence consistent with the validity of the instruments by, for example, testing the overidentifying restrictions in all IV regressions (Bascle, 2008; Moreira, 2003). In testing for weak instruments, we follow the work of Stock and Yogo (2005). The presence of instrument weakness may cause IV coefficient estimates that are biased towards OLS coefficients and to understate confidence intervals (Becker et al., 2011). For a model using a single endogenous variable, the first-stage F-statistic can be used to assess the weakness of instruments. The corresponding weakness is expressed as the size of the bias of our IV estimator relative to that of the corresponding OLS estimator.

Last, we also address an extension to deal with potentially weak instruments where the exclusion restrictions are potentially violated. We therefore employ the estimator suggested in Lewbel (2012) to identify the parameters of interest. The approach generates additional instruments and ensures identification through heteroskedastic covariance restrictions (Lewbel, 2012). In the case of Lewbel (2012), identification is achieved by introducing regressors that are uncorrelated with the product of the heteroskedastic errors. This is often the case if error correlations stem from unobserved common factors (as is hypothesized in our case). This extension supplements external instruments and subsequently helps to improve the efficiency of the standard IV estimator (Baum, & Lewbel, 2019; Lewbel, 2012).

4.4.5 Extension to white females, non-whites, and a follow-up study

Kuhn and Weinberger (2005: 399) noted that they "[…] study the labor market valuation of leaders without the confounding effects of race or gender discrimination or of the changing roles and expectations of women during this time period." The data which Kuhn and Weinberger (2005) drew upon includes many females, for which limits in their career advancement have been scientifically investigated subsequently (Matsa, & Miller, 2011; Ragins et al., 1998). The exclusion of females hence severely limits the generalizability of the results. As far as female leadership is concerned, studies, using instrumental variables to control for endogeneity, have shown positive effects of leadership opportunities in high school on women's later performance. Athletic participation increased both college attendance and labor force participation among women (Stevenson, 2010). This finding is similar to that of Eide and Ronan (2001), who found that athletic participation is associated with educational outcomes positively for white females but negatively for white males. In light of these findings, we study if the positive leadership effects reported for males in the Project TALENT sample also materialize in the white female subgroup.

Kuhn and Weinberger's (2005) explicit exclusion of non-whites might also affect the generalizability of the results. Among other things, the focus on white males limited the implications that could be drawn regarding relationship between leadership positions in high school and subsequent occupational earnings for a wider population, and thus for policy implications regarding leadership training initiatives. Using data from the Project TALENT and the NCES studies, Weinberger (2014b) explored the leadership activities of black students and analyzed the link between leadership engagement and subsequent occupational earnings. First, and in contrast to other work focusing on minorities (Lozano, 2008), Weinberger (2014b) found that black high school students engage at the same rate in extracurricular activities as white high school students, yet the earnings premium generally associated with leadership opportunities in high school was absent for black males. Importantly, Project TALENT was initiated in 1960, four years before the Civil Rights Act of 1964, which, in its Title VII, legally mandated equal pay for equal work, irrespective of the race of the employee. After 1964, the pay gap between Whites and Blacks decreased, but remained substantial (Brown, 1984). Investigations using quasi-experimental methods can help to better understand the distinctive relationships between leadership opportunities and wages for different ethnicities.

As a starting point, we provide the OLS regressions for white females, non-white males, and non-white females using the same format and employing the same control variables as the direct replication from Model 4 in Table 2. We then extend these OLS estimates by employing propensity score matching for white females and both non-white genders. To do so, we first calculate the Matching Models 1, 2 and 3 as well as the Outcome Assessment Model 1 for white females, non-white males, and non-white females to assess whether there are variables that simultaneously affect treatment and outcome. Subsequently, we estimate the propensity score and carry out the matching procedure for white females, non-white males and non-white females. This allows us to assess potential causal relationships between being granted leadership opportunities in high school and wages in later working life for groups not covered in Kuhn and Weinberger (2005). Again, we estimate the sensitivity of the treatment effects using Rosenbaum bounds for each sample. Additionally, we employ the instrumental variable regressions to control for other omitted causes.

We also investigate the long-term effect that leadership opportunities in high school have on later life income. Toward this end, we extend the original time horizon by including data from the 2011-12 Pilot Study for a subsample of the original data. This should provide a more detailed and robust understanding of whether being granted leadership opportunities in high school is a mere signal for subsequent job market performance (and for which group of leaders), or whether it equips individuals with skills that provide useful beyond the initially-studied 11-year window. The 2011-12 Pilot Study documents household income, which can be used to assess whether the effect of leadership opportunities in high school on later life income is consistent in strength and direction with the individual earnings results reported in Kuhn and Weinberger (2005). As the household income is reported on a five-scale ordinal variable, we employ an ordered logistic regression using the same independent variables as in the direct replication.

4.5 Results

Table 4.1 reports the descriptive statistics and the correlations of the variables employed in Table 2 in Kuhn and Weinberger (2005).[17]

4.5.1 Direct replication

We begin our replication and extension by exactly replicating the results from Kuhn and Weinberger (2005). Results are tabulated in Table 4.2. Model 1 represents the baseline model investigating the effects of leadership positions and controlling for mere club and team memberships on hourly wages. Like Kuhn and Weinberger (2005), we then add the math score in Model 2 to control for cognitive skills. Models 3 and 4 also include variables for lack of education on the part of the respondent's parents and on the part of the respondent, respectively. Furthermore, all models include grade and school attainment control variables, as well as (unreported) school dummies. Table 4.2 reports the findings for those four models for a sample size of 24,041 white males. This equals the sample size of Table 2 in Kuhn and Weinberger (2005).

[17] Correlations for all variables employed in the subsequent analyses and for the new samples are available as Tables 1A, 1B, 1C and 1D in the appendix.

Table 4.1: Descriptive statistics and correlations for KW sample

Variable Name	M	SD	1	2	3	4	5	6	7	8	9	10	11
Log (Hourly Earnings)	1.58	0.45											
Both Captain and President	0.21	0.41	0.07***										
Captain Only	0.13	0.33	0.01	-0.20***									
President Only	0.24	0.42	0.02*	-0.29***	-0.21***								
Both on Team and in Club	0.76	0.42	0.09***	0.17***	0.11***	0.05***							
On Team Only	0.02	0.15	-0.01	-0.04***	0.01†	-0.05***	-0.29***						
In Club Only	0.19	0.39	-0.08***	-0.15***	-0.11***	-0.01*	-0.88***	-0.08					
Math Score	53.15	28.53	0.22***	0.02**	-0.03***	0.10***	0.08***	-0.04***	-0.05***				
High School	0.50	0.50	0.03***	0.02**	0.00	0.00	0.03***	0.00	-0.02***	0.04***			
College Degree	0.20	0.40	0.06***	0.03**	-0.01*	0.09***	0.10***	-0.04***	-0.04***	0.23***	-0.50***		
Some College	0.23	0.42	-0.04***	0.00	0.00	-0.01*	-0.01†	0.01†	0.01	-0.08***	0.06***	-0.75***	
College Degree or Higher	0.40	0.49	0.19***	0.06***	-0.01	0.09***	0.10***	-0.05***	-0.06***	0.47***	0.00	0.23***	-0.45***

† $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Note: Pairwise correlation coefficients derived from Pearson-correlations. N. of obs. is 24,041.

Table 4.2: Direct Replication – *OLS of KW Table 2*

	Model 1 Log (Hourly Earnings)	**Model 2** Log (Hourly Earnings)	**Model 3** Log (Hourly Earnings)	**Model 4** Log (Hourly Earnings)
Leader				
Both Captain and President	0.054***	0.049***	0.049***	0.038**
	(0.012)	(0.012)	(0.012)	(0.012)
Captain Only	0.036**	0.036**	0.036**	0.035**
	(0.013)	(0.013)	(0.013)	(0.013)
President Only	0.036***	0.020†	0.019†	0.010
	(0.011)	(0.011)	(0.011)	(0.011)
Member				
Both on Team and in Club	0.107***	0.073**	0.070**	0.055*
	(0.025)	(0.024)	(0.025)	(0.025)
On Team only	0.083*	0.062†	0.060†	0.059†
	(0.035)	(0.034)	(0.034)	(0.034)
In Club Only	0.036	0.008	0.005	-0.006
	(0.026)	(0.026)	(0.026)	(0.026)
Controls				
Math Score		0.002***	0.002***	0.001***
		(0.000)	(0.000)	(0.000)
Parent´s Education				
High School			0.020†	0.011
			(0.011)	(0.011)
College Degree			0.015	-0.008
			(0.014)	(0.014)
Educational Attainment				
Some College				0.051***
				(0.012)
College Degree or Higher				0.136***
				(0.013)
School-fixed Effects	Yes	Yes	Yes	Yes
F-Value	23.58	45.63	34.87	39.30
p > F	0.000	0.000	0.000	0.000
Adjusted R^2	0.160	0.177	0.178	0.189
Observations	24,041	24,041	24,041	24,041

† p<0.1, * p<.05, ** p<.01, *** p<0.01

Note: Table 4.2 directly replicates the results depicted in Kuhn and Weinberger (2005: 405 (columns 5-8). The coefficients are derived from OLS regressions with standard errors in parentheses. All models include (unreported) grade and school attainment control variables as well as school dummies. Models 3 and 4 include an unreported dummy variable for lack of parent's education. Model 4 includes an unreported dummy variable for lack of educational attainment.

Importantly, we are able to derive almost the exact coefficients as in the original work; visible differences only exist in the third digit. The coefficient associated with being both, captain and president, is significant in all regressions (at the 1 percent level), and ranges between 0.038 and 0.54. Similarly, the coefficient for those who gained experience as captain only is also positive (ranging between 0.035 and 0.036) and significant at the 1 percent level throughout all specifications. Last, we find that the coefficient associated with those who act as president only is positive and significant in three out of four regression specifications (ranging between 0.019 and 0.036) and significant at least at the 10 percent level. To summarize, we can corroborate the original findings in Kuhn and Weinberger (2005) in significance and direction, and are almost exact in our coefficient and standard error estimations.[18]

4.5.2 Grounds for endogeneity concerns based on observable omitted selection

Regarding omitted selection based on observable variables, Table 4.3 depicts the results of Outcome Assessment Model 1 as well as the Matching Model 1, 2 and 3 based on a sample size of 22,095 white males.[19] Among the variables that affect both, the log of the hourly earnings and either of the three dichotomous leadership variables, we find several variables with a significant (positive and negative) influence on both the outcome variable and the three individual predictor variables. Among others, we find that those who are overweight are earning less and are less likely to act as both, president and captain. Wealthier students, those with a higher business score, and those more interested into businesses are more likely to earn more and are more likely to act as club president. A higher military score is associated with higher later life earnings and the likelihood of acting as captain and president. A higher English score is interestingly related to lower earnings, but a higher likelihood of acting as captain and president, as well as acting as a president only. Fine arts awards are associated with lower later life earnings, but a higher chance to act as a president and a lower chance to act as a captain of a team. Reading skills are related to higher earnings, but also induce a lower likelihood of acting as both, presi-

[18] Table 2A in the appendix reports the same models as Table 2 but with robust standard errors. Table 2B in the appendix reports effects for each leadership category individually (with the no leadership category as the reference group). The interpretation of both tables does not differ from the interpretation of Table 2.

[19] This sample size is slightly smaller compared to the sample size of 24,041 in Table 2 of Kuhn and Weinberger (2005), as we employ additional variables with more missing information and thus more missing observations. The slightly smaller sample size does not affect the interpretations and conclusions drawn from our matching analysis. We tested the susceptibility of the previous estimates to the sample reduction. In summary, we find no material differences in our estimations.

dent and captain. All in all, we find that there exist several variables that affect both, the outcome variable and the three predictor variables simultaneously which provides grounds to suspect an endogenous relation between the predictor variables and the outcome of interest.

4.5.3 Endogeneity adjustments based on observable omitted selection

We report how well the matching procedure balances the treatment and control sample in Balancing Model 1, 2 and 3 in Table 4.4.[20] Balancing Model 1, with a sample of 22,095 white males, provides the basis for comparing students who were team captains and club presidents during high school to all other students. When investigating those who were only president or captain it is important not to include non-leaders and those that were presidents and captains (those that were the main focus of the analysis in model 1) simultaneously in the control group as the effects of president and captain and non-leadership might cancel each other out. We therefore prune the analysis of those who were captain and president in the two subsequent analyses. Balancing Model 2, with a sample of 12,223 white males, provides the basis for comparing all students who were only team captains but not club presidents to those that were not leaders. Balancing Model 3, with a sample of 14,653, provides the basis for comparing all students who were only club presidents but not team captains to those that were not leaders. We also report the reduction in bias for each matching covariate that the procedure resulted in.

[20] The propensity-score matching approach does not include school dummies.

Table 4.3: Conceptual replication - *Endogeneity susceptibility of KW Table 2 estimates*

	Outcome Assessment Model 1 Log (Hourly Earnings)	**Matching Model 1** President and Captain	**Matching Model 2** Captain Only	**Matching Model 3** President Only
Leader (for Model 1 only)				
Both Captain and President	0.046**			
	(0.015)			
Captain Only	0.034*			
	(0.017)			
President Only	-0.005			
	(0.013)			
Member				
Both on Team and in Club	0.012	1.573***	1.161*	0.901
	(0.027)	(0.299)	(0.463)	(0.635)
On Team only	0.042	0.971*	1.019*	0.258
	(0.039)	(0.382)	(0.517)	(0.688)
In Club Only	-0.024	0.707*	0.089	0.788
	(0.028)	(0.317)	(0.477)	(0.638)
Personal Data				
Overweight	-0.122*	-0.611†	0.455	-0.138
	(0.048)	(0.332)	(0.462)	(0.279)
Underweight	-0.033	-0.388	-0.162	0.309
	(0.037)	(0.281)	(0.270)	(0.233)
Tall	-0.022*	-0.032	-0.020	-0.027
	(0.011)	(0.077)	(0.089)	(0.070)
Short	0.024	0.098	-0.259†	0.017
	(0.021)	(0.119)	(0.140)	(0.116)
Dates	-0.001	0.135***	-0.072*	0.041
	(0.004)	(0.029)	(0.033)	(0.028)
Comfortable	0.012	0.013	-0.096	0.115
	(0.016)	(0.104)	(0.115)	(0.094)
Wealthy	0.073***	0.042	-0.153	0.198†
	(0.019)	(0.124)	(0.136)	(0.113)
Personal Characteristics				
Sociability	0.021***	0.044	-0.045	0.085*
	(0.006)	(0.043)	(0.048)	(0.036)
Vigor	0.002	0.228***	0.164**	-0.134***
	(0.007)	(0.045)	(0.051)	(0.040)
Mature	-0.002	0.079†	-0.180***	0.126**
	(0.008)	(0.046)	(0.053)	(0.045)
Self-Confidence	0.008	0.023	0.019	0.113***
	(0.005)	(0.039)	(0.044)	(0.033)
Tidiness	0.007	-0.034	0.027	0.006
	(0.006)	(0.041)	(0.048)	(0.038)
Test Scores				
Math Score	0.002***	0.002	0.004†	0.001
	(0.000)	(0.002)	(0.002)	(0.002)
Vocabulary Score	-0.008	-0.067	-0.003	-0.227
	(0.029)	(0.191)	(0.224)	(0.179)
Social Studies Score	-0.024	-0.370*	0.143	0.103
	(0.027)	(0.167)	(0.204)	(0.167)
Science Score	-0.006	-0.274	-0.710***	0.451**
	(0.027)	(0.188)	(0.212)	(0.163)
Scientific Attitude Score	0.088***	-0.020	-0.155	0.134
	(0.020)	(0.131)	(0.166)	(0.129)
Law Score	0.013	-0.218	-0.253	0.069
	(0.022)	(0.145)	(0.177)	(0.137)
Military Score	0.058**	0.278*	-0.192	0.069

	(0.022)	(0.135)	(0.168)	(0.125)
Business Score	0.081***	-0.248	0.064	0.272*
	(0.021)	(0.154)	(0.178)	(0.135)
Etiquette Score	0.010	0.010	-0.170	0.062
	(0.018)	(0.125)	(0.148)	(0.114)
English Score	-0.056*	0.343†	-0.009	0.370*
	(0.025)	(0.180)	(0.205)	(0.174)
Awards				
Science Awards	-0.001	0.014	0.020	-0.020
	(0.007)	(0.039)	(0.054)	(0.037)
Fine Arts Awards	-0.006†	0.005	-0.088*	0.042*
	(0.003)	(0.021)	(0.035)	(0.020)
Sports Awards	0.001	0.109***	0.027	0.018
	(0.002)	(0.013)	(0.018)	(0.013)
Cognitive Skills				
Arithmetic Skills	-0.007	-0.004	-0.047	-0.210†
	(0.020)	(0.131)	(0.147)	(0.111)
Reading Skills	0.039*	-0.226†	0.097	0.096
	(0.018)	(0.120)	(0.145)	(0.109)
Clerical Skills	0.009	0.067	-0.076	-0.048
	(0.019)	(0.126)	(0.142)	(0.118)
Identification Skills	-0.002	-0.280*	-0.034	0.052
	(0.019)	(0.128)	(0.144)	(0.112)
Personal Interests				
Public Service Interest	0.007	0.111*	-0.086†	0.076†
	(0.007)	(0.048)	(0.050)	(0.040)
Business Management Interest	0.012†	0.014	0.139**	0.035
	(0.007)	(0.047)	(0.053)	(0.038)
School-fixed Effects	No	No	No	No
F / Chi^2	13.83	457.81	141.42	226.56
p > F / Chi^2	0.000	0.000	0.000	0.000
R^2 / Pseudo R^2	0.067	0.080	0.038	0.034
Observations	22,095	22,095	22,095	22,095

† p<0.1, * p<.05, ** p<.01, *** p<0.01

Note: Table 4.3 assesses the need for a potential endogeneity adjustment of the results depicted in Kuhn and Weinberger (2005). The coefficients for Outcome Assessment Model (1) are derived from OLS regressions, coefficients for Matching Model 1, 2 and 3 are derived from logit regression with standard errors in parentheses. The dependent variable in (1) is the natural logarithm of hourly earnings, the dependent variable in the matching models 1,2, and 3 correspond to a dichotomous indicator variable whether individuals acted as captain and president, captain only, or president only. Table 4.3 does not include school dummies as the propensity-score matching approach does not include school dummies. Table 3A in the online Appendix shows the same models with school dummies.

Table 4.4: Conceptual replication - *PSM Efficiency for KW sample*

	Balancing Model 1 Treatment: President and Captain			Balancing Model 2 Treatment: Captain Only			Balancing Model 3 Treatment: President Only		
	Mean		Bias Reduction	Mean		Bias Reduction	Mean		Bias Reduction
	Treated	Controls	in %	Treated	Controls	in %	Treated	Controls	in %
Member									
Both on Team and in Club	0.906	0.899	95.9	0.888	0.885	98.6	0.804	0.798	96.6
On Team only	0.012	0.012	98.4	0.029	0.028	76.3	0.010	0.011	96.8
In Club Only	0.077	0.083	96.3	0.077	0.081	97.8	0.181	0.185	96.0
Personal Data									
Overweight	0.008	0.008	98.8	0.014	0.016	0.9	0.013	0.013	90.0
Underweight	0.013	0.013	93.8	0.019	0.020	83.4	0.019	0.020	79.4
Tall	0.275	0.274	86.4	0.287	0.290	92.2	0.277	0.279	93.7
Short	0.089	0.087	85.6	0.083	0.0870	82.5	0.084	0.084	99.4
Dates	1.346	1.359	95.3	1.182	1.1860	98.3	1.172	1.194	89.2
Comfortable	0.686	0.685	86.3	0.674	0.670	82.2	0.697	0.693	-690.7
Wealthy	0.195	0.196	94.5	0.183	0.190	79.0	0.184	0.186	94.1
Personal Characteristics									
Sociability	0.272	0.274	99.3	0.108	0.113	98.6	0.155	0.154	99.6
Vigor	0.338	0.339	99.6	0.161	0.165	98.9	0.117	0.110	97.9
Mature	0.246	0.251	98.4	-0.013	0.007	85.1	0.238	0.232	98.5
Self-Confidence	0.181	0.167	93.0	0.000	0.011	92.3	0.200	0.183	95.0
Tidiness	0.188	0.201	92.8	0.037	0.035	98.9	0.171	0.176	98.2
Test Scores									
Math Score	54.698	54.788	93.3	51.561	51.957	42.8	58.778	58.340	94.5
Vocabulary Score	0.513	0.516	67.0	0.484	0.490	71.3	0.571	0.569	96.2
Social Studies Score	0.518	0.521	67.4	0.501	0.508	20.6	0.568	0.566	96.6
Science Score	0.512	0.516	64.9	0.480	0.487	77.1	0.572	0.568	92.4
Scientific Attitude Score	0.517	0.521	-206.6	0.497	0.504	-17.7	0.551	0.549	96.0
Law Score	0.514	0.514	89.1	0.492	0.497	67.2	0.552	0.548	90.7
Military Score	0.523	0.524	95.1	0.493	0.497	-371.0	0.541	0.542	98.5
Business Score	0.513	0.513	97.8	0.487	0.487	98.7	0.557	0.555	96.3
Etiquette Score	0.512	0.514	69.3	0.484	0.487	79.4	0.526	0.524	91.1

English Score	0.528	0.530	89.5	0.494	0.495	-65.1	0.573	0.571	97.4
Awards									
Science Awards	0.388	0.404	88.5	0.242	0.241	97.8	0.324	0.350	79.9
Fine Arts Awards	0.833	0.890	80.2	0.5059	0.5277	69.2	0.758	0.808	85.0
Sports Awards	2.725	2.762	97.1	2.1040	2.1693	94.0	1.837	1.901	92.2
Cognitive Skills									
Arithmetic Skills	0.509	0.509	99.3	0.5064	0.5034	26.1	0.528	0.530	90.2
Reading Skills	0.501	0.498	70.9	0.5048	0.5027	68.4	0.512	0.511	-186.1
Clerical Skills	0.520	0.520	97.2	0.5078	0.5055	74.7	0.534	0.536	92.0
Identification Skills	0.496	0.492	72.3	0.4973	0.4906	59.2	0.510	0.512	50.2
Personal Interests									
Public Service Interest	0.170	0.168	98.8	0.0147	0.0063	95.2	0.130	0.124	97.9
Business Management Interest	0.167	0.151	92.2	0.0586	0.0381	89.7	0.091	0.093	99.1
	LR Chi^2	p > Chi^2	Mean Bias	LR Chi^2	p > Chi^2	Mean Bias	LR Chi^2	p > Chi^2	Mean Bias
Unmatched	1821.64	0.000	13.5	1315.21	0.000	13.1	1708.29	0.000	18.9
Matched	8.82	1.000	0.8	8.48	1.000	1.3	8.81	1.000	1.0

Note: Table 4.4 depicts the mean values and bias reduction derived from Propensity Score Matching. The balancing models are based on a dichotomous treatment indicator variable that equals one if individuals acted as captain and president (balancing model 1), captain only (balancing model 2), or president only (balancing model 3), and zero otherwise.

Table 4.4 reports that the mean bias for the unmatched sample ranges from 13.5 percent (for the effect of presidents and team captain on earnings) to 13.1 percent (for those who are only captains) and 18.9 percent (for those who are presidents only). In summary, the estimates derived for the effect of leadership opportunities on earnings may still include the extent to which leadership experience reflects prior characteristics that determine leadership selection. The true causal effects that reflect the extent to which leadership opportunities induce the development of leadership skills that lead to higher earnings could be smaller or larger. Given the size of the coefficient estimates, this might result in overestimating its statistical significance. Importantly, the matching procedure reduces the bias to between 0.8 and 1.3 percent, which indicates that treatment effects derived from such an estimation would be close to the true parameter estimate when selection based on observable characteristics is taken into account. The model achieves substantial bias reductions in almost all variables. Large differences in model accuracy are reported for variables that are very close in values before matching, such as the scientific attitude score in balancing model 1, the military and English score in balancing model 2, and reading skills and comfortable assessment in balancing model 3. Importantly, none of the Chi^2 tests indicates that there are variables that affect the predictor variables of leadership positions significantly after matching has been applied. Consequently, we are confident that balance has been achieved through propensity score matching.

4.5.4 Treatment effect estimation

The treatment effect results are tabulated in Table 5.[21] We report the coefficient estimate and standard errors and also report the corresponding Rosenbaum bounds for the susceptibility of the derived estimates to potential unobserved confounders. We report the average treatment effect on treated (ATT). This measures the average effect of the treatment on the group of individuals that received treatment. Thus, the ATT informs us how much an individual that was granted a leadership opportunity in high school gained (or lost) in earnings as a consequence of having acted as president, captain, or both.

[21] Li (2013) pointed out that the choice of matching estimators might affect the estimates derived. We provide estimates from different matching estimators, kernel and radius matching, to supplement the propensity score matching results. The positive effects for those who were president and captain remain invariant. The results for those who were only president remain insignificant. The results also suggest a significant ATT for those who were only captain. In addition, we also employ a multinominal treatment regression to assess the validity of our matching results. Hereby, we create a four-point scaled treatment variable that is 0 for those who were neither team captains nor club presidents in high school, 1 for those who were only team captains, 2 for those who were only club presidents, and 3 for those who were team captains and club presidents. We conduct this analysis using the mtreatreg package for Stata (Deb, 2009). We find positive and significant coefficients for those who were president and captain and those who were only captain while those who were only president do not seem to have higher wages in later working life.

To begin with, we find that the treatment effect estimate (ATT=0.044) for those that acted as presidents and captains is similar to the OLS estimate reported in table 2 (ß=0.038) and the coefficient is still significant at the one-percent level.[22] Consequently, the original direction and interpretation remains invariant. Moreover, the coefficient is only slightly susceptible to a potential omitted confounder because only variables that increase the log of the odds of being president and captain by 1.2 and that simultaneously affect the log of the hourly earnings would render the treatment effect insignificant. Given the large number of variables included in the matching procedure and the original bias being only 13 percent, we consider this scenario unlikely.

For those who acted as captain only, we also find a smaller coefficient estimate (ATT=0.017) that is only significant at the 10 percent level. Moreover, the coefficient estimate is highly susceptible to a potential omitted confounder, as already small effects from confounding variables would render the treatment effect insignificant. Last, we did not find a significant treatment effect for those that acted as presidents only. The coefficient does not attain statistical significance. As such, we can corroborate the earnings effect for those acting as captains and presidents, but we did not find supportive evidence for those acting as presidents or captains only.

Table 5: Conceptual replication – *PSM Treatment Effect Assessment for KW sample*

	Average Treatment Effect on the Treated	**Sensitivity of Estimate (Rosenbaum Bounds)**
Treatment		
President and Captain	0.044***	1.21
(N=22,095)	(0.008)	
Captain Only	0.017†	1.04
(N=12,223)	(0.010)	
President Only	0.001	1.00
(N=14,653)	(0.008)	

† p<0.1, * p<.05, ** p<.01, *** p<0.01.

Note: Table 5 reports the Average Treatment Effect on the Treated derived from Propensity Score Matching with standard errors in parentheses. Rosenbaum bounds are estimated for the respective Average Treatment Effect on the Treated for p = 0.05. The sample for the treatment analysis is based on the sample of white males employed in Kuhn and Weinberger (2005).

22 The coefficient estimate for those who are president and captain derived from a model that uses each leadership category in separate regressions (with the category "no leadership" as the baseline) is 0.027. Results are reported in Table 2B in the appendix.

4.5.5 Endogeneity adjustments based on other omitted causes

To identify whether we need to test for the susceptibility of our estimates to potential other omitted causes by applying instrumental variable regressions, we investigated whether the estimates obtained by the least squares approach in Kuhn and Weinberger (2005) are consistent. Hereby, we employ the model specifications underlying Table 4.6 and estimate several augmented OLS regressions on the leadership and earnings measures. We treat each leadership categorization variable as dependent variable in separate regressions, respectively; each model tested includes all control variables and the instruments. We therefore first estimate a regression with the leadership categorization variable as dependent variable and the control variables and instruments as explanatory variables. We then store the residuals from each regression and include the residuals as predictors in the original OLS model (with the controls but without the instruments) (Davidson, & MacKinnon, 1993). The p-values derived for captain and president were small and suggest that the OLS estimates are not consistent. The p-values for captain or president only suggest no endogeneity concerns with respect to the leadership categories subject to our chosen instruments. Similar implications derive from using the endog option when employing an extended instrumental variables/2SLS framework using ivreg2 in Stata. We therefore subsequently report IV model extensions using conditional IV estimations (that allow for robust inference in the presence of weak instruments) and heteroskedasticity-based IV estimations that make use of generated instruments to supplement the chosen instruments.

Table 4.6 contains the results from instrumental variable regression. The samples for the IV-regressions correspond to the samples from the PSM to make the results comparable. [23] We therefore employ a sample size of 22,093 when instrumenting those who were team captains and club presidents, 12,221 when instrumenting those who were only team captains, and 14,847 when instrumenting those who were only club presidents. [24]

To begin with, we first checked the suitability of our instruments. In Model 1 of Table 4.6, the weak instrument test based on the first stage F-statistic ($F(4, 22054)=6.90$; $p<0.001$) indicates very little potential for weak instruments. The Cragg-Donald Wald statistic (31.31) in comparison with the critical values provided in Stock and Yogo (2005) reveals that a potential

[23] The corresponding OLS estimates using the leadership groups and the non-leadership category as the reference group are reported in Table 2B in the appendix.

[24] The instruments are derived from the self-reports of the questionnaire respondents and cannot be corroborated by the parents. This might invite the risk of common source bias. However, because we rely on the concept of vicarious learning here, perception of the children is of importance. As such, if the children do not know about their parent's prior behavior or erroneously believe that their parents might have done something they factually did not do, this could theoretically still impact on children's behavior.

bias is relatively small (less than 5% of OLS) and also that the potential size distortion is seemingly small (less than 10% of the maximal IV size).

As it relates to Model 2, the first stage F value ($F(4, 12183)=3.09$; $p<0.05$) suggests potentially weak instruments. Correspondingly, the weak instrument test based on the F-statistic (Cragg-Donald Wald=13.06) reports a moderate potential bias (10% of OLS) and also a larger potential size distortion (more than 15% of the maximal IV size). As for model 3, the first stage F-statistic ($F(4, 14613)=3.77$; $p<0.01$) again suggests some warranted caution about instrument weakness. The tests for Model 3 reveal a small potential bias (Cragg-Donald Wald=11.27; less than 10% of OLS) but a moderate potential size distortions (more than 15% of the maximal IV size). Given the potential susceptibility of the findings to the presence of weak instruments, we therefore provide estimates also on the basis of the conditional instrumental variable regression suggested in Moreira (2003) and recommended in Andrews et al. (2019).[25]

We also examine the exclusion restriction. In assessing the exclusion restriction, we rely on the Sargan-Hansen test to see whether we find evidence for the validity of at least one of the instruments tested. To begin with, the joint null hypothesis is that the instruments are valid instruments, uncorrelated with the error term, and that the excluded instruments are correctly excluded from the estimated equation. A rejection would cast doubt on the validity of the instruments. Sargan's statistic provides evidence consistent with the validity of the instruments in all three models, under the assumption that at least one of the instruments is valid in the first place.

Regarding the coefficient estimates, we find that the coefficient for those individuals that served as president in captain is significant in Model 1 of Table 4.6 when using the conventional IV estimates and when employing the conditional IV estimator that adjusts for the presence of weak instruments. However, both coefficients for those who are presidents or captains only, turn insignificant when we compare each leadership category with the corresponding non-leader sample and instrument both variables separately in the first stage regressions. As

[25] We also follow the advice in Andrews et al. (2019) who suggest the use of the effective F-statistic based on Olea and Pflueger (2013) for errors that are potentially non-homoscedastic. The test finds sufficient support against the presence of weak instruments (Model 1: Effective F-statistic= 31.45; TSLS critical value 10%=10.52; Model 2: Effective F-statistic= 12.89; TSLS critical value 10%=10.46; Model 3: Effective F-statistic= 11.88; TSLS critical value 10%=10.19).

such, we can corroborate the earnings effect for those students that act as presidents and captains, but again find only weak (at best) evidence that having experience as either president or captain affects the log of hourly earnings. [26]

In addition, Table 4.6 includes the estimators estimator suggested in Lewbel (2012). The coefficient for those who were team captains and club presidents remains significantly positive but is considerably smaller than the one reported in the conventional IV-regression while the coefficients for captain or president only remain insignificant. We discuss the interpretation of these coefficient estimates in light of the original findings and the replication findings in the discussion section.

4.5.6 Assessing the generalizability of findings through sample extensions

To assess the generalizability of our findings from the white male sample, we extend our analyses using white females, non-white females, and non-white males. Results are shown in Table 4.7. We report the coefficient estimates for all three samples following the format of Model 4 in Table 4.2 to allow for a better comparison with the original findings. The large difference between the sample size for white males and white females derives from the fact that many white females were not in the workforce eleven years after the initial Project TALENT interview. The even higher discrepancies between the white and the non-white samples stem from the fact that the base year questionnaire did not ask students to reveal their ethnicity (Weinberger, 2014b). Nevertheless, we try to include as many non-whites as possible by also assessing the ethnicity data from the three follow-up waves.

We find that only the coefficient associated with being a captain on a team is positive and significant for non-white males ($ß=0.135$, $p<0.05$). None of the other leadership variables attains statistical significance in any model. Hence, broadly speaking we cannot find evidence to generalize the main findings from the white male sample to the other three samples.

[26] Furthermore, Table 6C in the appendix incorporates school-dummies for the conventional IV-regression. This also reduces the size of the still significantly positive coefficient of *Both President and Captain.*

Table 4.6: Conceptual replication - *Conventional IV Regression for KW sample*

	Conventional IV- Estimation			Conditional IV- Estimation			Heteroskedasticity-based IV		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Instrumented Variables									
President and Captain	0.412*			0.321*			0.163*		
	(0.181)			(0.110)			(0.0685)		
Captain Only		0.008			0.114			0.058	
		(0.236)			(0.154)			(0.112)	
President Only			0.073			-0.038			0.293†
			(0.227)			(0.115)			(0.165)
School-fixed effects	No	No	No	No	No	No	No	No	No
First stage (Kleibergen-Paap) F-Statistic	6.90	3.09	3.77	66.89	7.73	49.13	15.26	5.78	3.71
p-value (F-Statistic)	0.000	0.015	0.005	0.000	0.000	0.000	-	-	-
Cragg-Donald-Wald F-Statistic	31.31	13.06	11.27	-	-	-	117.51	37.60	15.70
Effective F-Statistics	31.45	12.90	11.88	-	-	-	-	-	-
Sargan-Hansen statistic	0.933	3.87	2.69	-	-	-	7.54	4.74	8.44
p-value (Sargan-Hansen statistic)	0.812	0.276	0.442	-	-	-	0.375	0.692	0.295
Andersen-Rubin Confidence Set	-	-	-	[.06, .65]	[-.1, .38]	[-.33, .23]	-	-	-
p-value (Andersen-Rubin)	-	-	-	0.013	0.080	0.320	-	-	-
Observations	22,093	12,222	14,652	22,093	12,222	14,652	22,093	12,222	14,652

† p<0.1, * p<.05, ** p<.01, *** p<0.01

Note: Table 4.6 reports the coefficient estimates for the dichotomous indicator variables whether individuals acted as captain and president, captain only, or president only derived from a conventional IV estimator (Stata; ivreg2). President and Captain, Captain Only and President Only are the instrumented variables. Information whether each parent was a member of a club or a team represents the (unreported) instrumental variables. Instrumented coefficients derived from instrumental variable regressions with standard errors in parentheses. Estimations for the conventional IV-regressions are derived using the ivreg2 command in Stata13, estimations for the conditional IV-regressions are based on Moreira (2003), estimations for the heteroskedasticity-based IV-regressions are based on Lewbel (2012). All models include the variables listed in Table 11 in the Appendix as (unreported) controls. The sample for the treatment analysis is based on the sample of white males employed in Kuhn and Weinberger (2005). The models do not include school-fixed effects as this would reduce the sample size. Table 6A in the online appendix contains the full first-stage model for the conventional IV-estimations. Table 6B in the online appendix contains all full second-stage models. Table 6C in the online appendix contains the conventional IV-estimations with school-fixed effects.

† p<0.1, * p<.05, ** p<.01, *** p<0.01

Table 4.7: Extension – *OLS effects of High School Leadership Activities on Hourly Earnings for expanded samples*

	Model 1 White Females	**Model 2** Non-White Males	**Model 3** Non-White Females
Leader			
Both Captain and President	0.010	0.110	0.050
	(0.016)	(0.086)	(0.075)
Captain Only	0.023	0.135*	0.173
	(0.016)	(0.069)	(0.107)
President Only	0.009	-0.012	0.047
	(0.015)	(0.075)	(0.064)
Member			
Both on Team and in Club	0.110***	0.126	-0.313*
	(0.031)	(0.160)	(0.145)
On Team only	0.012	0.199	-0.267†
	(0.076)	(0.195)	(0.146)
In Club Only	0.086**	0.173	-0.285†
	(0.031)	(0.170)	(0.149)
Controls			
Math Score	0.002***	0.002*	0.004***
	(0.000)	(0.001)	(0.001)
Parents' Education			
High School	0.018	0.031	0.050
	(0.013)	(0.072)	(0.055)
College Degree	0.024	0.160†	-0.047
	(0.017)	(0.086)	(0.106)
Educational Attainment			
Some College	0.079***	-0.014	0.152*
	(0.015)	(0.073)	(0.062)
College Degree or Higher	0.321***	0.198**	0.445***
	(0.016)	(0.075)	(0.078)
School-fixed Effects	Yes	Yes	Yes
F-Value	67.87	3.33	7.91
p > F	0.000	0.000	0.000
Adjusted R^2	0.337	0.319	0.496
Observations	11,824	747	816

† p<0.1, * p<.05, ** p<.01, *** p<0.01

Note: Table 4.7 extends the results depicted in Kuhn and Weinberger (2005: 405 (columns 5-8). The coefficients are derived from OLS regressions with standard errors in parentheses. All models include (unreported) grade and school attainment control variables as well as school dummies. Models 3 and 4 include an unreported dummy variable for lack of parent's education. Model 4 includes an unreported dummy variable for lack of educational attainment. The sample for the models labelled 1, 2, and 3 are based on the sample of white females, non-white males, and non-white females that were excluded in the original Kuhn and Weinberger (2005) study.

Subsequently, we estimate the propensity score and carry out the matching procedure for white females, non-white males and non-white females. Table 4.8 thus provides evidence on the susceptibility to endogeneity due to observed variables. Noteworthy, all three leadership variables become statistically significant in the white females sample yet the ATTs are strongly susceptible to omitted confounding variables (Rosenbaum bound values derived range from 1.02 to 1.13). Hence, while there is some evidence that leadership may affect earnings for white females, the results might be impounded strongly by unobservable confounders. Additionally, we find that none of the leadership opportunities variables is significant after matching for non-white males and only the variable indicating presidency only is significant (and fairly robust to omitted confounders) for non-white females.

Given the potential susceptibility to other omitted causes, we also report the results from instrumental variable regressions for the three different samples and the three different leadership variables in Table 4.9. Noteworthy, none of the leadership variables is statistically significant for the sample of white females. Despite some evidence for leadership effects on earnings reported previously, the coefficient estimates for the white female sample is highly susceptible to omitted confounding variables and other omitted causes. However, we do find evidence for a potential leadership effect for those who were president and captain for the sample of non-white males. Yet this effect is not robust to the correction for potentially weak instruments in the conditional and the heteroskedasticity-based instrumental variable regression (results are reported in Table 9C in the online appendix). While we find some statistically significant effects in the non-white female sample, these effects are not robust to the corrections (reported in Table 9D in the online appendix). In summary, we do not find enough evidence to consider the leadership relation with later life earnings derived from the white male sample generalizable to white females, non-white males, or non-white females.

Table 4.8: Extension - *PSM Treatment Effect Assessment for expanded samples*

	Average Treatment Effect on the Treated	Sensitivity of Estimate (Rosenbaum Bounds)
Sample: White Females		
Treatment		
President and Captain	0.022†	1.02
(N=11,210)	(0.012)	
Captain Only	0.034**	1.06
(N=6,102)	(0.014)	
President Only	0.045***	1.13
(N=6,431)	(0.013)	
Sample: Non-White Males		
Treatment		
President and Captain	0.065	1.11
(N=666)	(0.046)	
Captain Only	-0.001	1.00
(N=314)	(0.065)	
President Only	-0.027	1.00
(N=420)	(0.052)	
Sample: Non-White Females		
Treatment		
President and Captain	0.074	1.04
(N=742)	(0.046)	
Captain Only	-0.015	1.00
(N=306)	(0.086)	
President Only	0.135*	1.32
(N=438)	(0.057)	

† p<0.1, * p<.05, ** p<.01, *** p<0.01

Note: Table 4.8 reports the Average Treatment Effect on the Treated derived from Propensity Score Matching with standard errors in parentheses. Rosenbaum bounds are estimated for the respective Average Treatment Effect on the Treated for p = 0.05. The models are based on the sample of white females, non-white males, and non-white females that were excluded in the original Kuhn and Weinberger (2005) study.

Table 4.9: Extension - *Conventional IV Regression for new samples*

	Conventional IV- Estimation								
	Sample: White Females			Sample: Non-White Males			Sample: Non-White Females		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Instrumented Variables									
President and Captain	-0.046			0.860*			0.588†		
	(0.337)			(0.361)			(0.352)		
Captain Only		0.455			-2.543			0.276	
		(0.453)			(4.348)			(0.440)	
President Only			0.032			-0.392			0.404
			(0.249)			(0.311)			(0.305)
First stage (Kleibergen-Paap) F-Statistic	7.77	0.91	2.94	1.92	0.10	1.86	1.76	2.06	3.86
p-value (F-Statistic)	0.10	0.46	0.02	0.10	0.98	0.12	0.14	0.09	0.01
Cragg-Donald-Wald F-Statistic	7.44	3.51	7.01	3.76	0.16	3.98	3.95	2.32	6.15
Effective F-Statistics	7.50	3.54	7.57	3.70	0.16	3.91	3.70	2.42	6.14
Sargan-Hansen statistic	4.98	0.44	0.28	0.34	0.31	3.22	0.35	7.74	2.39
p-value (Sargan-Hansen statistic)	0.174	0.930	0.964	0.953	0.956	0.359	0.951	0.05	0.50
Observations	11,210	6,102	6,431	680	325	420	745	309	442

† p<0.1, * p<.05, ** p<.01, *** p<0.01

Note: Table 4.9 reports the coefficient estimates for the dichotomous indicator variables whether individuals acted as captain and president, captain only, or president only derived from a conventional IV estimator (Stata; ivreg2). President and Captain, Captain Only and President Only are the instrumented variables. Information whether each parent was a member of a club or a team represents the (unreported) instrumental variables. Instrumented coefficients derived from instrumental variable regressions with standard errors in parentheses. All models include the confounder variables listed in Table 11 in the Appendix as (unreported) controls. A full list of first stage coefficients is available in Table 9A in the online appendix. The sample for the treatment analysis is based on the sample of white females, non-white males, and non-white females that were excluded in the original Kuhn and Weinberger (2005) study. The models do not include school-fixed effects as this would reduce the sample size. Table 9A in the online appendix contains the full first-stage model for the Conventional IV-estimations for all three samples. Tables 9B, 9C and 9D in the online appendix contain the full second-stage models for all three IV-regression types (conventional, conditional, heteroskedasticity-based). Table 9E in the online appendix contains the same models as reported here with school-fixed effects.

4.5.7 Assessing long-term effects using the Pilot study temporal extension

Table 10 reports on how the results from the original study change after employing data from the pilot study. As the original study reported income effects for white males only, we only include white males in this analysis. Our total sample size for this analysis is only 251 observations. The following findings should therefore not be used to judge on the replicability of the original effect, but rather to provide information about the temporal generalizability of the original effects and the replicated effects reported previously.

Table 10: Extension– *OLS-Regression effects of High School Leadership Activities on 2011/2012 Household Income for White Males*

	Model 1 Household Income	**Model 2** Household Income	**Model 3** Household Income	**Model 4** Household Income
Leader				
Both Captain and President	0.478*	0.487*	0.498*	0.371†
	(0.201)	(0.195)	(0.195)	(0.203)
Captain Only	-0.171	-0.139	-0.170	-0.192
	(0.255)	(0.256)	(0.266)	(0.270)
President Only	0.016	-0.053	-0.062	-0.114
	(0.176)	(0.177)	(0.178)	(0.181)
Member				
Both on Team and in Club	0.155	-0.098	-0.070	-0.095
	(0.260)	(0.285)	(0.287)	(0.286)
On Team only	0.033	-0.150	-0.153	-0.325
	(0.324)	(0.333)	(0.306)	(0.302)
In Club Only	0.067	-0.118	-0.101	-0.151
	(0.278)	(0.297)	(0.304)	(0.300)
Controls				
Math Score		0.008**	0.008**	0.005*
		(0.003)	(0.003)	(0.003)
Parent´s Education				
High School			0.025	0.070
			(0.187)	(0.185)
College Degree			0.350	0.313
			(0.237)	(0.234)
Educational Attainment				
Some College				0.172
				(0.251)
College Degree or Higher				0.514*
				(0.250)
School-fixed Effects	Yes	Yes	Yes	Yes
F-statistic	1.37	2.10	1.67	1.71
p > F	0.211	0.031	0.078	0.053
Observations	251	251	251	251
Adjusted R^2	0.083	0.117	0.119	0.132

† p<0.1, * p<.05, ** p<.01, *** p<0.01

Note: Table 10 extends the results from Kuhn and Weinberger (2005) Table 2 by employing household income as the dependent variable. The sample is based on the 2011-12 Pilot Study and only includes white males. Coefficients are derived from OLS regressions with standard errors in parentheses. All models include (unreported) grade and school attainment control variables. Models 3 and 4 include an unreported dummy variable for lack of parent's education. Model 4 includes an unreported dummy variable for lack of educational attainment. Results using an ordered logit approach and using the female sample are reported in the online appendix.

We find that the main effects for those who were granted leadership opportunities as both, president and captain is significantly positive (at least at the 10 percent level). Yet the effects reported for those individuals that reported to have acted as either president or captain only are insignificant. While we would not argue that this generally attests to the robustness of the original findings (given the small sample size), we certainly think that it attests to the generalizability of the main findings derived from the extended analyses. Having leadership experience as captain and president is beneficial and not only affects the log of the earnings some 11 years after high-school but also aligns with household income as reported in the 2011-2012 Pilot Study.[27]

4.6 Discussion

Is there such a thing as leadership skill? Do some individuals have the attributes that determine whether or not they are granted leadership opportunities, and are leadership skills developed because of the granting of leadership opportunities and because leaders practice leadership very early in life? To delineate the answers to these questions, we replicated and extended a seminal study by Kuhn and Weinberger (2005) using data from project TALENT. To begin with, we are able to exactly replicate the original findings by Kuhn and Weinberger (2005) up to the third digit, which is rare in replication studies (Ebersole et al., 2016; Klein et al., 2018). Our results support the notion that for white males a component of leadership skills is developed through the granting of leadership positions in high school.

Subsequently, we extended the analyses controlling for observable omitted selection and other unobservable omitted causes into leadership opportunities using propensity score matching and instrumental variable techniques. Our PSM results document that the original effects, reported for those who were captains and presidents, are robust to the presence of endogeneity. Subsequently, we also report in our instrumental variable regressions that these very coefficient estimates are also not susceptible to selection on unobservable factors. Yet the credibility of the IV results relies on the assumption that the instruments are valid; even though various tests corroborate this assumption, we cannot completely exclude the possibility that the large effects estimated by the IV regressions might be caused by an exclusion restriction failure. Noticeably, the coefficient estimates for the IV estimator are larger than the estimates derived

[27] Ordered-logistic regressions estimated on only the sample of white males for the 2011-2012 Pilot Study data return the same results (Table 10A in the appendix). Tables 10B and 10C in the appendix show that we cannot corroborate these implications for white females. Tables 10D and 10E show that the implications stay the same when we include *Living Alone*, a dummy variable capturing whether respondents lived alone at the time of the pilot study.

from the matching analysis and the results from the OLS regressions. Yet the inclusion of school-fixed effects in the IV-regressions and employing heteroskedasticity-based instruments to supplement the standard IV estimator reduce the size of the coefficients.

While the findings of our extended analyses provide support for the notion that being granted both team captaincy and club presidency increases subsequent wages for white males, we do not find evidence that having served as captain or president solely, affects subsequent wages. Therefore, we show that only a multiplicity of leadership positions held (as both team captain and president of a club) is associated with higher wages in later working life. Importantly, we report that the leadership skills developed due to the experience of being granted team captaincy and club presidency serve as a signal for later life earnings using an extended 50-year time horizon. As such, early life leadership skills are predictive of later life earnings. Noteworthy, we find that these effects only materialize for white males and are not robustly present in our analyses focusing on females and non-whites males.

4.6.1 Implications for future research

Our results have several implications for leadership theory and practice. First, our findings highlight the need for a life-course analysis of leadership skills development. Leadership biographies offer numerous accounts of early life leadership opportunities (Cunha et al., 2017). Following Reitan and Stenberg (2019), learning experiences and leadership opportunities granted materialize differently depending on when they occur. Understanding why and how leaders emerge might be better explained when a longer time horizon is applied to gain an understanding of when leadership skills are developed. This is especially important because exposure to leadership opportunities and the timing of these vary across individuals, in particular along gender and ethnicities (Fitzsimmons et al., 2017). Hence, future research on leadership skill development should investigate traits and behaviors that affect leadership exposure and selection as early as possible. In fact, being granted leadership roles might not only start in high school but already in primary school or even kindergarten (Bennett, & Derevensky, 1995).

Second, our results also emphasize the skill variety that makes for successful leaders. Lazear (2012) argued that effective leaders must solve a wide variety of technical, financial, organizational, and inter-personal problems that require diverse experience and skills. Frederiksen and Kato (2017) found that those individuals that occupied different roles in their careers were more likely to be appointed to top management positions. Our findings corroborate these notions and show that the granting of leadership opportunities in sports combined with another more cognitive-oriented domain is related to higher earnings subsequently, but neither leadership opportunity alone affects earnings. Hence, effective leadership skills developed in early life rests upon a variety of leadership experiences, and not a single source of leadership experience gained. It would therefore be very interesting to differentiate between different types of clubs and sports teams to assess whether there exist differences in the extent of leadership skill development.

Third, our findings also attest to the long-lasting impact of leadership selection in early life. Hambrick and Wowak (2012) addressed how changes in institutions bring about the fundamental characteristics of individuals who subsequently lead companies. They document a shift towards more individualistic and materialistic leaders chosen over the course of the 80s and 90s. Research along these lines could further illumine how early life leader selection affects the type of leaders (in management) that emerge subsequently. Future studies could investigate further to what extent selection processes in early life lead to a systematic favoritism of certain types of potential leaders (i.e., those individuals that possess the perceived features that are typically associated with leadership ability). This is especially important because potential future followers decide on whom they choose as prospective leaders (Bastardoz, & van Vugt, 2019). As it pertains to the school program participants granting other students an opportunity to lead, some of the variables included in the matching model (military or business scores, sociability, or overweight) influence whether individuals lead as team captains and/or club presidents and these variables simultaneously affect their earnings. This certainly provides grounds to better understand the type of managers that was ushered in because certain individuals that do not possess the corresponding traits drop out of the leader selection tournaments as early as in high school (if not earlier, e.g. Reitan, & Stenberg, 2019). Evidently, longitudinal and multi-country analyses could help to understand how the perception of leader prototypicality has changed and resulted in certain archetypes of leaders being developed over the different decades.

Fourth, it is important to note that we fail to generalize the effect of early life leadership experience on earnings for females and non-white males. While there is some evidence that (certain) leadership skills might relate to higher earnings for females, the effects vanish once we control for selection based on other omitted causes. Along these lines, it is important to be cognizant of the fact that developing a leader identity is important for the development of leadership skills. Individuals become aware of the fact that leaders exist and develop aspirations to become leaders themselves. In childhood and adolescence, individuals recognize that certain individuals have authority and leadership skills and might start to follow into their footsteps by experimenting with leadership positions (Reitan, & Stenberg, 2019). Along the development trajectory of leadership skills, mentors play a significant role. They act as role models and impact on the self-confidence of potential leaders (Fitzsimmons et al., 2017). Yet the number of potential female and non-white role models and mentors might be limited in the early life phase studied in our sample (Fitzsimmons et al., 2017; Hayes, 2008; Harper, 2018; Thomas, 2001). It would therefore be worthwhile to investigate further, how leadership experience translates into leadership skills conditional on mentoring and role modelling relations. Furthermore, engaging in new cohort studies that are similar to the Project TALENT could increase our understanding of leadership skill development and income effects for females and non-white males. This could help to further contextualize our findings as the share of female and non-white leaders has substantially increased since the 1980s (Heckman et al., 2017). Likewise, employing cohort study data from countries with smaller gender pay gaps like Scandinavia might unveil whether the leadership effects for males also apply to females if social policies reduce gender-based differences not only in labor markets but also in early child education (Polachek, & Xiang, 2009).

4.6.2 Implications for leadership development

Developing a leader identity and fostering leadership aspirations early is important as it has very long-lasting impact on the career prospects of high school students. White male leaders often equate their leadership training with integrity, courage, and self-efficacy (Fitzsimmons et al., 2017). As such, girls should be encouraged to lead early in life to increase exposure to leadership experience and the individual perceptions and ascriptions thereof. This is especially important, as earnings effects of leadership opportunities for males carry value beyond the initial 11-year horizon and are evident over the extended time horizon studied.

Initiatives in companies that aim to develop leadership skills might be more effective and lead to a more inclusive environment if they focus on early career females and non-white males and help to overcome the absence of leadership experiences in early life. This essentially could create confidence and willingness to take on subsequent leadership opportunities throughout their careers. For companies, it is important to recognize the leadership skills that are already developed and to identify those areas that are in need of development. This is important, because leadership ascriptions for males, females, and non-white males take place in different contexts and at different points in life. Simultaneously, talented females and non-white males need to be given access to mentors and role models early in their careers and possibly as early as high school, so that respective students could relate and develop their leadership self-image (Johnson, & Eby, 2011; Dougherty et al., 2013).

As it concerns women in particular, gender stereotypes might also be among the unobserved factors that explain why leadership opportunities in high school might not transpire into higher earnings subsequently. As a case in point, male CEOs predominantly report stay-at-homes wives with domestic responsibility while female CEOs report dual roles of working and shouldering domestic responsibility. Problematically, Stertz et al. (2017) emphasize the importance of couple dynamics for female careers. Men's gender-role attitudes explain women's work-involvement decisions after childbirth. If women are married to men with less egalitarian attitudes, they are less likely to increase their working hours after childbirth. For women aspiring careers and leadership roles, discussing and planning around domestic responsibility is important. Beyond the personal nature of these discussions, companies can disburden women of the dual role by facilitating childcare and facilitating workplace returns after maternity leaves (Wiese, & Ritter, 2012). This essentially might groom and increase the pool of potential female leaders.

Last, our results regarding the development of leadership skills through multiple sources of leadership experience also possess implications for school administrators and intendents to not only develop athletic programs but also rather complement these programs with other more cognitive-oriented programs in which students can strive (and vice versa). Arguably, developing future leaders along a multitude of dimensions is becoming more and more crucial. For employers and employees, the importance of non-cognitive skills associated with leadership skills (e.g. conflict resolution, communication, and the ability to integrate oneself into a team (Bacolod et al., 2010; Borghans et al., 2014)) is increasing with the economy-wide digitization of work (Deming, 2017). As a case in point, Bartel et al. (2007) show how technological changes induce adjustments of human resource policies in firms and the required skill profiles of employees. As such, equipping students with leadership skills drawn from a variety of experiences (including for example IT leadership, not covered in our data) to master these challenges becomes paramount.

4.6.3 Limitations

The results and implications of our replication and extension of "Leadership Skills and Wages" specifically have to be considered within the bounds of the limitations of our study. Besides the fact, that the Project TALENT represents "the largest and most comprehensive study of high school students in the history of the United States." (American Institutes for Research, 2016) many institutional settings have changed quite significantly over the recent decades. The base year in 1960 allows making life-course analyses of the initial respondents, but it would clearly be worthwhile to explore other cohorts that underwent early-life leadership training in subsequent decades. Curricula shifts and technological advancements might have granted leadership opportunities in areas such programing or desktop publishing; leadership skills in these areas might have gained in relevance with a focus on computerized work and the rise of the Internet.

As it concerns the generalizability of our findings, our samples for the analyses of non-whites are considerably smaller than for whites. This derives from the fact that the base year Project TALENT questionnaire did not ask respondents to reveal their ethnicity. Racial data can therefore only be drawn from the three follow-up waves. In these follow-up waves, whites represent an overwhelming majority (about 90%) causing some concerns of sample truncation, due to non-whites responding less often in the follow-up surveys (Weinberger 2014b). To deal with this issue, we employed Project TALENT weights considering the non-response bias for the 11-years follow up survey. Still, some selection effects exist as also the wage data contains missing observations among slightly more than 10% of the 11-years follow up study respondents. If, for some reason, there is selection on other unobservable variables (motivations or statistical discrimination), which correlates with the disturbance in the outcome, this would necessitate the use of sample selection models that can overcome the truncation in the dependent variable (Heckman, 1979; Killingsworth, & Heckman, 1986).

Furthermore, large demographic and societal shifts have occurred since the conduction of Project TALENT. As a case in point, the Civil Rights Act prohibiting racial segregation and ensuring same pay for same work came into effect in 1964 (Brown, 1984). Weinberger (2014b) points out that a substantial number of the high schools participating in the Project TALENT were segregated, which essentially could limit access to leadership opportunities. For non-whites, leadership opportunities as well as job development chances have substantially improved over the course of the past sixty years, which again would call for additional analyses using subsequent cohorts to further document the generalizability of leadership effects found in the original work and our replication (Hayes, 2008). These societal changes also apply to gender differences with the gender pay gap shrinking considerably (Bowler, 1999). Similarly, since female leadership in sports, but also in many other fields, has become more common, it could be possible that we might also find effects for women when investigating recent data (e.g. Yiamouyiannis, & Osborne, 2012; Chao, & Tian, 2011; Dormody, & Seevers, 1994). All in all, we currently miss a recent large scaled high school student cohort study that would allow us to test whether the implications derived from data gathered up to sixty years ago for white males also applies to non-white and female students attending high school nowadays. Along these lines, it is important to note that a new follow-up survey is currently underway that could be explored (Stone, et al. 2014).

4.7 Conclusion

Leadership is "a developmental journey that is a lifelong process" (Fitzsimmons et al., 2014: 245). Our study shows that adolescence is a foundational time of critical leadership development that equips individuals with leadership skills that provides benefits throughout an individual's career. Being granted a leadership position as both team captain and club president in high school increases individual earnings and even transpires into higher household income 50 years later. The effects however, only materialize for a large sample of white males, but not for females and non-white males. Similarly, certain characteristics of leader selection are already present as early as high school, lending support to the notion that potential future followers decide on whom they choose as prospective leaders and accept their influence and initiative taking already in early life (Bastardoz, & van Vugt, 2019; Antonakis et al., 2016; Reitan, & Stenberg, 2019). Our results support a life-course analysis of leadership development that highlights when leadership skills are developed. Leadership interventions should recognize the leadership skills that are already developed in individuals and identify those areas that are in need of further development. In doing so, it is important to be cognizant of diverse experiences with leadership and thus different aspirations to lead for individuals of different gender and ethnicities.

4.8 Appendix

The labelling (number followed by a letter) of all tables in the Appendix corresponds to the number of the related tables in the manuscript (e.g. Table 2A shows the same results as Table 4.2 but with robust standard errors). The corresponding letter denotes the different tables that were created. The only exception is Table 11 which stands on its own.

4.8.1 Technical appendix with model specifications

4.8.1.1 Exact Replication

We replicate the original results (see Kuhn, & Weinberger, 2005: 405; Table 2) using the log of hourly earnings as our dependent variable. Our sample size consists of 24,041 white men. This equals the sample size of Kuhn and Weinberger (2005). We begin our replication analysis by estimating an OLS model of the following form:

$$(1)\ Y_i = \alpha + \beta X_i + \mu_i$$

In (1), the index i (i=1,…N) denotes all individual students, X denotes all explanatory variables used, α is the intercept, and μ_i is the error term. We employ a model using school fixed effects. As in Kuhn and Weinberger (2005), the model is restricted to white males who earn between \$1 and \$50 per hour.

4.8.1.2 Observable Omitted Selection

Following the logic in Kuhn and Weinberger (2005), we assume that there is a binary set of options: either individuals are granted leadership opportunities, or they are not. Following Abadie et al. (2004) and Rubin (1974), let the two options denote D=1 for leaders and D=0 for non-leaders. Students with D=1 are considered "treated." The hourly earnings in subsequent occupations eleven years after high school is denoted Y_i for each individual. Hence, for an individual i, the earnings $Y_i(1)$ would be observed if D=1 was observed in high school. Alternatively, the earnings $Y_i(0)$ would be observed if D=0 was observed in high school. Based on these considerations, we want to estimate the average treatment effect on the treated, as shown in (2).

$$(2)\ \tau_{ATT} = E[Y(1)\ |D=1] - E[Y(0)\ |D=1]$$

However, we cannot observe the counterfactual of the hourly earnings for the treated individuals E[Y(0) |D=1], as the same individual cannot be both leader and non-leader in high school. Therefore, the counterfactual needs to be replaced by an estimate. A common estimate is the propensity score (the conditional predicted probability summarizing how likely each individual is to receive the treatment) (Rubin, 1974; Li, 2013). On a more general note, prior work (Rosenbaum, & Rubin, 1983) showed that one-to-one matching in large data sets is infeasible. Yet dimensionality problems can be overcome by conditioning on balancing scores such as the propensity score.

To compute the propensity score, $\hat{P}(x)$, we estimate a probit model with an individual's probability of being granted leadership opportunities high school as the dependent variable (Rosenbaum and Rubin, 1983). These equations should condition on all potential observable variables that could affect hourly earnings and leadership selection. Controlling for all variables is, of course, infeasible, but we try to include as many variables (observable in the dataset) as possible. The propensity score is equal to the conditional probability of receiving the treatment (being a leader). We estimate the propensity score as shown in (3). Hereby, $\hat{P}[D=1|x]$ denotes the probability of being a leader in high-school, Φ is the cumulative distribution function of a standard normal distribution, α is the intercept, and x and ß represent the input variables and the corresponding coefficients (Li, 2013).

$$(3)\quad \hat{P}[D=1|x] = \Phi\ (\alpha + x\ ß)$$

After calculating the propensity score, subjects with the same probability of being granted leadership opportunities are treated as statistical twins. After matching, given a set of observable characteristics, subjects are identical in all observable aspects except for receiving the treatment. The basic idea behind making causal inferences based on the propensity score matching procedure is that if two subjects have the same probability of receiving treatment (the same propensity score), yet are in different groups (leaders or non-leaders), we are, statistically speaking, comparing two individuals who were exogenously assigned to the leadership and non-leadership groups. Hence, after matching, we can directly infer the differences between the matched groups, as (4) shows. This allows us to estimate the net effect of the treatment on outcomes, much like in experiments that randomize the treatment assignment (Andersen and Lu, 2016).

$$(4)\quad \tau_{PSM} = E_{\hat{P}(x)D=1}\{E[Y(1)\ |D=1, \hat{P}(x)] - E[Y(0)\ |D=0, \hat{P}(x)]$$

with $\hat{P}(x)$ denoting the predicted probability of treatment using the variables in x.

4.8.1.3 Other unobservable omitted causes

Employing instrumental variable techniques is a strategy for the consistent estimation of causal effects between leadership positions and later-life earnings. In the corresponding first stage depicted in (5), the endogenous regressor (leadership variable) is regressed on the instruments and covariates (Bascle, 2008).

$$(5)\ X_i = \gamma_0 + \gamma_1 Z_i + \gamma_2 W_i + \upsilon_i$$

In (5), *Xi* is the *i*th observation of the endogenous explanatory variable and *Wi* represents the *i*th observation for each of the exogenous regressors (control variables). In comparison to the previous regression estimates employed in the exact replication above, we employ a set of variables Z_i that are correlated with the endogenous component of the leadership variables X_i but not with the error term of the outcome equation μ_i. The first stage therefore isolates the variation in X_i that is not correlated with u_i (error term of the outcome regression). The corresponding instrumented (fitted) value of the endogenous variable X_i is used in the second stage ($\hat{X}_i$) in lieu of the endogenous regressor, as (6) shows.

$$(6)\ Y_i = \alpha + ß_1 \hat{X}_i + ß_2 W_{i+\ldots} + \mu_i$$

As instruments, we rely on information about prior parental behavior. We suggest that if parents were members of clubs or teams, children would engage in similar behaviors, which in turn, might increase their chances of being granted leadership opportunities in similar teams or clubs. We therefore employ the information provided about whether each parent was a member of a club or a team as the four instrumental variables Z_1 to Z_4.

4.8.2 Information on variables

Table 11: Variables for the Propensity Score Matching and IV-regressions

Outcome Variable	
Natural Logarithm of Hourly Earnings	Average hourly pay of respondents; data from the 11-year follow-up study
Treatment Variables	
Captain and President	Dummy variable indicating whether respondents were team captains and club presidents in the three years leading up to the survey date
Captain only	Dummy variable indicating whether respondents were team captains but not club presidents in the three years leading up to the survey date
President only	Dummy variable indicating whether respondents were club presidents but not team captains in the three years leading up to the survey date
Confounder Variables	
Personal Data	
Team and Club Member	Dummy variable indicating whether respondents were team and club members in the three years leading up to the survey date
On Team only	Dummy variable indicating whether respondents were team members but not club members in the three years leading up to the survey date
In Club only	Dummy variable indicating whether respondents were club members but not team members in the three years leading up to the survey date
Overweight	Dummy variable indicating whether respondents possess a BMI above 31
Underweight	Dummy variable indicating whether respondents possess a BMI below or equal to 17
Tall	Dummy variable indicating whether respondents are 72 inches or more tall
Short	Dummy variable indicating whether respondents are 65 inches tall or less
Dates	Respondents' average number of dates per week
Comfortable	Dummy variable indicating whether the finances of respondents' families afford comfortable lifestyles
Wealthy	Dummy variable indicating whether the finances of respondents' families afford wealthy lifestyles
Personal Characteristics	
Sociability	Standardized score of the degree to which respondents enjoy socializing
Vigor	Standardized score of the degree to which respondents are physically energetic
Mature personality	Standardized score of the degree to which respondents are hardworking and reliable
Self-Confidence	Standardized score of the degree of self-confidence respondents possess
Tidiness	Standardized score of the degree to which respondents are neat and organized
Test Scores	
Math Score	Relative score in math test within same grade
Vocabulary Score	Relative score in vocabulary test within same grade
Social Studies Score	Relative score in social studies test within same grade

Science Score	Relative score in science test within same grade
Scientific Attitude Score	Relative score in scientific problem-solving test within same grade
Law Score	Relative score in law test within same grade
Military Score	Relative score in military test within same grade
Business Score	Relative score in business test within same grade
Etiquette Score	Relative score in etiquette test within same grade
English Score	Relative score in English test within same grade
Awards	
Science Awards	Number of science awards respondents won in the last three years
Fine Arts Awards	Number of fine arts awards respondents won in the last three years
Sports Awards	Number of sports awards respondents won in the last three years
Cognitive Skills	
Arithmetic Skills	Relative scoring in calculation test under time pressure
Reading Skills	Relative scoring in reading test under time pressure
Clerical Skills	Relative scoring in identifying spelling mistakes under time pressure
Identification Skills	Relative scoring in identifying differences in objects under time pressure
Personal Interests	
Public Service Interest	Standardized score of the degree to which respondents are interested in gaining a job in public services (e.g. mayor, governor, …)
Business Management Interest	Standardized score of the degree to which respondents are interested in gaining a job in business management (e.g. executive, foreman,…)

Note: Grey background indicates variables included in Kuhn and Weinberger (2005), either as dependent, independent, or control variables.

Table 1A: Full correlation matrix for KW sample

	Log (Hourly Earnings)	Both Captain and President	Captain Only	President Only	Both on Team and in Club	On Team Only	In Club Only	Math Score
Log (Hourly Earnings)	1.0000							
Both Captain and President	0.0703***	1.0000						
Captain Only	0.0079	-0.1965***	1.0000					
President Only	0.0162*	-0.2872***	-0.2121***	1.0000				
Both on Team and in Club	0.0866***	0.1707***	0.1102***	0.0504***	1.0000			
On Team Only	-0.0067	-0.0423***	0.0124†	-0.0511***	-0.2850***	1.0000		
In Club Only	-0.0791***	-0.1485***	-0.1102***	-0.0134*	-0.8764***	-0.0765***	1.0000	
Math Score	0.2197***	0.0193**	-0.0278***	0.1009***	0.0840***	-0.0382***	-0.0497***	1.0000
Parent High School	0.0300***	0.0175**	-0.0023	0.0043	0.0297***	-0.0029	-0.0226***	0.0387***
Parent College	0.0646***	0.0315***	-0.0132*	0.0578***	0.0717***	-0.0409***	-0.0442***	0.2300***
Some College	-0.0437***	-0.0014	-0.0026	-0.0144*	-0.0123†	0.0116†	0.0074	-0.0765***
Degree or Higher	0.1927***	0.0579***	-0.0105	0.0906***	0.1022***	-0.0486***	-0.0634***	0.4731***
Overweight	-0.0318***	-0.0258***	0.0025	-0.0026	-0.0212**	0.0070	0.0141*	-0.0479***
Underweight	-0.0144*	-0.0260***	-0.0045	-0.0059	-0.0548***	-0.0059	0.0589***	-0.0165*
Tall	-0.0008	0.0061	0.0154*	0.0098	0.0405***	-0.0010	-0.0364***	0.0352***
Short	-0.0173*	-0.0112†	-0.0161*	-0.0208**	-0.0494***	0.0021	0.0475***	-0.0611***
Dates	0.0029	0.1018***	0.0196**	0.0239***	0.1199***	-0.0392***	-0.0996***	-0.1450***
Comfortable	-0.0057	-0.0067	-0.0150*	0.0061	0.0050	-0.0057	0.0015	0.0522***
Wealthy	0.0663***	0.0330***	0.0125†	0.0199**	0.0480***	-0.0238***	-0.0353***	0.0235***
Sociability	0.0864***	0.1312***	0.0342***	0.0768***	0.1900***	-0.0487***	-0.1533***	-0.0081
Vigor	0.0779***	0.1583***	0.0498***	0.0484***	0.2185***	-0.0305***	-0.1918***	0.0872***
Mature	0.0993***	0.1038***	-0.0218**	0.1083***	0.0739***	-0.0601***	-0.0323***	0.2192***
Self-Confi-dence	0.0831***	0.0816***	-0.0081	0.0997***	0.1002***	-0.0417***	-0.0720***	0.1474***
Tidiness	0.0832***	0.0762***	-0.0008	0.0737***	0.0718***	-0.0497***	-0.0356***	0.0951***
Vocabulary Score	0.1656***	-0.0142*	-0.0497***	0.0998***	0.0233***	-0.0405***	0.0060	0.6710***
Social Studies	0.1534***	-0.0108	-0.0310***	0.0858***	0.0478***	-0.0206**	-0.0287***	0.6151***

Score								
Science Score	0.1485***	-0.0185**	-0.0562***	0.0993***	0.0163*	-0.0373***	0.0119†	0.6923***
Scientific Attitude Score	0.1414***	0.0017	-0.0261***	0.0681***	0.0372***	-0.0092	-0.0219**	0.4653***
Law Score	0.1331***	-0.0066	-0.0340***	0.0685***	0.0239***	-0.0336***	-0.0025	0.4719***
Military Score	0.1142***	0.0227***	-0.0246***	0.0594***	0.0817***	-0.0300***	-0.0619***	0.3585***
Business Score	0.1476***	-0.0073	-0.0409***	0.0785***	0.0061	-0.0346***	0.0191**	0.4537***
Etiquette Score	0.0580***	0.0109	-0.0304***	0.0399***	0.0027	-0.0196**	0.0113†	0.2283***
English Score	0.1487***	0.0158*	-0.0346***	0.1050***	0.0459***	-0.0266***	-0.0185**	0.6989***
Science Awards	0.0249***	0.0587***	-0.0134*	0.0293***	0.0386***	-0.0275***	-0.0250***	0.0675***
Fine Arts Awards	0.0032	0.0628***	-0.0205**	0.0478***	0.0462***	-0.0341***	-0.0258***	0.0037
Sports Awards	0.0553***	0.2185***	0.0630***	0.0308***	0.2619***	-0.0384***	-0.2350***	0.0592***
Arithmetic Skills	0.0765***	-0.0094	-0.0098	0.0282***	0.0197**	-0.0014	-0.0130†	0.3598***
Reading Skills	0.0040	-0.0148*	-0.0049	0.0070	-0.0148*	0.0119†	0.0101	-0.0037
Clerical Skills	0.0762***	0.0133*	-0.0062	0.0408***	0.0419***	-0.0075	-0.0355***	0.2544***
Identification Skills	0.0351***	-0.0213**	-0.0136*	0.0068	-0.0148*	0.0063	0.0101	0.1438***
Public Service Interest	0.1068***	0.0882***	0.0058	0.0729***	0.1298***	-0.0546***	-0.0918***	0.1525***
Business Management Interest	0.0891***	0.0847***	0.0210**	0.0489***	0.1247***	-0.0524***	-0.0881***	0.0555***

	Parent High School	Parent College	Some College	Degree or Higher	Overweight	Underweight	Tall	Short
Parent High School	1.0000							
Parent College	-0.5014***	1.0000						
Some College	0.0560***	-0.0747***	1.0000					
Degree or Higher	0.0010	0.2309***	-0.4465***	1.0000				
Overweight	0.0004	-0.0236***	0.0035	-0.0285***	1.0000			

Underweight	-0.0022	0.0033	-0.0058	-0.0072	-0.0168*	1.0000		
Tall	-0.0011	0.0431***	-0.0023	0.0272***	-0.0712***	0.0547***	1.0000	
Short	-0.0263***	-0.0129†	0.0075	-0.0392***	0.1211***	0.0555***	-0.1967***	1.0000
Dates	-0.0015	-0.0089	0.0147*	-0.1030***	-0.0272***	-0.0494***	0.0401***	-0.0819***
Comfortable	0.0634***	-0.0793***	0.0109	0.0233***	-0.0169*	-0.0051	0.0080	-0.0222***
Wealthy	-0.0613***	0.1840***	-0.0123†	0.0609***	0.0042	0.0041	0.0191**	0.0023
Sociability	0.0128†	0.0362***	0.0017	0.0560***	-0.0205**	-0.0407***	0.0014	-0.0262***
Vigor	0.0225***	0.0504***	-0.0120†	0.1002***	-0.0469***	-0.0296***	0.0437***	-0.0540***
Mature	0.0064	0.0888***	-0.0699***	0.1930***	-0.0188**	-0.0181**	0.0250***	-0.0474***
Self-Confidence	0.0032	0.0785***	-0.0174**	0.1159***	-0.0152*	-0.0284***	0.0366***	-0.0486***
Tidiness	0.0169*	0.0438***	-0.0208**	0.1227***	-0.0279***	-0.0041	0.0070	-0.0328***
Vocabulary Score	0.0426***	0.2209***	-0.0434***	0.3871***	-0.0364***	-0.0060	0.0519***	-0.0595***
Social Studies Score	0.0475***	0.1884***	-0.0619***	0.4025***	-0.0287***	-0.0163*	0.0462***	-0.0555***
Science Score	0.0342***	0.2020***	-0.0401***	0.3716***	-0.0319***	-0.0080	0.0387***	-0.0496***
Scientific Attitude Score	0.0272***	0.1422***	-0.0345***	0.2637***	-0.0308***	-0.0143*	0.0467***	-0.0573***
Law Score	0.0263***	0.1593***	-0.0269***	0.2899***	-0.0249***	-0.0138*	0.0462***	-0.0544***
Military Score	0.0311***	0.1504***	-0.0062	0.2491***	-0.0097	-0.0116†	0.0390***	-0.0539***
Business Score	0.0382***	0.1336***	-0.0334***	0.2731***	-0.0165*	-0.0047	0.0399***	-0.0402***
Etiquette Score	0.0120†	0.0670***	-0.0144*	0.1192***	-0.0086	-0.0030	0.0097	-0.0007
English Score	0.0251***	0.2021***	-0.0803***	0.4242***	-0.0436***	-0.0152*	0.0206**	-0.0542***
Science Awards	-0.0007	0.0327***	-0.0274***	0.0500***	0.0105	0.0129†	0.0139*	0.0166*
Fine Arts Awards	-0.0040	0.0519***	-0.0148*	0.0355***	0.0172*	0.0124†	0.0124†	0.0098
Sports Awards	0.0279***	0.0989***	-0.0092	0.1020***	-0.0070	-0.0255***	0.0657***	-0.0543***
Arithmetic Skills	0.0135*	0.0696***	-0.0413***	0.1880***	-0.0318***	-0.0098	0.0114†	-0.0340***
Reading Skills	0.0153*	-0.0077	0.0213**	-0.0117†	-0.0152*	-0.0052	-0.0020	-0.0227***
Clerical Skills	0.0204**	0.0462***	-0.0194**	0.1445***	-0.0192**	-0.0132†	0.0092	-0.0170*
Identification Skills	0.0180**	0.0142*	0.0004	0.0463***	-0.0209**	0.0000	0.0085	-0.0185**

Public Service Interest	0.0070	0.0675***	-0.0298***	0.1879***	-0.0074	-0.0080	0.0176**	-0.0071
Business Management Interest	0.0197**	0.0139*	-0.0050	0.0996***	-0.0140*	-0.0142*	0.0069	-0.0143*

	Dates	Comfortable	Wealthy	Sociability	Vigor	Mature	Self-Confidence	Tidiness
Dates	1.0000							
Comfortable	-0.0172*	1.0000						
Wealthy	0.0551***	-0.6797***	1.0000					
Sociability	0.2121***	-0.0102	0.0748***	1.0000				
Vigor	0.1130***	-0.0017	0.0563***	0.4873***	1.0000			
Mature	0.0571***	-0.0055	0.0570***	0.3687***	0.5291***	1.0000		
Self-Confidence	0.1084***	0.0116†	0.0545***	0.3519***	0.3132***	0.4093***	1.0000	
Tidiness	0.0875***	-0.0063	0.0843***	0.3651***	0.4121***	0.6123***	0.2825***	1.0000
Vocabulary Score	-0.1263***	0.0692***	0.0235***	-0.0340***	0.0525***	0.1692***	0.1443***	0.0729***
Social Studies Score	-0.1545***	0.0709***	-0.0006	-0.0400***	0.0576***	0.1504***	0.1165***	0.0564***
Science Score	-0.1550***	0.0608***	0.0117†	-0.0729***	0.0441***	0.1573***	0.1211***	0.0426***
Scientific Attitude Score	-0.0691***	0.0508***	-0.0010	-0.0077	0.0439***	0.1105***	0.1046***	0.0166*
Law Score	-0.0906***	0.0673***	0.0024	-0.0114†	0.0375***	0.1269***	0.1217***	0.0440***
Military Score	-0.0477***	0.0503***	0.0157*	0.0134*	0.0555***	0.0986***	0.1117***	0.0441***
Business Score	-0.0783***	0.0595***	0.0005	0.0017	0.0317***	0.1472***	0.1094***	0.0698***
Etiquette Score	-0.0225***	0.0143*	0.0170*	0.0207**	0.0212**	0.0664***	0.0539***	0.0531***
English Score	-0.1300***	0.0606***	0.0038	0.0147*	0.0786***	0.2005***	0.1435***	0.1364***
Science Awards	0.0255***	-0.0273***	0.0524***	0.0198**	0.0428***	0.0817***	0.0542***	0.0390***
Fine Arts Awards	0.0530***	-0.0246***	0.0496***	0.0568***	0.0518***	0.0712***	0.0594***	0.0486***
Sports Awards	0.1329***	-0.0130†	0.0598***	0.1731***	0.2356***	0.1282***	0.1328***	0.1039***
Arithmetic Skills	-0.0589***	0.0352***	-0.0111†	0.0077	0.0250***	0.0777***	0.0553***	0.0565***

Reading Skills	0.0085	0.0140*	-0.0058	0.0032	-0.0124†	-0.0177**	0.0029	0.0281***
Clerical Skills	-0.0221**	0.0351***	-0.0012	0.0392***	0.0546***	0.0758***	0.0770***	0.0779***
Identification Skills	-0.0116†	0.0233***	-0.0148*	-0.0376***	-0.0306***	-0.0111†	0.0125†	-0.0010
Public Service Interest	0.0282***	-0.0091	0.0487***	0.1758***	0.1258***	0.1631***	0.1270***	0.1212***
Business Management Interest	0.0972***	-0.0197**	0.0424***	0.2275***	0.1259***	0.1485***	0.0866***	0.1385***

	Vocabulary Score	Social Studies Score	Science Score	Scientific Attitude Score	Law Score	Military Score	Business Score	Etiquette Score
Vocabulary Score	1.0000							
Social Studies Score	0.6466***	1.0000						
Science Score	0.6860***	0.6286***	1.0000					
Scientific Attitude Score	0.4641***	0.4139***	0.4246***	1.0000				
Law Score	0.5312***	0.5262***	0.4744***	0.3652***	1.0000			
Military Score	0.4381***	0.4705***	0.3722***	0.2574***	0.3991***	1.0000		
Business Score	0.5135***	0.4745***	0.4217***	0.3522***	0.4483***	0.3521***	1.0000	
Etiquette Score	0.2402***	0.2062***	0.2100***	0.1269***	0.1640***	0.1370***	0.1823***	1.0000
English Score	0.6126***	0.5642***	0.5500***	0.4177***	0.4344***	0.3358***	0.4173***	0.2483***
Science Awards	0.0435***	0.0330***	0.0679***	0.0214**	0.0236***	0.0222***	0.0264***	0.0293***
Fine Arts Awards	0.0101	0.0037	0.0072	0.0001	-0.0028	0.0049	-0.0112†	0.0383***
Sports Awards	0.0212**	0.0403***	0.0274***	0.0184**	0.0192**	0.0769***	0.0037	0.0050
Arithmetic Skills	0.2373***	0.2464***	0.2295***	0.1786***	0.1707***	0.1263***	0.1662***	0.0790***
Reading Skills	0.0056	-0.0073	-0.0183**	0.0003	0.0062	0.0205**	0.0043	0.0056
Clerical Skills	0.2193***	0.2010***	0.1826***	0.1463***	0.1579***	0.1317***	0.1556***	0.1141***
Identification Skills	0.1190***	0.0926***	0.1024***	0.0920***	0.0788***	0.0683***	0.0767***	0.0542***

Public Service Interest	0.1364***	0.2077***	0.0946***	0.0927***	0.1641***	0.1588***	0.1609***	0.0508***
Business Management Interest	0.0148*	0.0426***	-0.0293***	0.0264***	0.0469***	0.0464***	0.0863***	0.0181**

	English Score	Science Awards	Fine Arts Awards	Sports Awards	Arithmetic Skills	Reading Skills	Clerical Skills	Identification Skills
English Score	1.0000							
Science Awards	0.0231***	1.0000						
Fine Arts Awards	0.0179**	0.6072***	1.0000					
Sports Awards	0.0205**	0.2271***	0.2813***	1.0000				
Arithmetic Skills	0.3492***	-0.0064	-0.0253***	0.0074	1.0000			
Reading Skills	0.0293***	-0.0350***	-0.0317***	-0.0091	0.0614***	1.0000		
Clerical Skills	0.3270***	-0.0060	-0.0026	0.0213**	0.2117***	0.1778***	1.0000	
Identification Skills	0.1528***	-0.0131†	-0.0235***	-0.0062	0.1542***	0.1235***	0.1704***	1.0000
Public Service Interest	0.1287***	0.0374***	0.0385***	0.0955***	0.0526***	-0.0114†	0.0604***	-0.0220**
Business Management Interest	0.0479***	0.0122†	0.0189**	0.0787***	0.0276***	0.0046	0.0316***	-0.0216**

	Public Service Interest	Business Management Interest
Public Service Interest	1.0000	
Business Management Interest	0.6355***	1.0000

† $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Note: Pairwise correlation coefficients derived from Pearson-correlations. N. of obs. is 20,095.

Table 1B: Correlation matrix for white females

	Log (Hourly Earnings)	Both Captain and President	Captain Only	President Only	Both on Team and in Club	On Team Only	In Club Only	Math Score
Log (Hourly Earnings)	1.0000							
Both Captain and President	0.0481***	1.0000						
Captain Only	-0.0043	-0.2693***	1.0000					
President Only	0.0417***	-0.2948***	-0.2555***	1.0000				
Both on Team and in Club	0.0651***	0.2244***	0.1575***	-0.0873***	1.0000			
On Team Only	-0.0143	-0.0190*	0.0315***	-0.0336***	-0.0919***	1.0000		
In Club Only	-0.0544***	-0.2140***	-0.1612***	0.1018***	-0.9629***	-0.0583***	1.0000	
Math Score	0.3233***	0.0208*	-0.0155	0.0651***	0.0650***	-0.0327***	-0.0378***	1.0000
Parent High School	0.0160†	0.0266**	0.0162†	0.0174†	0.0251**	-0.0116	-0.0173†	0.0664***
Parent College	0.1468***	0.0354***	-0.0199*	0.0234*	0.0537***	-0.0115	-0.0429***	0.1948***
Some College	-0.0512***	0.0283**	-0.0005	0.0067	0.0337***	-0.0107	-0.0254**	0.0055
Degree or Higher	0.3960***	0.0343***	-0.0297**	0.0700***	0.0224*	-0.0267**	-0.0054	0.4615***
Overweight	-0.0242*	-0.0212*	-0.0128	-0.0160†	-0.0327***	0.0096	0.0216*	-0.0242*
Underweight	0.0059	-0.0183†	-0.0164†	0.0034	-0.0213*	-0.0063	0.0183†	-0.0086
Tall	0.0094	0.0017	-0.0033	0.0043	0.0203*	-0.0042	-0.0184†	-0.0079
Short	-0.0115	-0.0122	-0.0136	0.0092	-0.0290**	0.0071	0.0276**	-0.0364***
Dates	-0.1365***	0.0848***	-0.0001	0.0315***	0.0711***	-0.0246**	-0.0609***	-0.1640***
Comfortable	0.0197*	-0.0057	-0.0080	0.0182†	-0.0128	-0.0026	0.0157†	0.0524***
Wealthy	0.0552***	0.0408***	0.0046	0.0050	0.0529***	-0.0004	-0.0499***	0.0104
Sociability	0.0190*	0.1288***	0.0149	0.0525***	0.1455***	-0.0320***	-0.1271***	-0.0580***
Vigor	0.0526***	0.1744***	0.0490***	0.0225*	0.2295***	-0.0325***	-0.2058***	0.0609***
Mature	0.1408***	0.0926***	-0.0227*	0.0806***	0.0737***	-0.0327***	-0.0575***	0.2626***
Self-Confi-dence	0.0731***	0.1173***	-0.0145	0.0734***	0.1199***	-0.0257**	-0.1026***	0.1107***
Tidiness	0.0359***	0.0405***	-0.0033	0.0490***	0.0248**	-0.0105	-0.0160†	0.0305**
Vocabulary Score	0.2792***	0.0128	-0.0329***	0.0593***	0.0271**	-0.0229*	-0.0059	0.6512***
Social Studies	0.2703***	0.0151	-0.0316***	0.0429***	0.0179†	-0.0198*	0.0011	0.6021***

Score								
Science Score	0.2389***	0.0183†	-0.0140	0.0514***	0.0446***	-0.0202*	-0.0251**	0.6319***
Scientific Attitude Score	0.1795***	0.0167†	0.0019	0.0214*	0.0103	-0.0200*	0.0064	0.4451***
Law Score	0.1606***	-0.0012	0.0006	0.0158†	0.0075	-0.0002	0.0026	0.3994***
Military Score	0.1337***	0.0059	-0.0176†	0.0328***	0.0162†	-0.0259**	-0.0031	0.3202***
Business Score	0.1951***	0.0046	-0.0132	0.0259**	0.0240*	-0.0056	-0.0113	0.4105***
Etiquette Score	0.1457***	0.0280**	-0.0198*	0.0350***	0.0108	-0.0339***	0.0012	0.3310***
English Score	0.2405***	0.0080	-0.0368***	0.0680***	0.0116	-0.0395***	0.0137	0.6408***
Science Awards	0.0271**	0.0280**	-0.0137	0.0050	0.0163†	0.0044	-0.0181†	0.0495***
Fine Arts Awards	0.0293**	0.0844***	-0.0317***	0.0462***	0.0366***	-0.0116	-0.0301**	0.0733***
Sports Awards	0.0356***	0.1753***	0.0280**	-0.0132	0.1982***	-0.0219*	-0.1859***	0.0364***
Arithmetic Skills	0.0861***	-0.0286**	0.0056	0.0122	-0.0064	-0.0124	0.0146	0.3165***
Reading Skills	-0.0300**	-0.0285**	-0.0066	-0.0127	-0.0248**	0.0082	0.0189*	-0.0748***
Clerical Skills	0.0847***	0.0096	-0.0124	0.0013	0.0134	-0.0098	-0.0086	0.1891***
Identification Skills	0.0504***	-0.0092	0.0065	-0.0003	0.0137	0.0046	-0.0113	0.1548***
Public Service Interest	0.1535***	0.0599***	-0.0282**	0.0449***	0.0853***	-0.0276**	-0.0708***	0.1875***
Business Management Interest	0.0733***	0.0478***	-0.0016	0.0469***	0.0775***	-0.0403***	-0.0614***	0.0871***

	Parent High School	Parent College	Some College	Degree or Higher	Overweight	Underweight	Tall	Short
Parent High School	1.0000							
Parent College	-0.4651***	1.0000						
Some College	0.0311***	-0.0028	1.0000					
Degree or Higher	-0.0025	0.2460***	-0.3532***	1.0000				
Overweight	-0.0079	-0.0363***	0.0044	-0.0276**	1.0000			

Underweight	-0.0034	-0.0092	-0.0036	-0.0032	-0.0296**	1.0000		
Tall	-0.0042	-0.0033	0.0067	-0.0029	-0.0073	0.0586***	1.0000	
Short	-0.0090	-0.0264**	-0.0118	-0.0290**	0.0401***	0.0771***	-0.0831***	1.0000
Dates	0.0159†	-0.0514***	-0.0076	-0.1850***	-0.0714***	-0.0136	-0.0098	0.0593***
Comfortable	0.0511***	-0.0223*	-0.0080	0.0249**	-0.0216*	-0.0032	-0.0075	-0.0059
Wealthy	-0.0375***	0.1377***	0.0284**	0.0584***	-0.0137	0.0030	-0.0080	-0.0101
Sociability	0.0342***	0.0099	0.0292**	-0.0338***	-0.0445***	-0.0226*	-0.0025	0.0129
Vigor	0.0392***	0.0518***	0.0168†	0.0537***	-0.0556***	-0.0212*	0.0064	-0.0342***
Mature	0.0369***	0.0633***	-0.0270**	0.1913***	-0.0252**	-0.0137	0.0082	-0.0323***
Self-Confidence	0.0132	0.0682***	0.0053	0.0902***	-0.0144	-0.0253**	0.0012	-0.0126
Tidiness	0.0449***	-0.0108	0.0064	0.0226*	-0.0505***	0.0109	0.0090	-0.0039
Vocabulary Score	0.0776***	0.1884***	0.0269**	0.3900***	-0.0226*	-0.0165†	-0.0183†	-0.0423***
Social Studies Score	0.0593***	0.1693***	0.0175†	0.3850***	-0.0101	-0.0264**	-0.0124	-0.0437***
Science Score	0.0645***	0.1684***	0.0113	0.3571***	-0.0123	-0.0029	-0.0061	-0.0400***
Scientific Attitude Score	0.0506***	0.1230***	0.0180†	0.2530***	-0.0343***	-0.0056	-0.0195*	-0.0231*
Law Score	0.0406***	0.1191***	0.0141	0.2452***	-0.0135	0.0004	-0.0206*	-0.0379***
Military Score	0.0503***	0.1134***	0.0247**	0.2039***	-0.0129	-0.0210*	0.0003	-0.0371***
Business Score	0.0554***	0.0850***	0.0027	0.2260***	-0.0164†	-0.0166†	-0.0163†	-0.0219*
Etiquette Score	0.0575***	0.0988***	0.0094	0.2000***	-0.0505***	-0.0059	-0.0124	-0.0120
English Score	0.0503***	0.1635***	-0.0006	0.3573***	-0.0479***	0.0010	-0.0236*	-0.0419***
Science Awards	-0.0079	0.0225*	-0.0227*	0.0496***	-0.0074	0.0205*	0.0358***	-0.0068
Fine Arts Awards	0.0207*	0.0630***	-0.0048	0.0865***	-0.0103	0.0101	0.0240*	-0.0203*
Sports Awards	0.0356***	0.0566***	0.0311***	0.0444***	-0.0186*	-0.0055	0.0167†	-0.0304**
Arithmetic Skills	0.0258**	0.0233*	-0.0101	0.1355***	-0.0003	0.0089	-0.0052	-0.0071
Reading Skills	0.0022	-0.0220*	0.0188*	-0.0666***	-0.0078	0.0223*	-0.0062	0.0045
Clerical Skills	0.0202*	0.0267**	0.0017	0.0937***	-0.0269**	0.0118	-0.0177†	-0.0056
Identification Skills	0.0269**	0.0258**	0.0019	0.0550***	-0.0175†	-0.0089	-0.0117	-0.0131

Public Service Interest	0.0167†	0.0604***	0.0056	0.1881***	-0.0199*	-0.0120	-0.0110	-0.0025
Business Management Interest	0.0096	-0.0012	-0.0052	0.0713***	-0.0237*	0.0001	-0.0043	0.0071

	Dates	Comfortable	Wealthy	Sociability	Vigor	Mature	Self-Confidence	Tidiness
Dates	1.0000							
Comfortable	-0.0056	1.0000						
Wealthy	0.0344***	-0.6302***	1.0000					
Sociability	0.2234***	0.0137	0.0580***	1.0000				
Vigor	0.1260***	-0.0029	0.0621***	0.5002***	1.0000			
Mature	0.0181†	0.0167†	0.0373***	0.2994***	0.4350***	1.0000		
Self-Confidence	0.1182***	0.0207*	0.0536***	0.4074***	0.3447***	0.3997***	1.0000	
Tidiness	0.0982***	0.0205*	0.0523***	0.2943***	0.3064***	0.5679***	0.2297***	1.0000
Vocabulary Score	-0.1504***	0.0660***	0.0002	-0.1003***	0.0149	0.2038***	0.1005***	-0.0059
Social Studies Score	-0.1930***	0.0569***	-0.0000	-0.1221***	-0.0017	0.1866***	0.0777***	-0.0173†
Science Score	-0.1577***	0.0488***	-0.0048	-0.0960***	0.0361***	0.1996***	0.0873***	-0.0170†
Scientific Attitude Score	-0.0622***	0.0407***	-0.0030	-0.0511***	0.0156†	0.1260***	0.0726***	-0.0258**
Law Score	-0.1039***	0.0583***	-0.0254**	-0.0918***	-0.0152	0.1157***	0.0695***	-0.0492***
Military Score	-0.0746***	0.0371***	0.0053	-0.0516***	0.0020	0.1006***	0.0638***	-0.0148
Business Score	-0.0804***	0.0525***	-0.0139	-0.0557***	-0.0249**	0.1384***	0.0685***	-0.0044
Etiquette Score	-0.0352***	0.0308**	0.0166†	-0.0081	0.0076	0.1128***	0.0625***	0.0434***
English Score	-0.1064***	0.0634***	-0.0167†	-0.0490***	0.0235*	0.2219***	0.0848***	0.0580***
Science Awards	0.0013	-0.0078	0.0211*	0.0132	0.0225*	0.0601***	0.0322***	0.0161†
Fine Arts Awards	0.0539***	-0.0031	0.0428***	0.0647***	0.0977***	0.1338***	0.0996***	0.0598***
Sports Awards	0.0796***	-0.0076	0.0542***	0.1203***	0.1957***	0.1194***	0.1091***	0.0626***
Arithmetic Skills	-0.0528***	0.0291**	-0.0268**	-0.0254**	-0.0120	0.0888***	0.0335***	0.0229*

Reading Skills	0.0179†	0.0101	-0.0111	-0.0181†	-0.0267**	-0.0669***	-0.0315***	-0.0058
Clerical Skills	-0.0095	0.0093	0.0094	0.0077	0.0189*	0.0720***	0.0402***	0.0392***
Identification Skills	-0.0130	0.0145	-0.0163†	-0.0373***	-0.0330***	-0.0107	-0.0173†	-0.0145
Public Service Interest	-0.0415***	0.0052	0.0354***	0.0790***	0.1164***	0.1659***	0.1397***	0.0367***
Business Management Interest	0.0602***	0.0045	0.0131	0.1510***	0.1220***	0.1651***	0.1295***	0.0898***

	Vocabulary Score	Social Studies Score	Science Score	Scientific Attitude Score	Law Score	Military Score	Business Score	Etiquette Score
Vocabulary Score	1.0000							
Social Studies Score	0.6625***	1.0000						
Science Score	0.6539***	0.6047***	1.0000					
Scientific Attitude Score	0.4623***	0.3999***	0.3842***	1.0000				
Law Score	0.4922***	0.4787***	0.4066***	0.3138***	1.0000			
Military Score	0.3860***	0.3848***	0.3232***	0.2087***	0.3249***	1.0000		
Business Score	0.4923***	0.4503***	0.3719***	0.3149***	0.3753***	0.2940***	1.0000	
Etiquette Score	0.3973***	0.3216***	0.3093***	0.2286***	0.2516***	0.2200***	0.2711***	1.0000
English Score	0.6109***	0.5337***	0.4953***	0.4021***	0.3799***	0.2863***	0.3918***	0.3795***
Science Awards	0.0439***	0.0314***	0.0595***	0.0295**	0.0281**	0.0133	0.0263**	0.0285**
Fine Arts Awards	0.0716***	0.0600***	0.0882***	0.0557***	0.0487***	0.0345***	0.0392***	0.0699***
Sports Awards	0.0197*	0.0264**	0.0451***	0.0036	-0.0014	0.0055	0.0118	0.0021
Arithmetic Skills	0.2297***	0.2109***	0.1932***	0.1524***	0.1302***	0.0927***	0.1665***	0.1336***
Reading Skills	-0.0406***	-0.0558***	-0.0490***	-0.0229*	-0.0056	-0.0305**	-0.0178†	-0.0092
Clerical Skills	0.1830***	0.1524***	0.1192***	0.1244***	0.1144***	0.0876***	0.1274***	0.1350***
Identification Skills	0.1404***	0.1103***	0.1117***	0.1229***	0.0786***	0.0601***	0.0882***	0.0977***

Public Service Interest	0.1815***	0.2164***	0.1598***	0.1016***	0.1652***	0.1439***	0.1597***	0.0803***
Business Management Interest	0.0578***	0.0785***	0.0462***	0.0583***	0.0743***	0.0535***	0.1086***	0.0268**

	English Score	Science Awards	Fine Arts Awards	Sports Awards	Arithmetic Skills	Reading Skills	Clerical Skills	Identification Skills
English Score	1.0000							
Science Awards	0.0183†	1.0000						
Fine Arts Awards	0.0742***	0.5925***	1.0000					
Sports Awards	-0.0262**	0.2845***	0.3562***	1.0000				
Arithmetic Skills	0.3029***	0.0030	0.0023	-0.0260**	1.0000			
Reading Skills	-0.0239*	-0.0359***	-0.0446***	-0.0319***	0.0302**	1.0000		
Clerical Skills	0.2700***	-0.0127	-0.0005	-0.0089	0.1681***	0.1204***	1.0000	
Identification Skills	0.1769***	-0.0055	-0.0105	-0.0244**	0.1440***	0.0839***	0.1355***	1.0000
Public Service Interest	0.1182***	0.0298**	0.0639***	0.0634***	0.0247**	-0.0333***	0.0295**	0.0037
Business Management Interest	0.0697***	-0.0075	0.0195*	0.0408***	0.0164†	-0.0215*	0.0326***	-0.0081

	Public Service Interest	Business Management Interest
Public Service Interest	1.0000	
Business Management Interest	0.6049***	1.0000

† $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Note: Pairwise correlation coefficients derived from Pearson-correlations. N. of obs. is 11,210.

Table 1C: Correlation matrix for non-white males

	Log (Hourly Earnings)	Both Captain and President	Captain Only	President Only	Both on Team and in Club	On Team Only	In Club Only	Math Score
Log (Hourly Earnings)	1.0000							
Both Captain and President	0.0539	1.0000						
Captain Only	0.0578	-0.2266***	1.0000					
President Only	-0.0286	-0.3528***	-0.2436***	1.0000				
Both on Team and in Club	0.1009**	0.1947***	0.0940*	-0.0184	1.0000			
On Team Only	0.0180	-0.0831*	0.0638†	-0.0197	-0.2749***	1.0000		
In Club Only	-0.1078**	-0.1557***	-0.1089**	0.0247	-0.8954***	-0.0685†	1.0000	
Math Score	0.3490***	-0.0558	0.0123	0.0339	0.1028**	0.0381	-0.1228**	1.0000
Parent High School	0.1098**	0.0682†	-0.0264	0.0257	0.1106**	-0.0017	-0.1183**	0.1177**
Parent College	0.1400***	0.1065**	-0.0463	-0.0272	0.0460	0.0019	-0.0347	0.1897***
Some College	0.0101	0.0968*	0.0453	-0.0112	0.0257	0.0304	-0.0416	0.0120
Degree or Higher	0.2417***	-0.0081	-0.0518	0.0618	0.0737†	-0.0556	-0.0484	0.3910***
Overweight	0.0270	-0.0628	-0.0301	0.0420	-0.0421	-0.0218	0.0587	-0.0299
Underweight	0.0019	0.0149	-0.0773*	0.0022	0.0462	-0.0283	-0.0315	0.0057
Tall	-0.0042	0.1212**	-0.0170	-0.0581	0.1268***	-0.0598	-0.0978*	-0.0447
Short	-0.0161	-0.0672†	0.0326	-0.0779*	-0.0532	0.0850*	0.0032	0.0323
Dates	-0.0630	0.0867*	-0.0107	0.0560	0.1129**	-0.0819*	-0.0651†	-0.2337***
Comfortable	0.0921*	-0.0438	0.0060	-0.0090	-0.0018	0.0186	-0.0065	0.1392***
Wealthy	0.0184	0.0254	0.0055	-0.0305	0.0144	-0.0138	-0.0124	-0.1182**
Sociability	0.0933*	0.0426	0.0673†	-0.0028	0.1526***	-0.1383***	-0.0844*	0.0543
Vigor	0.0883*	0.1940***	0.0235	-0.0364	0.1923***	-0.0678†	-0.1617***	0.0262
Mature	0.0694†	0.1165**	0.0216	0.0978*	0.0959*	-0.0682†	-0.0517	0.0569
Self-Confi-dence	0.1186**	0.0222	-0.0047	0.1108**	0.0533	-0.0882*	-0.0180	0.1309***
Tidiness	0.1165**	0.0827*	-0.0075	0.0962*	0.0769*	-0.1065**	-0.0262	0.0704†
Vocabulary Score	0.3088***	-0.0583	0.0281	0.0280	0.0446	-0.0065	-0.0398	0.6751***
Social Studies	0.3215***	-0.0596	0.0489	0.0356	0.0766*	0.0164	-0.0780*	0.6523***

Score								
Science Score	0.2748***	-0.0439	0.0233	0.0187	0.0177	0.0503	-0.0225	0.6862***
Scientific Attitude Score	0.2154***	-0.0508	0.0014	0.0207	0.0240	-0.0337	-0.0195	0.5270***
Law Score	0.2758***	-0.0526	0.0529	0.0079	0.0753*	0.0032	-0.0736†	0.5486***
Military Score	0.2271***	0.0021	0.0188	-0.0132	0.0314	-0.0193	-0.0199	0.4226***
Business Score	0.2089***	0.0375	0.0016	0.0244	0.0815*	0.0175	-0.0822*	0.4900***
Etiquette Score	0.0480	0.0214	0.0483	0.0114	-0.0137	0.0303	0.0127	0.1233**
English Score	0.2714***	-0.0858*	0.0032	0.0812*	0.0341	0.0753*	-0.0649†	0.6973***
Science Awards	-0.0151	0.1321***	-0.0294	0.0118	0.0300	-0.0131	-0.0243	-0.0809*
Fine Arts Awards	-0.0639†	0.1522***	-0.0662†	-0.0243	0.0826*	-0.0485	-0.0562	-0.1352***
Sports Awards	0.0532	0.2572***	0.1070**	-0.0691†	0.2598***	-0.0650†	-0.2254***	0.1005**
Arithmetic Skills	0.1008**	-0.1074**	0.0143	0.0247	0.0542	-0.0141	-0.0758*	0.4297***
Reading Skills	0.1058**	-0.0353	0.0175	0.0045	0.0496	-0.0532	-0.0402	0.1856***
Clerical Skills	0.0816*	0.0642†	-0.0319	-0.0196	0.0243	0.0043	-0.0211	0.2284***
Identification Skills	0.1162**	-0.0198	-0.0264	-0.0239	0.0759*	0.0141	-0.1035**	0.2220***
Public Service Interest	-0.0299	0.0997**	-0.0809*	0.0500	0.0918*	-0.0533	-0.0629	-0.0864*
Business Management Interest	-0.0941*	0.1200**	-0.0337	0.0431	0.1265***	-0.0516	-0.0966*	-0.1866***

	Parent High School	Parent College	Some College	Degree or Higher	Overweight	Underweight	Tall	Short
Parent High School	1.0000							
Parent College	-0.2567***	1.0000						
Some College	0.0483	-0.0665†	1.0000					
Degree or Higher	0.1125**	0.2346***	-0.4135***	1.0000				
Overweight	0.0580	0.0269	0.0007	0.0036	1.0000			

Underweight	0.0296	0.0350	0.0245	-0.0343	-0.0293	1.0000		
Tall	-0.0351	0.0884*	0.0075	0.0278	-0.0620	0.0080	1.0000	
Short	-0.0144	-0.0220	-0.0423	0.0138	0.1461***	0.0330	-0.2365***	1.0000
Dates	-0.0273	-0.0650†	0.0145	-0.1812***	-0.0072	0.0023	0.0779*	-0.1703***
Comfortable	0.0739†	0.0219	-0.0020	0.0708†	0.0063	0.0398	-0.0031	-0.0369
Wealthy	0.0125	0.0681†	0.0409	-0.0270	0.0342	-0.0099	0.0352	-0.0101
Sociability	0.0449	0.0512	0.0148	0.0460	-0.0147	0.0291	0.0594	-0.1222**
Vigor	0.0649†	-0.0322	0.0727†	0.0114	-0.0034	-0.0452	0.0999**	-0.1394***
Mature	0.0002	0.0465	-0.0284	0.0714†	-0.0024	-0.0235	0.1127**	-0.1549***
Self-Confidence	0.0363	0.1063**	-0.0290	0.1418***	-0.0388	0.0066	0.0825*	-0.1243**
Tidiness	0.0005	0.0800*	-0.0037	0.1014**	0.0152	-0.0347	0.0884*	-0.1301***
Vocabulary Score	0.1409***	0.1800***	0.0520	0.3784***	-0.0274	-0.0426	-0.0734†	0.0529
Social Studies Score	0.1088**	0.1236**	0.0162	0.3878***	-0.0245	-0.0347	-0.0874*	0.0671†
Science Score	0.0721†	0.1821***	0.0209	0.3738***	-0.0191	-0.0607	-0.1043**	0.0816*
Scientific Attitude Score	0.1145**	0.1141**	0.0481	0.2499***	-0.0053	0.0298	-0.0753*	0.0072
Law Score	0.0649†	0.1681***	0.0118	0.3320***	-0.0557	-0.0240	-0.0471	0.0187
Military Score	0.1034**	0.1498***	0.0436	0.2708***	-0.0314	-0.0370	-0.0696†	0.0499
Business Score	0.1004**	0.1344***	0.0029	0.2903***	-0.0103	0.0050	-0.0544	-0.0689†
Etiquette Score	-0.0118	0.1046**	-0.0265	0.1355***	0.0342	0.0086	-0.0088	0.0055
English Score	0.1439***	0.1607***	0.0287	0.4008***	-0.0704†	-0.0348	-0.0971*	0.0410
Science Awards	-0.0230	0.0125	-0.0091	-0.0503	0.0623	0.0154	0.1539***	-0.0281
Fine Arts Awards	0.0204	0.0069	-0.0124	-0.0475	-0.0088	0.0294	0.1428***	-0.0794*
Sports Awards	0.1140**	0.0360	0.0527	0.0877*	-0.0287	-0.0352	0.0769*	-0.0587
Arithmetic Skills	0.0481	0.0752*	0.0362	0.2084***	0.0337	-0.0107	-0.1056**	0.0865*
Reading Skills	0.0541	0.0489	0.0234	0.0326	-0.0160	-0.0475	0.0053	-0.0076
Clerical Skills	0.0948*	0.0540	0.0626	0.1744***	-0.0134	-0.0459	-0.0265	-0.0058
Identification Skills	0.0086	0.0448	0.0354	0.0909*	-0.0276	0.0462	0.0359	-0.0531

Public Service Interest	-0.0435	0.0307	-0.0498	0.0203	0.0294	-0.0219	0.0617	-0.0657†
Business Management Interest	-0.0496	-0.0513	0.0324	-0.0882*	-0.0088	0.0011	0.0637†	-0.0861*

	Dates	Comfortable	Wealthy	Sociability	Vigor	Mature	Self-Confidence	Tidiness
Dates	1.0000							
Comfortable	-0.0930*	1.0000						
Wealthy	0.1403***	-0.5488***	1.0000					
Sociability	0.1416***	0.0791*	-0.0412	1.0000				
Vigor	0.1467***	-0.0268	0.0502	0.4774***	1.0000			
Mature	0.0997**	-0.0504	0.0021	0.4209***	0.5407***	1.0000		
Self-Confidence	0.1207**	0.0597	-0.0471	0.3407***	0.2243***	0.3398***	1.0000	
Tidiness	0.1136**	-0.0035	-0.0069	0.4287***	0.4447***	0.6323***	0.3081***	1.0000
Vocabulary Score	-0.2283***	0.1767***	-0.1224**	0.0777*	0.0623	0.1049**	0.1750***	0.1345***
Social Studies Score	-0.2542***	0.2005***	-0.1297***	0.0495	0.0232	0.0569	0.1076**	0.0797*
Science Score	-0.2306***	0.1316***	-0.0993**	0.0071	0.0092	0.0350	0.0585	0.0538
Scientific Attitude Score	-0.1692***	0.1686***	-0.1375***	0.0794*	0.0464	0.0616	0.1101**	0.0923*
Law Score	-0.1841***	0.1345***	-0.1184**	0.0205	0.0522	0.0336	0.1026**	0.0363
Military Score	-0.1886***	0.0860*	-0.0155	0.0577	0.0128	0.0250	0.0645†	0.0307
Business Score	-0.1477***	0.1086**	-0.0884*	0.0924*	0.0819*	0.0887*	0.1118**	0.0706†
Etiquette Score	-0.0334	0.0462	-0.0415	0.0763*	0.0800*	0.0878*	0.0608	0.0980*
English Score	-0.2705***	0.1519***	-0.1309***	0.0936*	0.0055	0.0729†	0.1716***	0.1113**
Science Awards	0.1013**	-0.1336***	0.1149**	-0.0219	0.0845*	0.0900*	0.0196	0.0215
Fine Arts Awards	0.0810*	-0.0816*	0.0964*	0.0327	0.1128**	0.0449	0.0239	0.0231
Sports Awards	0.0703†	-0.0209	0.0177	0.2084***	0.2228***	0.1651***	0.0969*	0.1121**
Arithmetic Skills	-0.2103***	0.1073**	-0.1007**	0.1071**	0.0254	-0.0195	0.0444	0.0124

Reading Skills	-0.0568	0.0340	-0.0016	-0.0282	-0.0094	-0.0457	0.0325	-0.0088
Clerical Skills	-0.0437	0.0643†	-0.0037	0.0329	0.0161	-0.0709†	0.0710†	0.0612
Identification Skills	0.0245	0.0531	-0.1094**	0.0190	-0.0245	0.0032	0.1332***	0.0535
Public Service Interest	0.1104**	-0.0145	-0.0094	0.1355***	0.1071**	0.1636***	0.0350	0.0949*
Business Management Interest	0.2015***	-0.0560	0.0086	0.1954***	0.1539***	0.1846***	0.0520	0.1459***

	Vocabulary Score	Social Studies Score	Science Score	Scientific Attitude Score	Law Score	Military Score	Business Score	Etiquette Score
Vocabulary Score	1.0000							
Social Studies Score	0.7228***	1.0000						
Science Score	0.7284***	0.7227***	1.0000					
Scientific Attitude Score	0.6101***	0.5543***	0.5320***	1.0000				
Law Score	0.5889***	0.6165***	0.5498***	0.4502***	1.0000			
Military Score	0.4808***	0.5333***	0.4886***	0.3561***	0.4321***	1.0000		
Business Score	0.5624***	0.5159***	0.4983***	0.4508***	0.5106***	0.3331***	1.0000	
Etiquette Score	0.1731***	0.1468***	0.1513***	0.1337***	0.1560***	0.1350***	0.1065**	1.0000
English Score	0.7032***	0.6713***	0.6268***	0.5517***	0.5171***	0.3990***	0.5120***	0.1596***
Science Awards	-0.0840*	-0.0619	-0.0471	-0.0999**	-0.0712†	-0.0657†	-0.0389	0.0297
Fine Arts Awards	-0.1358***	-0.1178**	-0.1133**	-0.1500***	-0.1084**	-0.0338	-0.0573	-0.0193
Sports Awards	0.0834*	0.1082**	0.0846*	-0.0088	0.0753*	0.1363***	0.0857*	0.0008
Arithmetic Skills	0.3754***	0.3697***	0.3410***	0.3094***	0.2724***	0.1959***	0.2960***	0.0957*
Reading Skills	0.1916***	0.1819***	0.1416***	0.1680***	0.1951***	0.0874*	0.1376***	0.0662†
Clerical Skills	0.2868***	0.2700***	0.2195***	0.2167***	0.1762***	0.1683***	0.1912***	0.1083**
Identification Skills	0.2213***	0.1766***	0.1855***	0.1913***	0.1632***	0.0760*	0.1458***	0.0139

Public Service Interest	-0.1203**	-0.1142**	-0.1553***	-0.0878*	-0.0349	-0.0695†	-0.0580	-0.0178
Business Management Interest	-0.2113***	-0.2068***	-0.2591***	-0.1295***	-0.1297***	-0.1151**	-0.0752†	-0.0365

	English Score	Science Awards	Fine Arts Awards	Sports Awards	Arithmetic Skills	Reading Skills	Clerical Skills	Identification Skills
English Score	1.0000							
Science Awards	-0.1537***	1.0000						
Fine Arts Awards	-0.1687***	0.5893***	1.0000					
Sports Awards	0.0417	0.2539***	0.2893***	1.0000				
Arithmetic Skills	0.4910***	-0.1325***	-0.1525***	-0.0167	1.0000			
Reading Skills	0.2243***	-0.0762*	-0.0435	-0.0057	0.2440***	1.0000		
Clerical Skills	0.3177***	-0.0950*	-0.0503	0.0291	0.2679***	0.2373***	1.0000	
Identification Skills	0.2621***	-0.0900*	-0.0782*	-0.0347	0.2713***	0.1959***	0.2699***	1.0000
Public Service Interest	-0.1161**	0.0349	0.0587	0.0378	-0.0442	-0.0443	0.0180	-0.0512
Business Management Interest	-0.1857***	0.0442	0.0661†	0.0357	-0.0845*	-0.0703†	-0.0345	-0.0530

	Public Service Interest	Business Management Interest
Public Service Interest	1.0000	
Business Management Interest	0.6942***	1.0000

† p < 0.1; * p < 0.05; ** p < 0.01; *** p < 0.001

Note: Pairwise correlation coefficients derived from Pearson-correlations. N. of obs. is 680.

Table 1D: Correlation matrix for non-white females

	Log (Hourly Earnings)	Both Captain and President	Captain Only	President Only	Both on Team and in Club	On Team Only	In Club Only	Math Score
Log (Hourly Earnings)	1.0000							
Both Captain and President	0.0477	1.0000						
Captain Only	0.0029	-0.2456***	1.0000					
President Only	0.0326	-0.4116***	-0.2871***	1.0000				
Both on Team and in Club	-0.0133	0.1862***	0.1267***	-0.0178	1.0000			
On Team Only	-0.0260	-0.0377	0.0336	0.0464	-0.0834*	1.0000		
In Club Only	0.0213	-0.1741***	-0.1341***	0.0215	-0.9770***	-0.0474	1.0000	
Math Score	0.3849***	-0.0037	0.0319	-0.0331	-0.0362	-0.0117	0.0469	1.0000
Parent High School	0.1556***	0.0210	0.0490	0.0037	0.0673†	0.0503	-0.0713†	0.1718***
Parent College	0.1539***	0.0455	-0.0120	0.0185	0.0578	-0.0234	-0.0499	0.1646***
Some College	-0.0211	0.0718†	-0.0300	0.0005	0.1208***	0.0169	-0.1230***	0.0140
Degree or Higher	0.4833***	0.0219	-0.0698†	0.0481	-0.0989**	-0.0364	0.1124**	0.3305***
Overweight	-0.0811*	-0.0161	-0.0381	-0.0321	-0.0806*	-0.0141	0.0861*	-0.0440
Underweight	0.0407	-0.0089	0.0277	-0.0398	0.0334	-0.0145	-0.0285	0.0638†
Tall	-0.0270	0.0401	-0.0304	-0.0510	0.0179	-0.0047	-0.0164	-0.0441
Short	0.0385	-0.0028	-0.0483	-0.0234	-0.0778*	-0.0194	0.0737*	0.0422
Dates	-0.1867***	0.0587	0.0378	0.0393	0.0938*	0.0817*	-0.1103**	-0.2024***
Comfortable	0.0641†	0.0407	-0.0012	-0.0124	0.0333	0.0066	-0.0281	0.1186**
Wealthy	-0.0427	-0.0033	0.0006	0.0129	0.0401	-0.0263	-0.0469	-0.1601***
Sociability	-0.0160	0.1209***	-0.0354	0.0311	0.1409***	-0.0236	-0.1283***	0.0148
Vigor	0.0472	0.1131**	-0.0046	0.0756*	0.1866***	-0.0047	-0.1747***	-0.0204
Mature	0.1238***	0.1035**	-0.0562	0.0698†	0.1000**	-0.0131	-0.0915*	0.0705†
Self-Confidence	0.0605†	0.0895*	0.0139	-0.0161	0.1304***	0.0384	-0.1368***	0.0697†
Tidiness	0.0204	0.1063**	0.0200	0.0426	0.1486***	0.0145	-0.1493***	-0.0241
Vocabulary Score	0.3494***	-0.0217	0.0550	-0.0376	-0.0127	-0.0168	0.0234	0.6506***
Social Studies	0.3766***	0.0010	0.0559	-0.0453	-0.0052	-0.0395	0.0289	0.6453***

Score								
Science Score	0.3238***	-0.0067	0.0111	-0.0170	-0.0261	-0.0790*	0.0494	0.6086***
Scientific Attitude Score	0.2748***	0.0189	0.0330	-0.0358	-0.0641†	0.0331	0.0681†	0.4152***
Law Score	0.2877***	0.0363	0.0192	-0.0818*	-0.0394	-0.0166	0.0512	0.4631***
Military Score	0.1406***	-0.0447	0.0547	-0.0237	-0.0165	0.0139	0.0198	0.3121***
Business Score	0.2256***	0.0197	0.0869*	-0.0839*	0.0505	-0.0076	-0.0485	0.4064***
Etiquette Score	0.2005***	-0.0030	0.0448	-0.0647†	0.0094	0.0251	0.0002	0.3922***
English Score	0.3817***	-0.0342	0.0341	-0.0470	-0.0484	0.0037	0.0503	0.6692***
Science Awards	0.0208	0.0570	-0.0089	-0.0075	0.0476	-0.0162	-0.0422	-0.0317
Fine Arts Awards	-0.0025	0.1068**	-0.0020	-0.0142	0.0360	-0.0215	-0.0339	-0.0356
Sports Awards	0.0512	0.1129**	0.0215	0.0060	0.1618***	-0.0293	-0.1525***	0.0025
Arithmetic Skills	0.1821***	-0.0249	0.0963**	-0.1192**	-0.0956**	0.0433	0.0847*	0.3674***
Reading Skills	0.0456	-0.0178	0.0469	-0.0239	0.0604†	0.0130	-0.0684†	0.0703†
Clerical Skills	0.1397***	0.0063	0.0633†	-0.0072	-0.0183	0.0674†	0.0173	0.2640***
Identification Skills	0.2112***	0.0143	-0.0563	-0.0093	-0.0289	-0.0259	0.0501	0.2725***
Public Service Interest	0.0241	0.0412	-0.0532	0.0968**	0.0440	-0.0421	-0.0374	-0.0436
Business Management Interest	-0.0274	0.0847*	-0.0354	0.0296	0.0845*	0.0860*	-0.1061**	-0.1827***

	Parent High School	Parent College	Some College	Degree or Higher	Overweight	Underweight	Tall	Short
Parent High School	1.0000							
Parent College	-0.2431***	1.0000						
Some College	0.0115	0.0116	1.0000					
Degree or Higher	0.1163**	0.1729***	-0.3066***	1.0000				
Overweight	-0.0499	0.0551	-0.0732*	-0.0389	1.0000			

Underweight	0.0104	0.0491	-0.0185	0.0410	-0.0508	1.0000		
Tall	-0.0085	-0.0271	-0.0393	0.0005	-0.0163	0.1523***	1.0000	
Short	-0.0049	-0.0139	0.0457	-0.0113	0.0519	-0.0342	-0.1428***	1.0000
Dates	0.0156	-0.0605†	-0.0232	-0.2492***	-0.0933*	-0.0023	-0.0110	-0.0707†
Comfortable	0.0766*	0.0950**	-0.0482	0.0745*	0.0448	0.0026	-0.0553	-0.0467
Wealthy	-0.0088	-0.0354	0.0156	-0.0786*	-0.0022	0.0103	0.0216	0.0730*
Sociability	0.0775*	0.0045	0.0450	-0.0108	0.0396	-0.0467	-0.0532	0.0471
Vigor	0.0892*	0.0142	0.0644†	0.0550	-0.0692†	-0.0528	-0.0601	-0.0114
Mature	0.0444	0.0640†	-0.0065	0.1212***	0.0244	-0.0222	-0.0595	-0.0094
Self-Confidence	0.0510	-0.0129	0.0398	0.0722*	0.0199	-0.0591	-0.0534	-0.0194
Tidiness	0.0496	0.0029	0.0354	-0.0106	0.0019	0.0094	-0.0894*	-0.0071
Vocabulary Score	0.2306***	0.1171**	0.0118	0.3360***	-0.0373	-0.0090	-0.0671†	0.0725*
Social Studies Score	0.2012***	0.1017**	0.0489	0.2907***	-0.0612†	-0.0181	-0.0798*	0.0528
Science Score	0.1971***	0.1470***	0.0095	0.2667***	-0.0716†	0.0247	-0.0784*	0.0602
Scientific Attitude Score	0.1651***	0.0885*	-0.0032	0.2634***	-0.0382	-0.0403	-0.0352	-0.0416
Law Score	0.1742***	0.0878*	0.0428	0.2088***	-0.0814*	0.0069	-0.0083	0.0352
Military Score	0.1271***	0.0411	-0.0092	0.1163**	0.0336	0.0266	-0.0550	0.0785*
Business Score	0.1912***	0.0663†	0.0057	0.1786***	-0.0368	-0.0615†	-0.0207	0.0045
Etiquette Score	0.1218***	0.1252***	-0.0231	0.1623***	-0.0343	-0.0103	-0.0245	0.0022
English Score	0.1930***	0.1078**	0.0157	0.3152***	-0.0803*	0.0663†	-0.1151**	0.0408
Science Awards	-0.0513	0.0863*	-0.0075	0.0349	0.0413	0.0538	0.0811*	0.0172
Fine Arts Awards	-0.0014	0.0825*	0.0152	0.0839*	-0.0362	0.0072	0.1225***	-0.0299
Sports Awards	-0.0237	0.0841*	-0.0272	0.0537	-0.0444	-0.0227	0.0331	-0.0314
Arithmetic Skills	0.0756*	0.0619†	0.0119	0.1364***	-0.0657†	-0.0009	-0.0976**	0.0735*
Reading Skills	-0.0023	-0.0547	-0.0118	0.0276	-0.0120	0.0416	0.0315	-0.0225
Clerical Skills	0.1093**	0.0093	0.0366	0.0854*	-0.0112	0.0563	-0.0590	0.0374
Identification Skills	0.0757*	0.0283	0.0364	0.1114**	-0.0038	0.0405	-0.0682†	0.0001

Public Service Interest	-0.0020	-0.0475	0.0237	0.0870*	-0.0214	-0.0321	-0.0474	-0.0309
Business Management Interest	0.0084	-0.0464	-0.0122	0.0017	-0.0345	-0.0277	-0.0081	-0.0483

	Dates	Comfortable	Wealthy	Sociability	Vigor	Mature	Self-Confidence	Tidiness
Dates	1.0000							
Comfortable	0.0255	1.0000						
Wealthy	0.0231	-0.5245***	1.0000					
Sociability	0.0755*	-0.0166	0.1129**	1.0000				
Vigor	0.0700†	-0.0110	0.0776*	0.5156***	1.0000			
Mature	0.0128	-0.0245	0.0400	0.3626***	0.5354***	1.0000		
Self-Confidence	0.1063**	0.0336	0.0015	0.3664***	0.3073***	0.4177***	1.0000	
Tidiness	0.1093**	-0.0103	0.0347	0.4396***	0.4798***	0.6139***	0.3017***	1.0000
Vocabulary Score	-0.1875***	0.1596***	-0.1697***	0.0895*	0.0770*	0.1254***	0.1318***	0.0224
Social Studies Score	-0.2048***	0.1795***	-0.1995***	0.0349	0.0528	0.0791*	0.0771*	-0.0192
Science Score	-0.1543***	0.1381***	-0.1146**	0.0119	0.0603	0.1031**	0.0170	0.0074
Scientific Attitude Score	-0.1255***	0.0262	-0.0879*	0.0313	0.0260	0.0680†	0.1274***	-0.0169
Law Score	-0.1941***	0.0922*	-0.0890*	0.0541	0.0534	0.0824*	0.1163**	-0.0267
Military Score	-0.1088**	0.0630†	-0.0535	0.0439	0.0517	0.0621†	0.0602	0.0290
Business Score	-0.0909*	0.1102**	-0.0624†	0.0179	0.0459	0.0490	0.1150**	-0.0355
Etiquette Score	-0.0983**	0.0300	0.0026	0.0341	0.0452	0.0700†	0.0569	-0.0088
English Score	-0.1563***	0.1581***	-0.1718***	0.0825*	0.0319	0.0946**	0.1532***	0.0522
Science Awards	-0.0501	-0.0469	0.0771*	0.0206	0.0133	0.0728*	0.0695†	-0.0038
Fine Arts Awards	-0.0212	-0.0182	0.0446	0.0227	0.0172	0.0699†	0.0665†	0.0066
Sports Awards	-0.0096	-0.0119	0.0837*	0.1198**	0.1196**	0.1006**	0.1051**	0.0484
Arithmetic Skills	-0.1033**	0.0939*	-0.0839*	0.0491	-0.0347	0.0178	0.0407	0.0440

Reading Skills	0.0544	0.0758*	-0.0569	0.0915*	0.0552	-0.0094	0.0610†	0.0827*
Clerical Skills	-0.0395	0.0707†	-0.0563	0.0821*	0.0852*	0.0063	0.1030**	0.0135
Identification Skills	-0.0667†	0.0897*	-0.0616†	-0.0323	-0.0705†	0.0029	0.0635†	-0.0235
Public Service Interest	0.0811*	-0.0200	0.0042	0.0750*	0.1525***	0.1561***	0.1314***	0.0673†
Business Management Interest	0.1359***	-0.0962**	0.0684†	0.1104**	0.1873***	0.1837***	0.0873*	0.1052**

	Vocabulary Score	Social Studies Score	Science Score	Scientific Attitude Score	Law Score	Military Score	Business Score	Etiquette Score
Vocabulary Score	1.0000							
Social Studies Score	0.6623***	1.0000						
Science Score	0.5894***	0.6247***	1.0000					
Scientific Attitude Score	0.4181***	0.4144***	0.3060***	1.0000				
Law Score	0.5323***	0.5019***	0.4019***	0.3763***	1.0000			
Military Score	0.3431***	0.3500***	0.2909***	0.2142***	0.2919***	1.0000		
Business Score	0.4511***	0.4782***	0.3559***	0.3836***	0.4378***	0.2730***	1.0000	
Etiquette Score	0.4175***	0.3526***	0.3574***	0.2915***	0.2799***	0.2311***	0.2670***	1.0000
English Score	0.6821***	0.6144***	0.5535***	0.4794***	0.4779***	0.2887***	0.3987***	0.3911***
Science Awards	-0.0707†	-0.0848*	-0.0543	-0.0315	-0.0717†	-0.0618†	-0.0353	-0.0272
Fine Arts Awards	-0.0565	-0.0884*	-0.0959**	-0.0262	-0.0551	-0.0813*	-0.0389	-0.0251
Sports Awards	0.0258	0.0050	-0.0268	0.0506	0.0403	0.0125	0.0836*	-0.0125
Arithmetic Skills	0.3367***	0.3352***	0.2787***	0.2483***	0.2230***	0.1178**	0.1805***	0.2056***
Reading Skills	0.1068**	0.0408	0.0472	0.0345	0.0417	0.0434	0.0771*	0.0545
Clerical Skills	0.2712***	0.2733***	0.2110***	0.2313***	0.2556***	0.1346***	0.1690***	0.1451***
Identification Skills	0.2710***	0.2596***	0.2416***	0.1883***	0.1817***	0.0382	0.1828***	0.1764***

Public Service Interest	-0.0663†	-0.0446	-0.0655†	-0.0371	-0.0404	0.0038	-0.0479	-0.0641†
Business Management Interest	-0.1780***	-0.1490***	-0.1571***	-0.0628†	-0.1155**	-0.0218	-0.1352***	-0.0810*

	English Score	Science Awards	Fine Arts Awards	Sports Awards	Arithmetic Skills	Reading Skills	Clerical Skills	Identification Skills
English Score	1.0000							
Science Awards	-0.0786*	1.0000						
Fine Arts Awards	-0.0777*	0.6560***	1.0000					
Sports Awards	-0.0084	0.4546***	0.5351***	1.0000				
Arithmetic Skills	0.4882***	-0.0558	-0.0786*	-0.0635†	1.0000			
Reading Skills	0.1057**	-0.1357***	-0.1513***	-0.1006**	0.1625***	1.0000		
Clerical Skills	0.3688***	-0.1080**	-0.1025**	0.0092	0.2826***	0.2011***	1.0000	
Identification Skills	0.3244***	-0.0596	-0.0681†	0.0080	0.2476***	0.1196**	0.2042***	1.0000
Public Service Interest	-0.0884*	0.0063	-0.0018	0.0283	-0.1047**	-0.0364	-0.0654†	-0.0441
Business Management Interest	-0.1668***	0.0367	0.0329	-0.0105	-0.0996**	-0.0881*	-0.0674†	-0.0617†

	Public Service Interest	Business Management Interest
Public Service Interest	1.0000	
Business Management Interest	0.6237***	1.0000

† $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Note: Pairwise correlation coefficients derived from Pearson-correlations. N. of obs. is 745.

4.8.3 Additional models

Table 2A: Direct Replication of KW – *(robust standard errors)*

	Model 1 Log (Hourly Earnings)	**Model 2** Log (Hourly Earnings)	**Model 3** Log (Hourly Earnings)	**Model 4** Log (Hourly Earnings)
Leader				
Both Captain and President	0.054***	0.049***	0.049***	0.038**
	(0.012)	(0.012)	(0.012)	(0.012)
Captain Only	0.036**	0.036**	0.036**	0.035**
	(0.013)	(0.013)	(0.013)	(0.013)
President Only	0.036***	0.020†	0.019†	0.010
	(0.011)	(0.011)	(0.011)	(0.011)
Member				
Both on Team and in Club	0.107***	0.073**	0.070**	0.055*
	(0.025)	(0.024)	(0.025)	(0.025)
On Team only	0.083*	0.062†	0.060†	0.059†
	(0.035)	(0.034)	(0.034)	(0.034)
In Club Only	0.036	0.008	0.005	-0.006
	(0.026)	(0.026)	(0.026)	(0.026)
Controls				
Math Score		0.002***	0.002***	0.001***
		(0.000)	(0.000)	(0.000)
Parent´s Education				
High School			0.020†	0.011
			(0.011)	(0.011)
College Degree			0.015	-0.008
			(0.014)	(0.014)
Educational Attainment				
Some College				0.051***
				(0.012)
College Degree or Higher				0.136***
				(0.013)
School-fixed Effects	Yes	Yes	Yes	Yes
F-Value	23.58	45.63	34.87	39.30
p > F	0.000	0.000	0.000	0.000
Adjusted R^2	0.160	0.177	0.178	0.189
Observations	24041	24041	24041	24041

† p<0.1 * p<0.05, ** p<.01, *** p<.001

Note: Table 2A directly replicates the results depicted in Kuhn and Weinberger (2005: 405 (columns 5-8)). The coefficients are derived from OLS regressions. This table differs from Table 4.2 reported in the essay due to the inclusion of robust standard errors (reported in parentheses). All models include (unreported) grade and school attainment control variables as well as school dummies. Models 3 and 4 include an unreported dummy variable for lack of parent's education. Model 4 includes an unreported dummy variable for lack of educational attainment.

Table 2B: Direct Replication of KW - *(Individual Leadership Categorization)*

	Model 1 Log (Hourly Earnings)	**Model 2** Log (Hourly Earnings)	**Model 3** Log (Hourly Earnings)
Leader			
Both Captain and President	0.027*		
	(0.011)		
Captain Only		0.035**	
		(0.012)	
President Only			0.009
			(0.011)
Member			
Both on Team and in Club	0.062*	0.070**	0.065**
	(0.024)	(0.023)	(0.023)
On Team only	0.063†	0.041	0.062†
	(0.034)	(0.031)	(0.033)
In Club Only	-0.004	-0.002	0.003
	(0.026)	(0.025)	(0.024)
Controls			
Math Score	0.001***	0.002***	0.001***
	(0.000)	(0.000)	(0.000)
Parent´s Education			
High School	0.011	0.002	0.005
	(0.011)	(0.014)	(0.013)
College Degree	-0.008	-0.020	-0.010
	(0.014)	(0.018)	(0.016)
Educational Attainment			
Some College	0.051***	0.040**	0.033*
	(0.012)	(0.014)	(0.013)
College Degree or Higher	0.136***	0.115***	0.131***
	(0.013)	(0.016)	(0.015)
School-fixed Effects	Yes	Yes	Yes
F-Value	44.08	32.25	35.44
p > F	0.000	0.000	0.000
Adjusted R^2	0.189	0.255	0.200
Observations	24041	13315	15924

† $p<0.1$ * $p<0.05$, ** $p<.01$, *** $p<.001$

Note: Table 2B directly replicates the results depicted in Kuhn and Weinberger (2005: 405 (column 8)). The coefficients are derived from OLS regressions. This table differs from Table 4.2 reported in the paper due to the separation of estimation models: Each leadership category is individually included in separate regressions with the category "no leadership" as the baseline. All models include (unreported) grade and school attainment control variables as well as school dummies, unreported dummy variables for lack of parents' education and unreported dummy variables for lack of educational attainment.

Table 3A: Conceptual replication - *Endogeneity susceptibility of KW Table 2 estimates with school-dummies*

	Outcome Assessment Model 1 Log (Hourly Earnings)	**Matching Model 1** President and Captain	**Matching Model 2** Captain Only	**Matching Model 3** President Only
Leader (for Model 1 only)				
Both Captain and President	0.041**			
	(0.013)			
Captain Only	0.036**			
	(0.014)			
President Only	0.004			
	(0.012)			
Member				
Both on Team and in Club	0.042†	1.563***	1.176**	1.842***
	(0.025)	(0.255)	(0.447)	(0.269)
On Team only	0.047	0.869*	0.855†	1.240**
	(0.034)	(0.341)	(0.491)	(0.379)
In Club Only	-0.010	0.620*	0.088	1.666***
	(0.026)	(0.273)	(0.459)	(0.272)
Personal Data				
Overweight	-0.139**	-0.446	0.072	-0.030
	(0.048)	(0.329)	(0.378)	(0.239)
Underweight	-0.035	-0.453	-0.203	0.296
	(0.034)	(0.287)	(0.281)	(0.229)
Tall	-0.012	-0.037	0.056	-0.096
	(0.010)	(0.074)	(0.085)	(0.065)
Short	0.015	0.075	-0.194	0.040
	(0.016)	(0.120)	(0.133)	(0.109)
Dates	-0.002	0.161***	-0.008	0.042
	(0.004)	(0.029)	(0.032)	(0.029)
Comfortable	-0.002	0.094	-0.165	0.104
	(0.012)	(0.094)	(0.106)	(0.090)
Wealthy	0.053**	0.087	-0.242†	0.295**
	(0.017)	(0.115)	(0.132)	(0.109)
Personal Characteristics				
Sociability	0.020***	0.051	-0.016	0.095**
	(0.005)	(0.041)	(0.046)	(0.036)
Vigor	0.003	0.254***	0.189***	-0.145***
	(0.006)	(0.042)	(0.049)	(0.037)
Mature	-0.000	0.100*	-0.230***	0.111**
	(0.006)	(0.044)	(0.053)	(0.040)
Self-Confidence	0.007	0.016	0.016	0.122***
	(0.005)	(0.037)	(0.043)	(0.032)
Tidiness	0.001	-0.035	0.026	0.042
	(0.005)	(0.041)	(0.046)	(0.037)
Test Scores				
Math Score	0.002***	0.001	0.004†	0.002
	(0.000)	(0.002)	(0.002)	(0.002)
Vocabulary Score	-0.046†	-0.152	-0.070	0.171
	(0.026)	(0.201)	(0.221)	(0.180)
Social Studies Score	-0.036	-0.275	0.055	-0.002
	(0.026)	(0.172)	(0.212)	(0.167)
Science Score	0.025	-0.235	-0.651**	0.437**
	(0.025)	(0.184)	(0.207)	(0.168)
Scientific Attitude Score	0.080***	0.053	-0.159	0.120
	(0.018)	(0.128)	(0.159)	(0.128)

Law Score	0.016	-0.361*	-0.251	0.043
	(0.021)	(0.146)	(0.170)	(0.138)
Military Score	0.025	0.206	-0.223	0.125
	(0.019)	(0.136)	(0.162)	(0.129)
Business Score	0.080***	-0.234	0.071	0.250†
	(0.019)	(0.145)	(0.162)	(0.133)
Etiquette Score	-0.013	0.010	-0.149	0.097
	(0.016)	(0.121)	(0.147)	(0.114)
English Score	-0.002	0.593**	-0.123	0.320†
	(0.024)	(0.181)	(0.199)	(0.166)
Awards				
Science Awards	-0.003	0.006	0.041	-0.030
	(0.006)	(0.039)	(0.055)	(0.035)
Fine Arts Awards	-0.004	0.013	-0.089**	0.043*
	(0.003)	(0.021)	(0.033)	(0.018)
Sports Awards	0.001	0.127***	0.022	0.021
	(0.002)	(0.013)	(0.017)	(0.013)
Cognitive Skills				
Arithmetic Skills	0.010	-0.014	0.045	-0.209†
	(0.017)	(0.119)	(0.143)	(0.110)
Reading Skills	0.035*	-0.215†	0.053	0.104
	(0.017)	(0.121)	(0.141)	(0.111)
Clerical Skills	-0.019	-0.006	-0.188	0.049
	(0.018)	(0.122)	(0.137)	(0.114)
Identification Skills	-0.005	-0.235†	-0.111	-0.094
	(0.016)	(0.124)	(0.141)	(0.112)
Personal Interests				
Public Service Interest	0.002	0.106*	-0.113*	0.099*
	(0.006)	(0.045)	(0.051)	(0.039)
Business Management Interest	0.011†	0.017	0.129*	0.004
	(0.006)	(0.045)	(0.051)	(0.039)
School-fixed Effects	Yes	Yes	Yes	Yes
F / Chi^2	13.08	1934.03	1242.90	1615.37
p > F /Chi^2	0.000	0.000	0.000	0.000
R^2 / Pseudo R^2	0.183	0.1846	0.1427	0.1272
Observations	22095	21658	20710	21735

† p<0.1 * p<0.05, ** p<.01, *** p<.001

Note: Table 3A assesses the need for a potential endogeneity adjustment of the results depicted in Kuhn and Weinberger (2005). The coefficients for Outcome Assessment Model (1) are derived from OLS regressions, coefficients for Matching Model 1, 2 and 3 are derived from logit regression with standard errors in parentheses. The dependent variable in (1) is the natural logarithm of hourly earnings, the dependent variable in the matching models 1,2, and 3 correspond to a dichotomous indicator variable whether individuals acted as captain and president, captain only, or president only. Table 3A differs from Table 4.3 reported in the paper as it includes school dummies.

Table 6A: Conceptual replication – *First stage results for the Conventional IV regressions for KW sample in Table 6*

	(1) First-stage regression: Both Captain and President	(2) First-stage regression: Captain Only	(3) First-stage regression: President Only
Instruments			
Father in Club	0.007	-0.010	-0.034
	(0.016)	(0.021)	(0.022)
Mother in Club	0.023	0.011	0.054*
	(0.015)	(0.021)	(0.021)
Father in Team	0.042**	0.055**	0.044*
	(0.013)	(0.019)	(0.017)
Mother in Team	0.039*	0.018	-0.024
	(0.018)	(0.024)	(0.022)
Covariates			
Both on Team and in Club	0.103***	0.160***	0.156**
	(0.016)	(0.029)	(0.048)
On Team only	0.042	0.086†	0.022
	(0.028)	(0.046)	(0.059)
In Club Only	0.017	-0.006	0.063
	(0.017)	(0.028)	(0.048)
Overweight	-0.083*	0.034	-0.049
	(0.036)	(0.083)	(0.047)
Underweight	-0.044	-0.018	0.048
	(0.032)	(0.035)	(0.048)
Tall	-0.006	-0.010	-0.013
	(0.012)	(0.017)	(0.016)
Short	0.011	-0.048*	-0.010
	(0.018)	(0.023)	(0.025)
Dates	0.021***	-0.001	0.018**
	(0.005)	(0.006)	(0.007)
Comfortable	-0.003	-0.017	0.020
	(0.015)	(0.020)	(0.019)
Wealthy	-0.004	-0.022	0.039
	(0.020)	(0.025)	(0.025)
Sociability	0.005	-0.001	0.019*
	(0.007)	(0.009)	(0.008)
Vigor	0.033***	0.040***	-0.003
	(0.007)	(0.009)	(0.009)
Mature	0.014†	-0.017	0.035***
	(0.007)	(0.010)	(0.011)
Self-Confidence	0.004	0.017*	0.032***
	(0.006)	(0.008)	(0.008)
Tidiness	-0.006	-0.001	-0.006
	(0.007)	(0.009)	(0.009)
Math Score	0.000	0.001*	0.001
	(0.000)	(0.000)	(0.000)
Vocabulary Score	-0.008	-0.022	-0.058
	(0.030)	(0.042)	(0.040)
Social Studies Score	-0.056*	-0.007	-0.007
	(0.026)	(0.037)	(0.037)
Science Score	-0.040	-0.119**	0.037
	(0.029)	(0.039)	(0.038)
Scientific Attitude Score	-0.002	-0.028	0.026
	(0.021)	(0.031)	(0.030)
Law Score	-0.030	-0.060†	-0.016
	(0.023)	(0.033)	(0.031)
Military Score	0.047*	-0.004	0.036
	(0.021)	(0.031)	(0.029)
Business Score	-0.037	0.020	0.043

	(0.024)	(0.032)	(0.030)
Etiquette Score	-0.000	-0.027	0.004
	(0.020)	(0.028)	(0.026)
English Score	0.064*	0.069†	0.133***
	(0.028)	(0.039)	(0.040)
Science Awards	0.002	0.006	-0.006
	(0.007)	(0.011)	(0.010)
Fine Arts Awards	0.000	-0.014*	0.005
	(0.004)	(0.006)	(0.005)
Sports Awards	0.020***	0.026***	0.025***
	(0.003)	(0.004)	(0.004)
Arithmetic Skills	0.001	-0.027	-0.062*
	(0.020)	(0.027)	(0.025)
Reading Skills	-0.033†	0.012	0.003
	(0.019)	(0.027)	(0.024)
Clerical Skills	0.008	-0.005	-0.007
	(0.019)	(0.026)	(0.026)
Identification Skills	-0.041*	-0.033	-0.004
	(0.020)	(0.027)	(0.025)
Public Service Interest	0.017*	-0.003	0.028**
	(0.008)	(0.010)	(0.009)
Business Management Interest	0.000	0.029**	0.018*
	(0.007)	(0.009)	(0.009)
School-fixed effects	No	No	No
F	6.90	3.09	3.77
p > F	0.000	0.015	0.005
Observations	22093	12222	14652

† p<0.1 * p<0.05, ** p<.01, *** p<.001

Note: Table 6A reports the first stage coefficient estimates for the dichotomous indicator variables whether individuals acted as captain and president, captain only, or president only derived from a conventional IV estimator (Stata; ivreg2). Abbreviated second stage results were reported in Table 4.6 in the paper. Information whether each parent was a member of a club of a team represents the instrumental variables. The sample for the treatment analysis is based on the sample of white males employed in Kuhn and Weinberger (2005).

Table 6B: *Conceptual replication – Full Second stage results for the IV regressions (conventional, conditional and heteroskedasticity-based) for KW sample*

	(1) Conventional	(2) Conventional	(3) Conventional	(4) Conditional	(5) Conditional	(6) Conditional	(7) Heteroske-dasticity	(8) Heteroske-dasticity	(9) Heteroske-dasticity
	Log(Hourly Earnings)	Log(Hourly Earnings)	Log(Hourly Earnings)	Log(Hourly Earnings)	Log(Hourly Earnings)	Log(Hourly Earnings)	Log(Hourly Earnings)	Log(Hourly Earnings)	Log(Hourly Earnings)
Instrumented Variable									
President and Captain	0.412*			0.321**			0.163*		
	(0.181)			(0.110)			(0.068)		
Captain Only		0.008			0.114			0.058	
		(0.236)			(0.154)			(0.112)	
President Only			0.073			-0.038			0.293†
			(0.227)			(0.115)			(0.165)
Covariates									
Both on Team and in Club	-0.028	0.023	0.022	-0.006	0.005	0.027	-0.002	0.009	-0.003
	(0.035)	(0.051)	(0.047)	(0.022)	(0.032)	(0.031)	(0.028)	(0.035)	(0.043)
On Team only	0.030	0.019	0.067†	0.042	0.022	0.049†	0.037	0.017	0.073†
	(0.039)	(0.044)	(0.036)	(0.026)	(0.031)	(0.027)	(0.038)	(0.040)	(0.042)
In Club Only	-0.033	-0.033	-0.008	-0.029	-0.034	-0.025	-0.029	-0.036	-0.013
	(0.029)	(0.030)	(0.032)	(0.020)	(0.021)	(0.024)	(0.028)	(0.029)	(0.036)
Overweight	-0.092†	-0.120†	-0.090	-0.048†	-0.067*	-0.075**	-0.119*	-0.127†	-0.078
	(0.053)	(0.064)	(0.056)	(0.025)	(0.028)	(0.026)	(0.048)	(0.065)	(0.058)
Underweight	-0.018	-0.071	-0.036	-0.001	-0.032	-0.013	-0.025	-0.070	-0.045
	(0.041)	(0.046)	(0.044)	(0.020)	(0.023)	(0.022)	(0.037)	(0.045)	(0.043)
Tall	-0.020†	-0.017	-0.015	-0.013*	-0.019*	-0.016*	-0.023*	-0.015	-0.012
	(0.012)	(0.015)	(0.013)	(0.006)	(0.008)	(0.007)	(0.011)	(0.014)	(0.014)
Short	0.018	0.028	0.031	0.001	0.007	0.012	0.021	0.031	0.028
	(0.023)	(0.031)	(0.028)	(0.010)	(0.012)	(0.011)	(0.022)	(0.029)	(0.029)
Dates	-0.009	-0.001	0.001	-0.004	0.002	0.004	-0.003	-0.001	-0.003
	(0.006)	(0.005)	(0.006)	(0.003)	(0.003)	(0.004)	(0.004)	(0.005)	(0.006)
Comfortable	0.011	0.030	0.006	0.024**	0.021*	0.020*	0.015	0.035†	0.007
	(0.016)	(0.020)	(0.020)	(0.008)	(0.010)	(0.009)	(0.016)	(0.019)	(0.020)
Wealthy	0.070***	0.087***	0.066*	0.073***	0.061***	0.069***	0.076***	0.092***	0.061*
	(0.021)	(0.025)	(0.026)	(0.010)	(0.013)	(0.012)	(0.019)	(0.024)	(0.026)

Sociability	0.018**	0.021**	0.021*	0.016***	0.018***	0.023***	0.020**	0.020**	0.017*
	(0.007)	(0.008)	(0.008)	(0.004)	(0.005)	(0.005)	(0.006)	(0.008)	(0.008)
Vigor	-0.010	-0.002	0.003	-0.009*	-0.006	-0.005	-0.001	-0.003	0.003
	(0.010)	(0.014)	(0.008)	(0.005)	(0.007)	(0.004)	(0.007)	(0.010)	(0.009)
Mature	-0.008	0.009	0.003	-0.005	0.004	0.003	-0.006	0.008	-0.007
	(0.009)	(0.012)	(0.013)	(0.004)	(0.005)	(0.006)	(0.008)	(0.012)	(0.012)
Self-Confidence	0.007	0.009	-0.000	0.003	0.002	0.000	0.009	0.008	-0.007
	(0.006)	(0.008)	(0.009)	(0.003)	(0.004)	(0.005)	(0.005)	(0.008)	(0.008)
Tidiness	0.009	-0.007	0.010	0.014***	0.008†	0.015***	0.007	-0.005	0.012
	(0.007)	(0.008)	(0.007)	(0.004)	(0.005)	(0.004)	(0.006)	(0.007)	(0.008)
Math Score	0.002***	0.002***	0.002***	0.002***	0.003***	0.003***	0.002***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Vocabulary Score	-0.005	-0.036	0.011	0.029†	0.036†	0.047*	-0.005	-0.037	0.019
	(0.031)	(0.035)	(0.036)	(0.016)	(0.021)	(0.019)	(0.029)	(0.033)	(0.037)
Social Studies Score	-0.002	-0.001	-0.007	0.018	0.011	-0.000	-0.012	-0.000	-0.014
	(0.030)	(0.032)	(0.033)	(0.015)	(0.019)	(0.017)	(0.027)	(0.031)	(0.035)
Science Score	0.006	0.036	-0.008	-0.013	-0.012	-0.020	-0.004	0.045	-0.010
	(0.031)	(0.043)	(0.032)	(0.016)	(0.023)	(0.019)	(0.027)	(0.035)	(0.035)
Scientific Attitude Score	0.088***	0.094***	0.062*	0.061***	0.066***	0.063***	0.082***	0.091***	0.060*
	(0.021)	(0.027)	(0.025)	(0.011)	(0.015)	(0.013)	(0.020)	(0.026)	(0.027)
Law Score	0.025	0.006	0.026	0.023†	0.022	0.009	0.014	0.016	0.032
	(0.025)	(0.032)	(0.027)	(0.012)	(0.016)	(0.015)	(0.022)	(0.028)	(0.027)
Military Score	0.040†	0.028	0.047†	0.016	0.005	0.023†	0.051*	0.028	0.041
	(0.024)	(0.028)	(0.028)	(0.012)	(0.015)	(0.013)	(0.022)	(0.027)	(0.028)
Business Score	0.095***	0.104***	0.069*	0.066***	0.043**	0.061***	0.088***	0.103***	0.052†
	(0.023)	(0.028)	(0.028)	(0.012)	(0.015)	(0.014)	(0.021)	(0.026)	(0.027)
Etiquette Score	0.009	0.021	-0.005	-0.002	0.003	0.004	0.012	0.023	-0.005
	(0.020)	(0.025)	(0.023)	(0.011)	(0.014)	(0.012)	(0.018)	(0.025)	(0.024)
English Score	-0.077*	-0.070†	-0.090*	-0.077***	-0.055**	-0.055*	-0.066*	-0.080*	-0.128***
	(0.030)	(0.037)	(0.043)	(0.016)	(0.019)	(0.022)	(0.026)	(0.033)	(0.039)
Science Awards	-0.002	0.002	0.006	0.003	0.002	0.007	-0.001	0.003	0.007
	(0.007)	(0.009)	(0.007)	(0.004)	(0.005)	(0.004)	(0.007)	(0.008)	(0.008)
Fine Arts Awards	-0.006	-0.002	-0.008*	-0.003	0.001	-0.003	-0.006†	-0.002	-0.009*
	(0.004)	(0.006)	(0.004)	(0.002)	(0.003)	(0.002)	(0.004)	(0.005)	(0.004)
Sports Awards	-0.007	-0.002	-0.002	-0.007*	-0.004	0.002	-0.002	-0.004	-0.008
	(0.005)	(0.007)	(0.007)	(0.003)	(0.005)	(0.003)	(0.003)	(0.004)	(0.005)
Arithmetic Skills	-0.006	0.007	-0.007	0.003	0.002	-0.009	-0.005	0.007	0.012

	(0.021)	(0.025)	(0.027)	(0.011)	(0.014)	(0.013)	(0.020)	(0.024)	(0.026)
Reading Skills	0.052**	0.032	0.036†	0.006	0.001	0.005	0.042*	0.032	0.032
	(0.020)	(0.024)	(0.021)	(0.010)	(0.013)	(0.012)	(0.018)	(0.023)	(0.023)
Clerical Skills	0.005	0.017	0.027	0.021*	0.026*	0.026*	0.008	0.023	0.029
	(0.020)	(0.024)	(0.023)	(0.010)	(0.013)	(0.012)	(0.019)	(0.024)	(0.024)
Identification Skills	0.014	-0.013	-0.019	0.017	0.011	0.004	0.002	-0.009	-0.018
	(0.022)	(0.025)	(0.021)	(0.010)	(0.013)	(0.012)	(0.019)	(0.023)	(0.023)
Public Service Interest	-0.000	-0.006	-0.011	0.005	0.006	0.002	0.005	-0.004	-0.015
	(0.008)	(0.009)	(0.010)	(0.004)	(0.005)	(0.006)	(0.007)	(0.009)	(0.010)
Business Management Interest	0.012	0.013	0.016†	0.013***	0.009†	0.016***	0.011	0.010	0.009
	(0.008)	(0.011)	(0.009)	(0.004)	(0.005)	(0.004)	(0.007)	(0.009)	(0.009)
School-fixed effects	No	No	No	No	No	No	No	No	No
Second Stage F-Statistic	12.64	9.20	9.65	46.66	29.74	34.70	14.27	9.41	8.70
p > F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Observations	22,093	12,222	14,652	22,093	12,222	14,652	22,093	12,222	14,652

† p<0.1 * p<0.05, ** p<.01, *** p<.001

Note: Table 6B reports the second stage coefficients and standard deviations from a conventional, a conditional and a heteroskedasticity-based IV estimation. Abbreviated second stage results were reported in Table 4.6 in the paper. The instrumented variables are whether individuals acted as captain and president, captain only, or president. Information whether each parent was a member of a club of a team represents the instrumental variables. The sample for the treatment analysis is based on the sample of white males employed in Kuhn and Weinberger (2005).

Table 6C: Conceptual replication – *Full Second stage results for the IV regressions for KW sample with school dummies*

	Conventional IV- Estimation		
	Model 1	Model 2	Model 3
Instrumented Variable			
President and Captain	0.412*		
	(0.181)		
Captain Only		0.008	
		(0.236)	
President Only			0.073
			(0.227)
Covariates			
Both on Team and in Club	-0.028	0.023	0.022
	(0.035)	(0.051)	(0.047)
On Team only	0.030	0.019	0.067†
	(0.039)	(0.044)	(0.036)
In Club Only	-0.033	-0.033	-0.008
	(0.029)	(0.030)	(0.032)
Overweight	-0.092†	-0.120†	-0.090
	(0.053)	(0.064)	(0.056)
Underweight	-0.018	-0.071	-0.036
	(0.041)	(0.046)	(0.044)
Tall	-0.020†	-0.017	-0.015
	(0.012)	(0.015)	(0.013)
Short	0.018	0.028	0.031
	(0.023)	(0.031)	(0.028)
Dates	-0.009	-0.001	0.001
	(0.006)	(0.005)	(0.006)
Comfortable	0.011	0.030	0.006
	(0.016)	(0.020)	(0.020)
Wealthy	0.070***	0.087***	0.066*
	(0.021)	(0.025)	(0.026)
Sociability	0.018**	0.021**	0.021*
	(0.007)	(0.008)	(0.008)
Vigor	-0.010	-0.002	0.003
	(0.010)	(0.014)	(0.008)
Mature	-0.008	0.009	0.003
	(0.009)	(0.012)	(0.013)
Self-Confidence	0.007	0.009	-0.000
	(0.006)	(0.008)	(0.009)
Tidiness	0.009	-0.007	0.010
	(0.007)	(0.008)	(0.007)
Math Score	0.002***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)
Vocabulary Score	-0.005	-0.036	0.011
	(0.031)	(0.035)	(0.036)
Social Studies Score	-0.002	-0.001	-0.007
	(0.030)	(0.032)	(0.033)
Science Score	0.006	0.036	-0.008
	(0.031)	(0.043)	(0.032)
Scientific Attitude Score	0.088***	0.094***	0.062*
	(0.021)	(0.027)	(0.025)
Law Score	0.025	0.006	0.026
	(0.025)	(0.032)	(0.027)
Military Score	0.040†	0.028	0.047†

	(0.024)	(0.028)	(0.028)
Business Score	0.095***	0.104***	0.069*
	(0.023)	(0.028)	(0.028)
Etiquette Score	0.009	0.021	-0.005
	(0.020)	(0.025)	(0.023)
English Score	-0.077*	-0.070†	-0.090*
	(0.030)	(0.037)	(0.043)
Science Awards	-0.002	0.002	0.006
	(0.007)	(0.009)	(0.007)
Fine Arts Awards	-0.006	-0.002	-0.008*
	(0.004)	(0.006)	(0.004)
Sports Awards	-0.007	-0.002	-0.002
	(0.005)	(0.007)	(0.007)
Arithmetic Skills	-0.006	0.007	-0.007
	(0.021)	(0.025)	(0.027)
Reading Skills	0.052**	0.032	0.036†
	(0.020)	(0.024)	(0.021)
Clerical Skills	0.005	0.017	0.027
	(0.020)	(0.024)	(0.023)
Identification Skills	0.014	-0.013	-0.019
	(0.022)	(0.025)	(0.021)
Public Service Interest	-0.000	-0.006	-0.011
	(0.008)	(0.009)	(0.010)
Business Management Interest	0.012	0.013	0.016†
	(0.008)	(0.011)	(0.009)
School-fixed effects	Yes	Yes	Yes
First stage (Kleibergen-Paap) F-Statistic	11.95	6.14	6.44
p-value (F-Statistic)	0.000	0.000	0.000
Cragg-Donald-Wald F-Statistic	37.01	19.02	13.88
Sargan-Hansen statistic	1.891	5.896	3.666
p-value (Sargan-Hansen statistic)	0.595	0.118	0.299
Observations	22048	12158	14587

† p<0.1 * p<0.05, ** p<.01, *** p<.001

Note: Table 6C reports the second stage coefficients and standard deviations from conventional IV estimations. The instrumented variables are whether individuals acted as captain and president, captain only, or president. Information whether each parent was a member of a club of a team represents the instrumental variables. The sample for the treatment analysis is based on the sample of white males employed in Kuhn and Weinberger (2005). This table differs from Table 4.6 reported in the paper, as it additionally includes school-fixed effects. The sample for the treatment analysis is based on the sample of white males employed in Kuhn and Weinberger (2005).

Table 6D: Conceptual replication – *Full Second stage results for the IV regressions for KW sample with robust standard errors*

	Conventional IV- Estimation		
	Model 1	Model 2	Model 3
Instrumented Variable			
President and Captain	0.284*		
	(0.141)		
Captain Only		-0.004	
		(0.157)	
President Only			0.190
			(0.172)
Covariates			
Both on Team and in Club	0.022	0.075*	0.019
	(0.029)	(0.035)	(0.043)
On Team only	0.043	0.032	0.045
	(0.034)	(0.031)	(0.034)
In Club Only	-0.011	-0.003	-0.007
	(0.027)	(0.026)	(0.030)
Overweight	-0.127*	-0.118†	-0.137*
	(0.050)	(0.064)	(0.053)
Underweight	-0.025	-0.051	-0.071†
	(0.036)	(0.045)	(0.041)
Tall	-0.011	-0.006	-0.011
	(0.010)	(0.012)	(0.012)
Short	0.012	0.023	0.024
	(0.017)	(0.018)	(0.019)
Dates	-0.008†	-0.002	-0.002
	(0.005)	(0.004)	(0.006)
Comfortable	-0.006	-0.001	-0.015
	(0.013)	(0.014)	(0.015)
Wealthy	0.050**	0.056**	0.036
	(0.017)	(0.019)	(0.022)
Sociability	0.019***	0.018**	0.023**
	(0.006)	(0.006)	(0.007)
Vigor	-0.005	0.000	0.004
	(0.008)	(0.009)	(0.007)
Mature	-0.005	0.007	-0.004
	(0.007)	(0.008)	(0.009)
Self-Confidence	0.007	0.008	-0.005
	(0.005)	(0.007)	(0.008)
Tidiness	0.003	-0.007	0.001
	(0.006)	(0.006)	(0.006)
Math Score	0.002***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)
Vocabulary Score	-0.040	-0.044	-0.063*
	(0.027)	(0.030)	(0.030)
Social Studies Score	-0.026	-0.029	-0.014
	(0.026)	(0.028)	(0.031)
Science Score	0.029	0.038	0.019
	(0.026)	(0.034)	(0.028)
Scientific Attitude Score	0.078***	0.066**	0.059**
	(0.018)	(0.022)	(0.022)
Law Score	0.027	0.025	0.042†
	(0.022)	(0.026)	(0.024)
Military Score	0.017	-0.026	0.012

	(0.020)	(0.023)	(0.024)
Business Score	0.089***	0.101***	0.079***
	(0.020)	(0.024)	(0.023)
Etiquette Score	-0.014	-0.016	-0.030
	(0.016)	(0.021)	(0.020)
English Score	-0.022	-0.032	-0.061†
	(0.028)	(0.030)	(0.034)
Science Awards	-0.003	0.005	0.008
	(0.006)	(0.008)	(0.007)
Fine Arts Awards	-0.005	-0.002	-0.004
	(0.004)	(0.005)	(0.003)
Sports Awards	-0.004	-0.003	-0.007
	(0.004)	(0.005)	(0.005)
Arithmetic Skills	0.011	0.025	0.013
	(0.017)	(0.020)	(0.021)
Reading Skills	0.042*	0.021	0.022
	(0.018)	(0.019)	(0.019)
Clerical Skills	-0.019	-0.012	0.011
	(0.018)	(0.021)	(0.020)
Identification Skills	0.003	-0.010	-0.009
	(0.018)	(0.021)	(0.021)
Public Service Interest	-0.002	-0.011	-0.020*
	(0.007)	(0.007)	(0.009)
Business Management Interest	0.011†	0.016*	0.016*
	(0.006)	(0.008)	(0.007)
School-fixed effects	Yes	Yes	Yes
First stage (Kleibergen-Paap) F-Statistic	6.90	3.09	3.77
p-value (F-Statistic)	0.000	0.015	0.005
Cragg-Donald-Wald F-Statistic	31.31	13.06	11.27
Sargan-Hansen statistic	0.933	3.87	2.69
p-value (Sargan-Hansen statistic)	0.812	0.276	0.442
Observations	22,093	12,222	14,652

† p<0.1 * p<0.05, ** p<.01, *** p<.001

Note: Table 6D reports the second stage coefficients and standard deviations from conventional IV estimations. The instrumented variables are whether individuals acted as captain and president, captain only, or president. Information whether each parent was a member of a club of a team represents the instrumental variables. The sample for the treatment analysis is based on the sample of white males employed in Kuhn and Weinberger (2005). This table differs from Table 4.6 reported in the paper, as it includes robust standard errors. The sample for the treatment analysis is based on the sample of white males employed in Kuhn and Weinberger (2005).

Table 7A: Extension – *OLS effects of High School Leadership Activities on Hourly Earnings for expanded samples with robust standard errors*

	Model 1 White Females	**Model 2** Non-White Males	**Model 3** Non-White Females
Leader			
Both Captain and President	0.010	0.110	0.050
	(0.016)	(0.086)	(0.075)
Captain Only	0.023	0.135*	0.173
	(0.016)	(0.069)	(0.107)
President Only	0.009	-0.012	0.047
	(0.015)	(0.075)	(0.064)
Member			
Both on Team and in Club	0.110***	0.126	-0.313*
	(0.031)	(0.160)	(0.145)
On Team only	0.012	0.199	-0.267†
	(0.076)	(0.195)	(0.146)
In Club Only	0.086**	0.173	-0.285†
	(0.031)	(0.170)	(0.149)
Controls			
Math Score	0.002***	0.002*	0.004***
	(0.000)	(0.001)	(0.001)
Parents' Education			
High School	0.018	0.031	0.050
	(0.013)	(0.072)	(0.055)
College Degree	0.024	0.160†	-0.047
	(0.017)	(0.086)	(0.106)
Educational Attainment			
Some College	0.079***	-0.014	0.152*
	(0.015)	(0.073)	(0.062)
College Degree or Higher	0.321***	0.198**	0.445***
	(0.016)	(0.075)	(0.078)
School-fixed Effects	Yes	Yes	Yes
F-Value	67.87	3.33	7.91
p > F	0.000	0.000	0.000
Adjusted R^2	0.337	0.319	0.496
Observations	11824	747	816

† p<0.1 * p<0.05, ** p<.01, *** p<.001

Note: Table 7A extends the results depicted in Kuhn and Weinberger (2005: 405 (columns 5-8). This table differs from Table 4.7 reported in the paper due to the inclusion of robust standard errors (reported in parentheses). All models include (unreported) grade and school attainment control variables as well as school dummies. Models 3 and 4 include an unreported dummy variable for lack of parent's education. Model 4 includes an unreported dummy variable for lack of educational attainment. The sample for the models labelled 1, 2 and 3 are based on the sample of white females, non-white males, and non-white females that were excluded in the original Kuhn and Weinberger (2005) study.

Table 7B: Extension – *OLS effects of High School Leadership Activities on Hourly Earnings for expanded samples with school average income*

	Model 1 White Females	**Model 2** Non-White Males	**Model 3** Non-White Females
Leader			
Both Captain and President	-0.005	0.083	0.166**
	(0.017)	(0.060)	(0.064)
Captain Only	0.020	0.052	0.140
	(0.018)	(0.069)	(0.099)
President Only	0.000	-0.076	0.150*
	(0.017)	(0.057)	(0.066)
Member			
Both on Team and in Club	0.065*	0.013	-0.416***
	(0.032)	(0.086)	(0.101)
On Team only	0.006	0.086	-0.460***
	(0.066)	(0.108)	(0.113)
In Club Only	0.034	0.071	-0.309**
	(0.032)	(0.102)	(0.101)
Controls			
Math Score	0.002***	0.002†	0.003**
	(0.000)	(0.001)	(0.001)
Parents' Education			
High School	0.022	0.003	0.058
	(0.015)	(0.057)	(0.055)
College Degree	0.015	0.184**	0.061
	(0.020)	(0.068)	(0.077)
Educational Attainment			
Some College	0.066***	0.036	0.188***
	(0.017)	(0.058)	(0.056)
College Degree or Higher	0.318***	0.224***	0.483***
	(0.017)	(0.062)	(0.069)
School Controls			
School Average Income	0.825***	0.395***	0.537***
	(0.041)	(0.105)	(0.107)
School-fixed Effects	No	No	No
F-Value	97.63	6.23	16.82
p > F	0.000	0.000	0.000
R^2	0.260	0.245	0.396
Observations	11824	747	816

† p<0.1 * p<0.05, ** p<.01, *** p<.001

Note: Table 7B extends the results depicted in Kuhn and Weinberger (2005: 405 (columns 5-8). This table differs from Table 4.7 reported in the paper due to the inclusion of school mean wages instead of school dummies to avoid problems arising from the small sample sizes for non-whites. All models include (unreported) grade and school attainment control variables. Models 3 and 4 include an unreported dummy variable for lack of parent's education. Model 4 includes an unreported dummy variable for lack of educational attainment. The sample for the models labelled 1,2, and 3 are based on the sample of white females, non-white males, and non-white females that were excluded in the original Kuhn and Weinberger (2005) study.

Table 9A: Instrumental Variable Regressions for new Samples – *First Stage Estimates*

	Sample: White Females			Sample: Non-White Males			Sample: Non-White Females		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	First-stage regression: Both Captain and President	First-stage regression: Captain Only	First-stage regression: President Only	First-stage regression: Both Captain and President	First-stage regression: Captain Only	First-stage regression: President Only	First-stage regression: Both Captain and President	First-stage regression: Captain Only	First-stage regression: President Only
Instruments									
Father in Club	0.006	-0.011	0.057*	0.151*	-0.044	-0.181*	-0.059	0.026	0.145
	(0.022)	(0.032)	(0.028)	(0.068)	(0.085)	(0.091)	(0.076)	(0.086)	(0.092)
Mother in Club	-0.004	0.003	0.060*	-0.093	0.001	-0.026	-0.039	-0.201**	-0.090
	(0.024)	(0.031)	(0.031)	(0.090)	(0.077)	(0.096)	(0.069)	(0.076)	(0.084)
Father in Team	0.039*	0.029	-0.012	0.061	0.019	0.094	0.153*	-0.033	0.183*
	(0.019)	(0.027)	(0.026)	(0.070)	(0.083)	(0.077)	(0.067)	(0.080)	(0.073)
Mother in Team	0.025	0.048	0.020	-0.003	-0.031	-0.170†	0.038	0.018	0.099
	(0.026)	(0.038)	(0.034)	(0.079)	(0.104)	(0.088)	(0.091)	(0.101)	(0.099)
Covariates									
Both on Team and in Club	0.155***	0.231***	0.083	0.182	0.116	-0.209	0.028	0.018	0.431*
	(0.027)	(0.044)	(0.068)	(0.132)	(0.160)	(0.190)	(0.140)	(0.175)	(0.167)
On Team only	-0.002	0.044	-0.145*	-0.110	-0.140	-0.513*	-0.293*	0.380†	0.946***
	(0.068)	(0.097)	(0.074)	(0.138)	(0.215)	(0.238)	(0.144)	(0.213)	(0.221)
In Club Only	0.014	-0.011	0.058	0.015	0.003	-0.328	-0.091	-0.312†	0.244
	(0.025)	(0.041)	(0.067)	(0.121)	(0.164)	(0.208)	(0.134)	(0.163)	(0.162)
Overweight	-0.021	0.009	-0.135*	0.042	-0.116	0.453†	-0.149†	-0.046	-0.178
	(0.041)	(0.063)	(0.056)	(0.162)	(0.225)	(0.232)	(0.078)	(0.091)	(0.120)
Underweight	-0.033	-0.039	0.029	0.128	-0.284**	0.232	-0.163†	0.040	-0.151
	(0.027)	(0.044)	(0.046)	(0.157)	(0.102)	(0.149)	(0.093)	(0.121)	(0.136)
Tall	0.113	-0.102	0.013	-0.017	-0.073	-0.157*	0.245	-0.407*	-0.180
	(0.115)	(0.108)	(0.101)	(0.081)	(0.082)	(0.078)	(0.240)	(0.169)	(0.200)
Short	0.008	-0.028	0.022	-0.004	0.107	-0.108	-0.015	-0.123	-0.045
	(0.016)	(0.024)	(0.022)	(0.066)	(0.083)	(0.071)	(0.069)	(0.083)	(0.089)
Dates	0.016**	0.011	0.028***	0.014	-0.034	-0.016	0.021	0.006	0.050
	(0.006)	(0.009)	(0.008)	(0.019)	(0.026)	(0.026)	(0.024)	(0.026)	(0.031)
Comfortable	0.030	-0.017	0.037	-0.041	-0.008	0.055	0.062	0.112	-0.019
	(0.019)	(0.028)	(0.026)	(0.080)	(0.075)	(0.076)	(0.065)	(0.077)	(0.071)
Wealthy	0.029	0.001	0.016	-0.116	-0.008	-0.111	0.011	0.050	-0.293**

	(0.027)	(0.043)	(0.040)	(0.084)	(0.101)	(0.090)	(0.084)	(0.092)	(0.095)
Sociability	0.014†	0.020†	0.034**	-0.055†	0.035	0.041	0.052	-0.049	0.029
	(0.008)	(0.012)	(0.013)	(0.032)	(0.036)	(0.042)	(0.034)	(0.031)	(0.036)
Vigor	0.042***	0.033*	0.004	0.056†	-0.052	-0.088*	0.016	0.049	0.073†
	(0.009)	(0.014)	(0.013)	(0.032)	(0.036)	(0.041)	(0.037)	(0.041)	(0.039)
Mature	0.004	-0.009	0.024†	0.017	0.078*	0.097*	0.038	-0.113*	-0.029
	(0.010)	(0.014)	(0.014)	(0.037)	(0.038)	(0.042)	(0.037)	(0.045)	(0.050)
Self-Confidence	0.016†	0.010	0.039**	0.018	0.043	0.015	0.014	0.003	-0.008
	(0.008)	(0.014)	(0.012)	(0.028)	(0.035)	(0.033)	(0.030)	(0.038)	(0.039)
Tidiness	-0.019*	0.010	-0.010	-0.006	-0.104*	0.010	-0.049	0.103*	0.088†
	(0.009)	(0.013)	(0.012)	(0.037)	(0.040)	(0.042)	(0.035)	(0.041)	(0.048)
Math Score	-0.000	0.000	0.002**	0.000	-0.005*	-0.005*	-0.001	0.001	0.000
	(0.000)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)
Vocabulary Score	-0.020	-0.117†	0.002	0.030	-0.048	-0.575**	-0.030	0.300†	0.163
	(0.041)	(0.062)	(0.059)	(0.181)	(0.212)	(0.186)	(0.142)	(0.169)	(0.188)
Social Studies Score	0.037	0.015	-0.009	-0.312†	0.220	0.324†	0.170	0.137	0.086
	(0.036)	(0.055)	(0.053)	(0.183)	(0.189)	(0.194)	(0.160)	(0.165)	(0.187)
Science Score	-0.005	0.068	0.005	0.147	0.173	0.257	-0.100	-0.220	-0.111
	(0.035)	(0.053)	(0.054)	(0.162)	(0.223)	(0.214)	(0.127)	(0.149)	(0.159)
Scientific Attitude Score	-0.003	0.064	-0.045	-0.020	0.105	0.018	-0.060	0.063	0.151
	(0.029)	(0.045)	(0.043)	(0.144)	(0.154)	(0.166)	(0.105)	(0.121)	(0.124)
Law Score	0.035	0.074†	0.006	-0.259*	-0.193	-0.027	0.051	-0.030	0.007
	(0.033)	(0.045)	(0.045)	(0.129)	(0.147)	(0.150)	(0.120)	(0.124)	(0.145)
Military Score	-0.049†	-0.039	0.010	0.154	-0.096	-0.185	0.059	0.058	-0.012
	(0.028)	(0.040)	(0.040)	(0.117)	(0.140)	(0.121)	(0.101)	(0.114)	(0.133)
Business Score	-0.025	-0.018	-0.026	0.364**	0.276*	0.041	-0.154	0.170	-0.138
	(0.030)	(0.043)	(0.044)	(0.131)	(0.140)	(0.151)	(0.109)	(0.116)	(0.127)
Etiquette Score	0.033	0.056	0.016	0.060	0.271†	-0.001	0.035	0.059	-0.127
	(0.031)	(0.044)	(0.041)	(0.102)	(0.148)	(0.113)	(0.110)	(0.131)	(0.134)
English Score	0.069†	-0.083	0.004	-0.114	0.020	0.491**	0.305†	-0.440*	-0.301†
	(0.038)	(0.052)	(0.054)	(0.168)	(0.200)	(0.183)	(0.157)	(0.181)	(0.182)
Science Awards	-0.033**	-0.033*	-0.042*	-0.007	0.086*	0.044	0.025	-0.017	0.029
	(0.010)	(0.016)	(0.017)	(0.021)	(0.037)	(0.035)	(0.027)	(0.033)	(0.040)
Fine Arts Awards	0.011*	-0.002	0.016*	0.029*	-0.041†	-0.024	0.018	0.045*	0.022
	(0.005)	(0.008)	(0.008)	(0.012)	(0.025)	(0.025)	(0.017)	(0.018)	(0.020)
Sports Awards	0.022***	0.023**	0.029***	0.029†	0.069***	0.066***	-0.023	0.057†	0.015
	(0.005)	(0.007)	(0.008)	(0.015)	(0.016)	(0.016)	(0.019)	(0.029)	(0.020)

Arithmetic Skills	-0.026	0.100*	-0.030	-0.146	-0.167	-0.027	-0.010	0.332*	-0.001
	(0.028)	(0.040)	(0.039)	(0.118)	(0.124)	(0.144)	(0.123)	(0.131)	(0.135)
Reading Skills	-0.004	0.007	-0.016	0.084	0.219†	0.099	-0.085	0.121	-0.185
	(0.025)	(0.037)	(0.037)	(0.097)	(0.116)	(0.120)	(0.105)	(0.114)	(0.123)
Clerical Skills	-0.024	-0.005	-0.007	0.233*	-0.115	0.096	-0.019	0.132	0.105
	(0.027)	(0.039)	(0.037)	(0.100)	(0.127)	(0.117)	(0.102)	(0.111)	(0.131)
Identification Skills	-0.017	0.005	0.009	-0.076	-0.131	-0.287*	0.035	-0.073	-0.002
	(0.027)	(0.040)	(0.038)	(0.103)	(0.117)	(0.125)	(0.116)	(0.120)	(0.107)
Public Service Interest	0.001	-0.006	0.006	0.013	-0.020	0.003	-0.044	-0.006	0.092*
	(0.009)	(0.013)	(0.013)	(0.034)	(0.038)	(0.041)	(0.034)	(0.040)	(0.039)
Business Management Interest	0.002	0.015	0.024†	-0.013	-0.005	0.026	-0.030	0.047	-0.069
	(0.010)	(0.013)	(0.013)	(0.035)	(0.036)	(0.043)	(0.036)	(0.039)	(0.046)
School-fixed Effects	No	No	No	No	No	No	No	No	No
F	1.95	0.91	2.94	1.92	0.1	1.86	1.76	2.06	3.86
p > F	0.09	0.45	0.02	0.10	0.98	0.12	0.13	0.09	0.00
Observations	11210	6102	6431	680	325	420	745	309	442

† p<0.1 * p<0.05, ** p<.01, *** p<.001

Note: Table 9A reports the first stage coefficient estimates for the dichotomous indicator variables whether individuals acted as captain and president, captain only, or president only derived from a conventional IV estimator (Stata; ivreg2). Abbreviated second stage results for the conventional IV estimates were reported in Table 4.9 in the paper. Information whether each parent was a member of a club of a team represents the instrumental variables. The sample for the treatment analysis is based on the sample of white females, non-white males and non-white females respectively.

Table 9B: Extension – *Full Second stage results for the IV regressions (conventional, conditional and heteroskedasticity-based) for white females*

	(1) Conventional	(2) Conventional	(3) Conventional	(4) Conditional	(5) Conditional	(6) Conditional	(7) Heteroske-dasticity	(8) Heteroske-dasticity	(9) Heteroske-dasticity
	Log(Hourly Earnings)	Log(Hourly Earnings)	Log(Hourly Earnings)	Log(Hourly Earnings)	Log(Hourly Earnings)	Log(Hourly Earnings)	Log(Hourly Earnings)	Log(Hourly Earnings)	Log(Hourly Earnings)
Instrumented Variable									
President and Captain	-0.046			0.355			-0.172		
	(0.337)			(0.268)			(0.130)		
Captain Only		0.456			0.164			0.231	
		(0.453)			(0.373)			(0.233)	
President Only			0.032			0.036			-0.033
			(0.249)			(0.113)			(0.115)
Covariates									
Both on Team and in Club	0.033	-0.117	-0.003	-0.025	-0.054	0.023	0.050	-0.068	0.010
	(0.068)	(0.118)	(0.055)	(0.053)	(0.092)	(0.045)	(0.045)	(0.072)	(0.046)
On Team only	-0.073	-0.126	-0.105	-0.014	-0.083	-0.045	-0.075	-0.106	-0.104
	(0.088)	(0.095)	(0.122)	(0.067)	(0.106)	(0.082)	(0.090)	(0.089)	(0.121)
In Club Only	-0.022	-0.042	-0.039	-0.006	-0.039	-0.006	-0.020	-0.042	-0.033
	(0.041)	(0.050)	(0.051)	(0.038)	(0.042)	(0.043)	(0.040)	(0.046)	(0.045)
Overweight	-0.003	0.009	-0.063	-0.049	-0.013	-0.056	-0.014	0.008	-0.074
	(0.049)	(0.043)	(0.055)	(0.033)	(0.042)	(0.038)	(0.050)	(0.042)	(0.046)
Underweight	0.016	0.057	0.008	0.027	0.020	0.021	0.009	0.031	0.009
	(0.032)	(0.046)	(0.033)	(0.019)	(0.031)	(0.022)	(0.030)	(0.040)	(0.032)
Tall	0.178†	0.166†	0.058	0.115	0.157	0.061	0.200†	0.140	0.051
	(0.104)	(0.101)	(0.083)	(0.073)	(0.103)	(0.089)	(0.107)	(0.091)	(0.084)
Short	0.017	0.030	0.033	0.011	0.009	0.017	0.016	0.015	0.031
	(0.015)	(0.025)	(0.021)	(0.009)	(0.012)	(0.011)	(0.015)	(0.021)	(0.019)
Dates	-0.027**	-0.035***	-0.034***	-0.038***	-0.035***	-0.039***	-0.026***	-0.036***	-0.033***
	(0.008)	(0.009)	(0.010)	(0.005)	(0.005)	(0.005)	(0.006)	(0.007)	(0.007)
Comfortable	-0.015	0.036	-0.024	0.045***	0.074***	0.041**	-0.011	0.028	-0.020
	(0.022)	(0.029)	(0.025)	(0.012)	(0.015)	(0.015)	(0.021)	(0.026)	(0.024)
Wealthy	0.059*	0.114**	0.065†	0.095***	0.131***	0.090***	0.061*	0.131***	0.074*
	(0.029)	(0.041)	(0.034)	(0.017)	(0.023)	(0.020)	(0.027)	(0.037)	(0.032)
Sociability	0.020†	-0.001	0.017	0.014*	0.015*	0.021**	0.021*	0.002	0.019†
	(0.010)	(0.014)	(0.014)	(0.006)	(0.007)	(0.007)	(0.009)	(0.012)	(0.011)

Vigor	-0.001	-0.018	0.001	-0.014	-0.012	-0.008	0.006	-0.012	-0.000
	(0.017)	(0.020)	(0.010)	(0.010)	(0.019)	(0.007)	(0.010)	(0.014)	(0.010)
Mature	0.020*	0.030*	0.014	0.012*	0.016*	0.018*	0.021*	0.035**	0.019†
	(0.010)	(0.014)	(0.012)	(0.006)	(0.008)	(0.008)	(0.010)	(0.013)	(0.011)
Self-Confidence	-0.000	-0.001	-0.001	-0.004	0.000	-0.002	0.001	0.001	0.003
	(0.009)	(0.013)	(0.014)	(0.006)	(0.007)	(0.007)	(0.008)	(0.011)	(0.011)
Tidiness	-0.001	-0.009	0.001	0.004	0.006	0.004	-0.004	-0.009	-0.002
	(0.011)	(0.013)	(0.011)	(0.006)	(0.007)	(0.006)	(0.009)	(0.012)	(0.011)
Math Score	0.002***	0.003***	0.003***	0.003***	0.003***	0.003***	0.002***	0.003***	0.003***
	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Vocabulary Score	0.150***	0.168*	0.129*	0.101***	0.101**	0.068*	0.147***	0.142*	0.119*
	(0.038)	(0.078)	(0.050)	(0.025)	(0.036)	(0.031)	(0.038)	(0.061)	(0.049)
Social Studies Score	0.120***	0.136**	0.110*	0.089***	0.123***	0.104***	0.122***	0.123**	0.115**
	(0.036)	(0.050)	(0.046)	(0.023)	(0.032)	(0.027)	(0.036)	(0.047)	(0.045)
Science Score	-0.032	-0.105†	-0.009	-0.015	-0.032	-0.022	-0.037	-0.062	-0.002
	(0.036)	(0.062)	(0.047)	(0.021)	(0.029)	(0.026)	(0.036)	(0.051)	(0.045)
Scientific Attitude Score	0.030	0.000	0.006	0.022	0.007	0.028	0.027	0.005	0.006
	(0.029)	(0.051)	(0.039)	(0.018)	(0.028)	(0.021)	(0.029)	(0.041)	(0.036)
Law Score	-0.010	-0.083	-0.031	-0.022	-0.030	-0.030	-0.009	-0.060	-0.034
	(0.033)	(0.056)	(0.040)	(0.018)	(0.030)	(0.023)	(0.030)	(0.044)	(0.039)
Military Score	-0.005	0.038	0.030	-0.009	-0.004	0.004	-0.010	0.041	0.033
	(0.032)	(0.045)	(0.035)	(0.017)	(0.021)	(0.020)	(0.029)	(0.039)	(0.035)
Business Score	0.008	0.018	-0.026	0.059***	0.051*	0.048*	0.000	-0.002	-0.030
	(0.033)	(0.047)	(0.038)	(0.018)	(0.022)	(0.022)	(0.031)	(0.041)	(0.037)
Etiquette Score	0.006	-0.027	-0.014	0.010	0.008	0.017	0.005	-0.001	-0.009
	(0.031)	(0.049)	(0.040)	(0.020)	(0.025)	(0.021)	(0.030)	(0.042)	(0.038)
English Score	0.018	0.082	0.046	0.019	0.051	0.016	0.040	0.061	0.045
	(0.041)	(0.060)	(0.044)	(0.021)	(0.034)	(0.027)	(0.036)	(0.048)	(0.042)
Science Awards	0.017	0.020	0.030	0.013	0.011	0.013	0.014	0.016	0.029
	(0.017)	(0.025)	(0.020)	(0.008)	(0.013)	(0.009)	(0.013)	(0.019)	(0.018)
Fine Arts Awards	-0.009	-0.012	-0.010	-0.008*	-0.008†	-0.004	-0.007	-0.014†	-0.010
	(0.006)	(0.009)	(0.008)	(0.004)	(0.004)	(0.005)	(0.005)	(0.008)	(0.007)
Sports Awards	0.004	-0.004	-0.006	-0.006	0.001	-0.002	0.007	0.002	-0.002
	(0.009)	(0.013)	(0.011)	(0.007)	(0.011)	(0.004)	(0.006)	(0.009)	(0.008)
Arithmetic Skills	-0.028	-0.087	-0.054†	-0.019	-0.045*	-0.053**	-0.035	-0.083*	-0.050
	(0.028)	(0.059)	(0.033)	(0.019)	(0.023)	(0.020)	(0.026)	(0.039)	(0.031)
Reading Skills	0.014	-0.008	0.020	0.004	-0.009	0.031†	0.007	-0.016	0.017

	(0.025)	(0.037)	(0.031)	(0.016)	(0.023)	(0.019)	(0.025)	(0.034)	(0.031)
Clerical Skills	0.004	0.024	0.013	0.022	0.023	0.035†	0.002	0.028	0.003
	(0.026)	(0.037)	(0.032)	(0.016)	(0.022)	(0.019)	(0.026)	(0.034)	(0.031)
Identification Skills	0.013	0.029	-0.009	0.003	0.027	-0.010	0.007	0.039	-0.006
	(0.026)	(0.038)	(0.032)	(0.016)	(0.021)	(0.019)	(0.026)	(0.034)	(0.032)
Public Service Interest	0.029***	0.023†	0.041***	0.031***	0.036**	0.039***	0.026**	0.019†	0.040***
	(0.009)	(0.013)	(0.011)	(0.006)	(0.011)	(0.007)	(0.009)	(0.011)	(0.011)
Business Management Interest	0.002	-0.000	0.005	-0.005	-0.009	-0.006	0.005	0.005	0.006
	(0.009)	(0.014)	(0.013)	(0.005)	(0.010)	(0.007)	(0.009)	(0.011)	(0.011)
School-fixed effects	No	No	No	No	No	No	No	No	No
Second Stage F-Statistic	18.53	9.39	13.87	47.87	31.52	35.27	18.25	12.08	14.87
p > F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Observations	11210	6102	6431	11210	6102	6431	11210	6102	6431

† p<0.1 * p<0.05, ** p<.01, *** p<.001

Note: Table 9B reports the second stage coefficients and standard deviations from a conventional, a conditional and a heteroskedasticity-based IV estimation. Abbreviated second stage results for the conventional IV estimates were reported in Table 4.9 in the paper. The instrumented variables are whether individuals acted as captain and president, captain only, or president. Information whether each parent was a member of a club of a team represents the instrumental variables. The sample for the treatment analysis is based on the sample of white females.

Table 9C: Extension – *Full Second stage results for the IV regressions (conventional, conditional and heteroskedasticity-based) for non-white males*

	(1) Conventional	(2) Conventional	(3) Conventional	(4) Conditional	(5) Conditional	(6) Conditional	(7) Heteroske-dasticity	(8) Heteroske-dasticity	(9) Heteroske-dasticity
	Log(Hourly Earnings)	Log(Hourly Earnings)	Log(Hourly Earnings)	Log(Hourly Earnings)	Log(Hourly Earnings)	Log(Hourly Earnings)	Log(Hourly Earnings)	Log(Hourly Earnings)	Log(Hourly Earnings)
Instrumented Variable									
President and Captain	0.860*			-0.158			0.189†		
	(0.361)			(1.466)			(0.098)		
Captain Only		-2.543			0.159			0.021	
		(4.348)			(0.486)			(0.197)	
President Only			-0.392			-0.203			-0.206
			(0.311)			(0.558)			(0.204)
Covariates									
Both on Team and in Club	-0.142	0.122	-0.028	0.078	-0.083	0.034	0.023	-0.149	0.015
	(0.179)	(0.621)	(0.158)	(0.248)	(0.191)	(0.137)	(0.161)	(0.115)	(0.129)
On Team only	0.235	-0.390	0.071	0.103	-0.013	0.076	0.244	0.007	0.089
	(0.189)	(0.865)	(0.273)	(0.220)	(0.247)	(0.195)	(0.191)	(0.137)	(0.208)
In Club Only	-0.007	-0.192	-0.135	-0.013	-0.131	-0.059	0.037	-0.206	-0.084
	(0.164)	(0.467)	(0.201)	(0.143)	(0.155)	(0.151)	(0.171)	(0.130)	(0.156)
Overweight	0.092	-0.522	0.089	0.100	-0.053	0.169	0.174	-0.184	-0.011
	(0.139)	(0.741)	(0.191)	(0.275)	(0.164)	(0.127)	(0.190)	(0.161)	(0.167)
Underweight	-0.339*	-0.650	-0.150	0.023	0.166	0.055	-0.241**	0.077	-0.194
	(0.133)	(1.283)	(0.131)	(0.112)	(0.177)	(0.102)	(0.089)	(0.133)	(0.119)
Tall	0.105	-0.003	0.068	-0.002	0.016	-0.052	0.098†	0.135*	0.099
	(0.078)	(0.371)	(0.084)	(0.117)	(0.067)	(0.099)	(0.052)	(0.064)	(0.077)
Short	-0.015	0.369	-0.014	-0.018	-0.004	-0.047	-0.022	0.064	0.012
	(0.065)	(0.518)	(0.061)	(0.043)	(0.053)	(0.089)	(0.046)	(0.055)	(0.056)
Dates	-0.035†	-0.089	0.003	-0.002	-0.017	0.029	-0.020	-0.013	0.006
	(0.020)	(0.187)	(0.029)	(0.014)	(0.021)	(0.028)	(0.014)	(0.017)	(0.026)
Comfortable	0.031	-0.054	0.067	0.069†	0.067	0.064	0.061	0.001	0.036
	(0.065)	(0.199)	(0.068)	(0.040)	(0.059)	(0.071)	(0.047)	(0.049)	(0.062)
Wealthy	0.226**	0.031	0.079	0.116†	0.075	0.114	0.220***	0.063	0.087
	(0.081)	(0.251)	(0.094)	(0.062)	(0.074)	(0.099)	(0.053)	(0.065)	(0.083)
Sociability	0.069*	0.128	0.026	0.012	0.042	-0.001	0.042†	0.048*	0.007

	(0.035)	(0.158)	(0.036)	(0.073)	(0.030)	(0.042)	(0.023)	(0.024)	(0.031)
Vigor	-0.025	-0.101	-0.042	0.027	0.006	-0.030	0.004	0.010	-0.035
	(0.036)	(0.216)	(0.047)	(0.092)	(0.029)	(0.029)	(0.028)	(0.026)	(0.035)
Mature	-0.067*	0.144	0.030	-0.021	-0.011	0.030	-0.049†	-0.040	0.033
	(0.034)	(0.349)	(0.054)	(0.025)	(0.038)	(0.051)	(0.027)	(0.029)	(0.044)
Self-Confidence	0.009	0.102	-0.017	0.015	-0.022	0.025	0.006	-0.010	-0.018
	(0.030)	(0.199)	(0.032)	(0.019)	(0.025)	(0.040)	(0.023)	(0.026)	(0.029)
Tidiness	0.029	-0.238	0.050	0.034	0.023	0.066†	0.021	0.017	0.042
	(0.032)	(0.471)	(0.033)	(0.021)	(0.028)	(0.039)	(0.024)	(0.030)	(0.030)
Math Score	0.001	-0.008	0.001	0.003**	0.004*	0.003*	0.002†	0.004**	0.002
	(0.002)	(0.020)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Vocabulary Score	0.248	-0.214	0.055	0.083	-0.195	0.020	0.269**	-0.134	0.114
	(0.155)	(0.544)	(0.247)	(0.184)	(0.153)	(0.157)	(0.101)	(0.123)	(0.202)
Social Studies Score	0.353†	0.580	0.227	0.170	0.079	0.217	0.177	-0.017	0.156
	(0.181)	(1.038)	(0.215)	(0.177)	(0.158)	(0.147)	(0.120)	(0.140)	(0.190)
Science Score	-0.233	0.186	0.003	-0.057	-0.083	-0.017	-0.146	-0.241	0.018
	(0.169)	(0.775)	(0.189)	(0.158)	(0.135)	(0.139)	(0.102)	(0.147)	(0.164)
Scientific Attitude Score	-0.017	0.235	-0.059	-0.035	-0.084	-0.040	-0.076	-0.058	-0.065
	(0.141)	(0.617)	(0.159)	(0.079)	(0.107)	(0.106)	(0.087)	(0.109)	(0.141)
Law Score	0.386*	-0.371	0.230†	0.077	0.120	0.201*	0.179*	0.182†	0.182†
	(0.163)	(0.828)	(0.119)	(0.176)	(0.110)	(0.101)	(0.078)	(0.104)	(0.103)
Military Score	0.021	0.085	0.026	0.090	0.228*	0.046	0.166*	0.298***	0.064
	(0.116)	(0.496)	(0.121)	(0.103)	(0.093)	(0.098)	(0.074)	(0.090)	(0.107)
Business Score	-0.324†	0.905	-0.129	0.010	0.153	-0.129	-0.081	0.203†	-0.115
	(0.174)	(1.313)	(0.145)	(0.261)	(0.101)	(0.095)	(0.088)	(0.113)	(0.129)
Etiquette Score	-0.138	0.574	-0.127	-0.025	-0.113	-0.023	-0.101	-0.141	-0.114
	(0.102)	(1.348)	(0.110)	(0.096)	(0.104)	(0.079)	(0.076)	(0.092)	(0.098)
English Score	0.108	0.081	0.022	-0.019	0.057	0.058	0.037	-0.017	-0.064
	(0.175)	(0.496)	(0.250)	(0.119)	(0.137)	(0.183)	(0.131)	(0.133)	(0.201)
Science Awards	0.046	0.318	0.036	0.019	0.061†	0.012	0.060**	0.104***	0.038
	(0.031)	(0.368)	(0.031)	(0.025)	(0.033)	(0.030)	(0.023)	(0.029)	(0.028)
Fine Arts Awards	-0.044*	-0.152	-0.036†	-0.011	-0.028	-0.016	-0.029*	-0.053**	-0.043*
	(0.021)	(0.188)	(0.021)	(0.012)	(0.026)	(0.018)	(0.014)	(0.019)	(0.018)
Sports Awards	-0.044**	0.179	-0.004	0.001	-0.005	-0.007	-0.025**	0.005	-0.017
	(0.017)	(0.298)	(0.025)	(0.051)	(0.030)	(0.021)	(0.009)	(0.017)	(0.019)
Arithmetic Skills	-0.073	-0.805	-0.132	-0.160	-0.289**	-0.032	-0.209*	-0.364***	-0.129
	(0.133)	(0.816)	(0.127)	(0.188)	(0.093)	(0.089)	(0.084)	(0.093)	(0.108)

Reading Skills	-0.024	0.635	0.100	0.063	0.006	-0.005	0.067	0.009	0.027
	(0.109)	(1.073)	(0.116)	(0.082)	(0.086)	(0.080)	(0.075)	(0.086)	(0.100)
Clerical Skills	-0.166	-0.241	0.012	-0.032	-0.055	-0.107	-0.078	0.066	0.019
	(0.122)	(0.620)	(0.103)	(0.260)	(0.088)	(0.085)	(0.072)	(0.074)	(0.091)
Identification Skills	0.107	-0.233	-0.008	0.079	0.105	0.017	0.080	0.094	0.015
	(0.108)	(0.611)	(0.147)	(0.063)	(0.102)	(0.149)	(0.070)	(0.086)	(0.116)
Public Service Interest	0.021	-0.016	0.035	0.022	0.031	0.013	0.020	0.043	0.047
	(0.033)	(0.160)	(0.035)	(0.031)	(0.042)	(0.030)	(0.023)	(0.026)	(0.032)
Business Management Interest	-0.018	-0.130	-0.063	-0.028	-0.066†	-0.038	-0.034	-0.103***	-0.064†
	(0.034)	(0.094)	(0.038)	(0.036)	(0.035)	(0.037)	(0.025)	(0.027)	(0.034)
School-fixed effects	No	No	No	No	No	No	No	No	No
Second Stage F-Statistic	3.44	0.48	3.51	4.01	2.20	2.79	5.78	7.11	3.93
p > F	0.000	0.995	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Observations	680	325	420	680	325	420	680	325	420

† p<0.1 * p<0.05, ** p<.01, *** p<.001

Note: Table 9C reports the second stage coefficients and standard deviations from a conventional, a conditional and a heteroskedasticity-based IV estimation. Abbreviated second stage results for the conventional IV estimates were reported in Table 4.9 in the paper. The instrumented variables are whether individuals acted as captain and president, captain only, or president. Information whether each parent was a member of a club of a team represents the instrumental variables. The sample for the treatment analysis is based on the sample of non-white males.

Table 9D: Extension – *Full Second stage results for the IV regressions (conventional, conditional and heteroskedasticity-based) for non-white females*

	(1) Conventional Log(Hourly Earnings)	(2) Conventional Log(Hourly Earnings)	(3) Conventional Log(Hourly Earnings)	(4) Conditional Log(Hourly Earnings)	(5) Conditional Log(Hourly Earnings)	(6) Conditional Log(Hourly Earnings)	(7) Heteroske-dasticity Log(Hourly Earnings)	(8) Heteroske-dasticity Log(Hourly Earnings)	(9) Heteroske-dasticity Log(Hourly Earnings)
Instrumented Variable									
President and Captain	0.588†			0.113			0.369†		
	(0.352)			(0.322)			(0.208)		
Captain Only		0.276			0.053			0.197	
		(0.440)			(0.375)			(0.343)	
President Only			0.404			0.030			0.435*
			(0.305)			(0.319)			(0.194)
Covariates									
Both on Team and in Club	-0.355**	-0.332†	-0.354†	-0.089	-0.119	-0.018	-0.375***	-0.390*	-0.340*
	(0.118)	(0.173)	(0.182)	(0.217)	(0.241)	(0.282)	(0.104)	(0.171)	(0.166)
On Team only	-0.102	-0.539*	-0.373	-0.097	-0.201	-0.029	-0.227†	-0.411†	-0.522*
	(0.169)	(0.244)	(0.283)	(0.322)	(0.544)	(0.490)	(0.129)	(0.232)	(0.257)
In Club Only	-0.195	-0.186	-0.159	-0.068	-0.101	0.023	-0.234*	-0.193	-0.218
	(0.126)	(0.229)	(0.156)	(0.205)	(0.205)	(0.262)	(0.106)	(0.158)	(0.188)
Overweight	0.115	0.148	0.077	-0.130†	-0.243*	-0.181	0.053	0.088	0.148
	(0.125)	(0.112)	(0.133)	(0.078)	(0.109)	(0.113)	(0.111)	(0.124)	(0.112)
Underweight	0.173†	-0.153	0.102	0.045	-0.141	0.001	0.124†	0.094	-0.152
	(0.091)	(0.097)	(0.097)	(0.076)	(0.107)	(0.114)	(0.070)	(0.086)	(0.096)
Tall	-0.277	0.189	-0.188	0.041	0.055	0.003	-0.178	-0.120	0.152
	(0.219)	(0.308)	(0.180)	(0.240)	(0.337)	(0.354)	(0.154)	(0.166)	(0.275)
Short	0.010	0.153	0.006	0.028	0.046	0.053	0.007	-0.003	0.143†
	(0.058)	(0.100)	(0.061)	(0.042)	(0.075)	(0.058)	(0.046)	(0.062)	(0.084)
Dates	-0.031	0.014	-0.058*	-0.043**	-0.041	-0.082***	-0.040*	-0.069**	0.013
	(0.023)	(0.032)	(0.028)	(0.015)	(0.027)	(0.024)	(0.019)	(0.026)	(0.032)
Comfortable	0.042	-0.021	0.038	0.031	0.045	0.069	0.041	0.072	-0.011
	(0.077)	(0.107)	(0.086)	(0.041)	(0.060)	(0.053)	(0.062)	(0.076)	(0.115)
Wealthy	0.083	-0.003	0.254	0.080	0.054	0.158*	0.031	0.238†	0.000
	(0.094)	(0.139)	(0.167)	(0.055)	(0.085)	(0.075)	(0.073)	(0.128)	(0.139)
Sociability	-0.088*	-0.148**	-0.079†	-0.043†	-0.033	-0.065*	-0.079*	-0.077*	-0.152***

	(0.040)	(0.051)	(0.041)	(0.022)	(0.034)	(0.028)	(0.033)	(0.039)	(0.046)
Vigor	-0.000	0.107†	-0.036	0.009	0.007	-0.001	0.007	-0.042	0.109†
	(0.031)	(0.060)	(0.053)	(0.021)	(0.033)	(0.033)	(0.026)	(0.037)	(0.057)
Mature	0.073†	0.083	0.080†	0.043†	-0.002	0.038	0.086*	0.069	0.074
	(0.041)	(0.071)	(0.046)	(0.023)	(0.040)	(0.032)	(0.035)	(0.049)	(0.060)
Self-Confidence	-0.032	0.025	-0.016	-0.005	0.035	-0.005	-0.019	-0.020	0.024
	(0.035)	(0.040)	(0.034)	(0.019)	(0.031)	(0.029)	(0.027)	(0.031)	(0.040)
Tidiness	0.020	-0.091	-0.014	-0.004	-0.002	0.011	0.007	0.003	-0.081
	(0.040)	(0.056)	(0.042)	(0.023)	(0.041)	(0.033)	(0.033)	(0.038)	(0.054)
Math Score	0.002	-0.002	0.000	0.002*	-0.001	0.002	0.002	0.000	-0.002
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
Vocabulary Score	0.189	-0.243	-0.030	0.032	-0.024	-0.127	0.168	-0.141	-0.216
	(0.155)	(0.224)	(0.158)	(0.097)	(0.161)	(0.128)	(0.130)	(0.152)	(0.195)
Social Studies Score	0.014	0.485**	0.108	0.222*	0.407**	0.278*	0.039	0.134	0.493**
	(0.180)	(0.186)	(0.161)	(0.092)	(0.143)	(0.126)	(0.145)	(0.151)	(0.183)
Science Score	0.130	0.137	0.049	0.049	-0.083	-0.106	0.100	-0.007	0.123
	(0.122)	(0.182)	(0.147)	(0.081)	(0.124)	(0.112)	(0.098)	(0.144)	(0.177)
Scientific Attitude Score	0.304**	0.159	0.245*	0.116	0.157	0.132	0.278**	0.210†	0.163
	(0.098)	(0.132)	(0.108)	(0.072)	(0.112)	(0.094)	(0.088)	(0.108)	(0.134)
Law Score	0.113	0.265*	0.237*	0.064	0.134	0.048	0.086	0.242*	0.262*
	(0.109)	(0.124)	(0.103)	(0.083)	(0.117)	(0.113)	(0.084)	(0.108)	(0.125)
Military Score	0.015	0.015	0.080	-0.044	0.081	0.011	0.039	0.078	0.017
	(0.104)	(0.136)	(0.119)	(0.069)	(0.097)	(0.088)	(0.088)	(0.112)	(0.137)
Business Score	0.050	-0.085	0.038	-0.023	-0.021	-0.031	0.026	-0.016	-0.072
	(0.124)	(0.151)	(0.120)	(0.069)	(0.127)	(0.096)	(0.101)	(0.117)	(0.138)
Etiquette Score	0.089	0.229†	0.082	-0.007	0.037	-0.056	0.128	0.097	0.234†
	(0.103)	(0.131)	(0.109)	(0.065)	(0.102)	(0.095)	(0.086)	(0.105)	(0.133)
English Score	-0.068	0.485	0.292	0.254*	0.360†	0.394**	0.018	0.393†	0.442
	(0.186)	(0.300)	(0.252)	(0.107)	(0.185)	(0.139)	(0.146)	(0.222)	(0.269)
Science Awards	0.022	0.065	0.047	0.021	0.025	0.022	0.034	0.044	0.062
	(0.028)	(0.047)	(0.044)	(0.020)	(0.035)	(0.029)	(0.023)	(0.036)	(0.049)
Fine Arts Awards	-0.018	-0.054	-0.060*	-0.008	0.012	-0.029	-0.016	-0.067**	-0.051
	(0.016)	(0.035)	(0.027)	(0.012)	(0.022)	(0.018)	(0.013)	(0.024)	(0.035)
Sports Awards	0.008	-0.010	0.017	0.009	-0.007	0.047*	-0.001	0.028	-0.004
	(0.020)	(0.039)	(0.023)	(0.013)	(0.024)	(0.020)	(0.017)	(0.021)	(0.039)
Arithmetic Skills	-0.126	-0.336	-0.069	-0.052	-0.228†	-0.048	-0.126	-0.053	-0.308
	(0.104)	(0.217)	(0.139)	(0.070)	(0.128)	(0.119)	(0.091)	(0.122)	(0.195)

Reading Skills	0.171†	0.152	0.136	0.067	0.046	0.037	0.169*	0.126	0.157
	(0.098)	(0.116)	(0.118)	(0.065)	(0.102)	(0.088)	(0.083)	(0.104)	(0.116)
Clerical Skills	0.070	-0.065	-0.124	-0.033	-0.072	0.046	0.100	-0.071	-0.054
	(0.106)	(0.128)	(0.120)	(0.064)	(0.124)	(0.112)	(0.090)	(0.105)	(0.132)
Identification Skills	0.076	0.196	0.188	0.125†	0.115	0.135	0.074	0.183†	0.190
	(0.099)	(0.137)	(0.116)	(0.065)	(0.123)	(0.086)	(0.083)	(0.108)	(0.137)
Public Service Interest	0.090*	-0.038	-0.015	0.015	-0.007	-0.007	0.071*	0.007	-0.040
	(0.044)	(0.065)	(0.040)	(0.022)	(0.038)	(0.037)	(0.036)	(0.037)	(0.062)
Business Management Interest	-0.024	0.051	0.023	0.017	-0.001	0.022	-0.025	0.010	0.057
	(0.038)	(0.058)	(0.041)	(0.025)	(0.035)	(0.033)	(0.031)	(0.040)	(0.047)
School-fixed effects	No	No	No	No	No	No	No	No	No
Second Stage F-Statistic	5.11	6.37	7.49	6.33	2.89	4.00	8.76	6.27	6.56
p > F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Observations	745	309	442	745	309	442	745	309	442

† p<0.1 * p<0.05, ** p<.01, *** p<.001

Note: Table 9D reports the second stage coefficients and standard deviations from a conventional, a conditional and a heteroskedasticity-based IV estimation. Abbreviated second stage results for the conventional IV estimates were reported in Table 4.9 in the paper. The instrumented variables are whether individuals acted as captain and president, captain only, or president. Information whether each parent was a member of a club of a team represents the instrumental variables. The sample for the treatment analysis is based on the sample of non-white females.

Table 9E: Extension – *Full Second stage results for conventional IV-regressions with school dummies for new samples*

	Sample: White Females			Sample: Non-White Males			Sample: Non-White Females		
	Second Stage Regression: Log(Hourly Earnings)	Second Stage Regression: Log(Hourly Earnings)	Second Stage Regression: Log(Hourly Earnings)	Second Stage Regression: Log(Hourly Earnings)	Second Stage Regression: Log(Hourly Earnings)	Second Stage Regression: Log(Hourly Earnings)	Second Stage Regression: Log(Hourly Earnings)	Second Stage Regression: Log(Hourly Earnings)	Second Stage Regression: Log(Hourly Earnings)
Instrumented Variable									
President and Captain	-0.039			1.072			0.194		
	(0.253)			(0.712)			(0.213)		
Captain Only		0.183			0.170			0.524	
		(0.285)			(0.212)			(1.394)	
President Only			-0.039			-0.023			0.258
			(0.207)			(0.205)			(0.236)
Covariates									
Both on Team and in Club	0.092†	0.022	0.073	-0.150	0.234	0.263**	-0.306**	-0.028	-0.137
	(0.048)	(0.081)	(0.056)	(0.354)	(0.206)	(0.098)	(0.107)	(0.373)	(0.321)
On Team only	-0.057	-0.006	0.017	0.360	0.448*	0.386*	-0.100	-0.755	0.006
	(0.101)	(0.090)	(0.096)	(0.227)	(0.204)	(0.195)	(0.223)	(1.721)	(0.355)
In Club Only	0.057	0.044	0.050	0.002	0.096	0.218†	-0.250*	0.007	0.030
	(0.036)	(0.041)	(0.047)	(0.284)	(0.192)	(0.125)	(0.105)	(0.717)	(0.321)
Overweight	0.011	0.031	-0.045	0.174	-0.005	0.087	-0.184*	-0.127	-0.142
	(0.048)	(0.041)	(0.040)	(0.171)	(0.162)	(0.134)	(0.086)	(0.338)	(0.117)
Underweight	0.021	0.064*	0.037	-0.325*	0.012	-0.095	0.105	-0.205	0.022
	(0.027)	(0.031)	(0.026)	(0.157)	(0.196)	(0.165)	(0.084)	(0.136)	(0.145)
Tall	0.124†	0.152	0.089	-0.071	0.170*	0.111	-0.125		-0.092
	(0.065)	(0.104)	(0.095)	(0.109)	(0.077)	(0.119)	(0.178)		(0.282)
Short	0.014	0.022	0.022	-0.028	-0.017	0.024	0.052	0.125	0.021
	(0.013)	(0.017)	(0.016)	(0.079)	(0.069)	(0.061)	(0.058)	(0.399)	(0.066)
Dates	-0.023***	-0.027***	-0.027***	-0.056	0.022	0.046*	-0.065***	-0.087	-0.108***
	(0.006)	(0.006)	(0.007)	(0.034)	(0.032)	(0.020)	(0.019)	(0.130)	(0.026)
Comfortable	-0.017	0.032†	-0.007	0.144	0.012	0.123†	0.010	0.091	-0.001
	(0.018)	(0.019)	(0.018)	(0.115)	(0.058)	(0.063)	(0.054)	(0.217)	(0.069)
Wealthy	0.046†	0.097**	0.060*	0.332*	0.067	0.218**	0.103	0.293	0.288**
	(0.025)	(0.036)	(0.028)	(0.164)	(0.076)	(0.083)	(0.070)	(0.215)	(0.097)
Sociability	0.021*	0.006	0.020	0.109†	0.034	0.000	-0.054†	-0.032	-0.070†
	(0.009)	(0.010)	(0.013)	(0.060)	(0.030)	(0.029)	(0.032)	(0.049)	(0.039)

Vigor	0.011	0.004	0.003	-0.027	-0.045	-0.009	0.043†	0.034	0.041
	(0.013)	(0.015)	(0.010)	(0.050)	(0.035)	(0.028)	(0.023)	(0.059)	(0.036)
Mature	0.015†	0.022*	0.012	-0.070†	-0.073*	0.012	0.022	0.084	0.013
	(0.008)	(0.011)	(0.009)	(0.042)	(0.032)	(0.036)	(0.025)	(0.116)	(0.043)
Self-Confidence	-0.004	-0.005	0.002	0.032	-0.010	-0.019	0.017	0.036	0.014
	(0.007)	(0.010)	(0.010)	(0.031)	(0.024)	(0.022)	(0.028)	(0.069)	(0.031)
Tidiness	-0.001	-0.005	-0.001	0.036	0.090**	0.030	0.010	-0.087	0.002
	(0.008)	(0.009)	(0.010)	(0.038)	(0.035)	(0.030)	(0.031)	(0.121)	(0.039)
Math Score	0.002***	0.003***	0.003***	0.001	0.005**	0.003*	0.002†	0.000	0.001
	(0.000)	(0.000)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.005)	(0.002)
Vocabulary Score	0.053	0.031	0.022	0.037	-0.202	0.162	0.318*	0.159	0.148
	(0.033)	(0.051)	(0.044)	(0.196)	(0.193)	(0.189)	(0.129)	(0.203)	(0.179)
Social Studies Score	0.099**	0.133***	0.070†	0.428	0.063	0.122	0.226†	0.395	0.218
	(0.031)	(0.039)	(0.041)	(0.320)	(0.206)	(0.179)	(0.125)	(0.274)	(0.177)
Science Score	0.023	-0.013	0.039	-0.271	-0.263	-0.104	0.144	0.077	0.016
	(0.032)	(0.040)	(0.037)	(0.182)	(0.173)	(0.156)	(0.099)	(0.301)	(0.156)
Scientific Attitude Score	0.036	0.018	0.005	0.065	0.104	-0.012	0.206**	0.089	0.098
	(0.022)	(0.035)	(0.028)	(0.146)	(0.110)	(0.113)	(0.080)	(0.322)	(0.120)
Law Score	0.008	-0.048	-0.005	0.400*	0.174	0.205†	0.097	-0.232	0.122
	(0.027)	(0.036)	(0.033)	(0.189)	(0.106)	(0.108)	(0.090)	(0.178)	(0.119)
Military Score	0.001	0.023	0.033	-0.085	0.116	0.023	0.021	0.186	0.093
	(0.028)	(0.035)	(0.030)	(0.164)	(0.135)	(0.110)	(0.091)	(0.223)	(0.120)
Business Score	-0.009	0.024	-0.004	-0.091	0.316*	-0.037	-0.108	-0.050	-0.076
	(0.026)	(0.032)	(0.030)	(0.181)	(0.148)	(0.103)	(0.094)	(0.178)	(0.117)
Etiquette Score	-0.002	-0.013	0.010	-0.165	-0.081	-0.164†	0.022	-0.041	0.070
	(0.025)	(0.032)	(0.031)	(0.130)	(0.121)	(0.086)	(0.080)	(0.197)	(0.111)
English Score	0.058†	0.078†	0.069†	0.095	-0.258†	-0.054	0.053	0.791	0.273
	(0.034)	(0.040)	(0.035)	(0.221)	(0.140)	(0.151)	(0.139)	(0.917)	(0.198)
Science Awards	-0.004	0.001	0.007	0.025	0.087**	0.006	0.025	-0.013	0.030
	(0.013)	(0.021)	(0.017)	(0.032)	(0.033)	(0.034)	(0.019)	(0.111)	(0.033)
Fine Arts Awards	0.004	-0.003	0.000	-0.007	-0.050**	-0.034†	-0.017	0.006	-0.014
	(0.005)	(0.007)	(0.008)	(0.022)	(0.017)	(0.019)	(0.011)	(0.062)	(0.021)
Sports Awards	0.001	-0.003	0.002	-0.055	0.002	-0.001	0.006	-0.032	0.000
	(0.007)	(0.010)	(0.008)	(0.035)	(0.015)	(0.015)	(0.015)	(0.050)	(0.024)
Arithmetic Skills	-0.011	-0.043	-0.030	-0.179	-0.217*	-0.040	-0.149†	-0.353*	-0.177
	(0.022)	(0.044)	(0.027)	(0.151)	(0.092)	(0.108)	(0.081)	(0.159)	(0.124)
Reading Skills	0.001	-0.020	0.008	-0.163	-0.202*	-0.096	0.105	-0.054	0.147

	(0.022)	(0.033)	(0.027)	(0.124)	(0.093)	(0.083)	(0.081)	(0.150)	(0.118)
Clerical Skills	-0.010	0.044†	0.033	-0.113	0.018	0.011	0.017	-0.244	-0.203
	(0.021)	(0.027)	(0.027)	(0.124)	(0.104)	(0.087)	(0.088)	(0.306)	(0.131)
Identification Skills	-0.013	0.014	-0.023	0.187	0.094	0.061	0.064	0.127	0.235†
	(0.022)	(0.026)	(0.026)	(0.124)	(0.083)	(0.099)	(0.089)	(0.177)	(0.134)
Public Service Interest	0.018*	0.016†	0.031**	0.010	0.002	0.032	-0.010	-0.124	-0.025
	(0.007)	(0.009)	(0.010)	(0.041)	(0.038)	(0.032)	(0.030)	(0.103)	(0.040)
Business Management Interest	0.003	0.008	0.010	-0.011	-0.108**	-0.080*	0.055†	0.121	0.050
	(0.007)	(0.010)	(0.010)	(0.050)	(0.038)	(0.032)	(0.032)	(0.101)	(0.046)
School-fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
First stage (Kleibergen-Paap) F-Statistic	3.44	2.05	5.30	0.69	3.37	2.57	3.46	0.18	4.30
p-value (F-Statistic)	0.008	0.085	0.000	0.597	0.011	0.038	0.008	0.946	0.000
Cragg-Donald-Wald F-Statistic	8.81	4.66	7.77	1.09	3.26	3.04	4.54	0.15	4.52
Sargan-Hansen statistic	6.01	4.42	0.55	1.34	12.26	13.14	3.49	n/a	n/a
p-value (Sargan-Hansen statistic)	0.11	0.22	0.91	0.72	0.01	0.00	0.32	n/a	n/a
Observations	11147	6005	6346	596	247	342	649	221	351

† p<0.1 * p<0.05, ** p<.01, *** p<.001

Note: Table 9E reports the second stage coefficients and standard deviations from IV estimations. The instrumented variables are whether individuals acted as captain and president, captain only, or president. Information whether each parent was a member of a club of a team represents the instrumental variables. The sample for the treatment analysis are white females, non-white males and non-white females respectively. For the last two columns of the non-white female sample, due to the large number of variables the Sargan-Hansen statistic cannot be directly estimated, as the covariance matrix of moment conditions is not of full rank. This table differs from Table 4.9 as it includes school dummies.

Table 9F: Extension – *Full Second stage results for conventional IV-regressions with robust standard errors for new samples*

	Sample: White Females			Sample: Non-White Males			Sample: Non-White Females		
	Second Stage Regression: Log(Hourly Earnings)	Second Stage Regression: Log(Hourly Earnings)	Second Stage Regression: Log(Hourly Earnings)	Second Stage Regression: Log(Hourly Earnings)	Second Stage Regression: Log(Hourly Earnings)	Second Stage Regression: Log(Hourly Earnings)	Second Stage Regression: Log(Hourly Earnings)	Second Stage Regression: Log(Hourly Earnings)	Second Stage Regression: Log(Hourly Earnings)
Instrumented Variable									
President and Captain	-0.046			0.860*			0.588†		
	(0.337)			(0.361)			(0.352)		
Captain Only		0.456			-2.543			0.276	
		(0.453)			(4.348)			(0.440)	
President Only			0.032			-0.392			0.404
			(0.249)			(0.311)			(0.305)
Covariates									
Both on Team and in Club	0.033	-0.117	-0.003	-0.142	0.122	-0.028	-0.355**	-0.332†	-0.354†
	(0.068)	(0.118)	(0.055)	(0.179)	(0.621)	(0.158)	(0.118)	(0.173)	(0.182)
On Team only	-0.073	-0.126	-0.105	0.235	-0.390	0.071	-0.102	-0.539*	-0.373
	(0.088)	(0.095)	(0.122)	(0.189)	(0.865)	(0.273)	(0.169)	(0.244)	(0.283)
In Club Only	-0.022	-0.042	-0.039	-0.007	-0.192	-0.135	-0.195	-0.186	-0.159
	(0.041)	(0.050)	(0.051)	(0.164)	(0.467)	(0.201)	(0.126)	(0.229)	(0.156)
Overweight	-0.003	0.009	-0.063	0.092	-0.522	0.089	0.115	0.148	0.077
	(0.049)	(0.043)	(0.055)	(0.139)	(0.741)	(0.191)	(0.125)	(0.112)	(0.133)
Underweight	0.016	0.057	0.008	-0.339*	-0.650	-0.150	0.173†	-0.153	0.102
	(0.032)	(0.046)	(0.033)	(0.133)	(1.283)	(0.131)	(0.091)	(0.097)	(0.097)
Tall	0.178†	0.166†	0.058	0.105	-0.003	0.068	-0.277	0.189	-0.188
	(0.104)	(0.101)	(0.083)	(0.078)	(0.371)	(0.084)	(0.219)	(0.308)	(0.180)
Short	0.017	0.030	0.033	-0.015	0.369	-0.014	0.010	0.153	0.006
	(0.015)	(0.025)	(0.021)	(0.065)	(0.518)	(0.061)	(0.058)	(0.100)	(0.061)
Dates	-0.027**	-0.035***	-0.034***	-0.035†	-0.089	0.003	-0.031	0.014	-0.058*
	(0.008)	(0.009)	(0.010)	(0.020)	(0.187)	(0.029)	(0.023)	(0.032)	(0.028)
Comfortable	-0.015	0.036	-0.024	0.031	-0.054	0.067	0.042	-0.021	0.038
	(0.022)	(0.029)	(0.025)	(0.065)	(0.199)	(0.068)	(0.077)	(0.107)	(0.086)
Wealthy	0.059*	0.114**	0.065†	0.226**	0.031	0.079	0.083	-0.003	0.254
	(0.029)	(0.041)	(0.034)	(0.081)	(0.251)	(0.094)	(0.094)	(0.139)	(0.167)
Sociability	0.020†	-0.001	0.017	0.069*	0.128	0.026	-0.088*	-0.148**	-0.079†
	(0.010)	(0.014)	(0.014)	(0.035)	(0.158)	(0.036)	(0.040)	(0.051)	(0.041)
Vigor	-0.001	-0.018	0.001	-0.025	-0.101	-0.042	-0.000	0.107†	-0.036

	(0.017)	(0.020)	(0.010)	(0.036)	(0.216)	(0.047)	(0.031)	(0.060)	(0.053)
Mature	0.020*	0.030*	0.014	-0.067*	0.144	0.030	0.073†	0.083	0.080†
	(0.010)	(0.014)	(0.012)	(0.034)	(0.349)	(0.054)	(0.041)	(0.071)	(0.046)
Self-Confidence	-0.000	-0.001	-0.001	0.009	0.102	-0.017	-0.032	0.025	-0.016
	(0.009)	(0.013)	(0.014)	(0.030)	(0.199)	(0.032)	(0.035)	(0.040)	(0.034)
Tidiness	-0.001	-0.009	0.001	0.029	-0.238	0.050	0.020	-0.091	-0.014
	(0.011)	(0.013)	(0.011)	(0.032)	(0.471)	(0.033)	(0.040)	(0.056)	(0.042)
Math Score	0.002***	0.003***	0.003***	0.001	-0.008	0.001	0.002	-0.002	0.000
	(0.000)	(0.001)	(0.001)	(0.002)	(0.020)	(0.002)	(0.002)	(0.002)	(0.002)
Vocabulary Score	0.150***	0.168*	0.129*	0.248	-0.214	0.055	0.189	-0.243	-0.030
	(0.038)	(0.078)	(0.050)	(0.155)	(0.544)	(0.247)	(0.155)	(0.224)	(0.158)
Social Studies Score	0.120***	0.136**	0.110*	0.353†	0.580	0.227	0.014	0.485**	0.108
	(0.036)	(0.050)	(0.046)	(0.181)	(1.038)	(0.215)	(0.180)	(0.186)	(0.161)
Science Score	-0.032	-0.105†	-0.009	-0.233	0.186	0.003	0.130	0.137	0.049
	(0.036)	(0.062)	(0.047)	(0.169)	(0.775)	(0.189)	(0.122)	(0.182)	(0.147)
Scientific Attitude Score	0.030	0.000	0.006	-0.017	0.235	-0.059	0.304**	0.159	0.245*
	(0.029)	(0.051)	(0.039)	(0.141)	(0.617)	(0.159)	(0.098)	(0.132)	(0.108)
Law Score	-0.010	-0.083	-0.031	0.386*	-0.371	0.230†	0.113	0.265*	0.237*
	(0.033)	(0.056)	(0.040)	(0.163)	(0.828)	(0.119)	(0.109)	(0.124)	(0.103)
Military Score	-0.005	0.038	0.030	0.021	0.085	0.026	0.015	0.015	0.080
	(0.032)	(0.045)	(0.035)	(0.116)	(0.496)	(0.121)	(0.104)	(0.136)	(0.119)
Business Score	0.008	0.018	-0.026	-0.324†	0.905	-0.129	0.050	-0.085	0.038
	(0.033)	(0.047)	(0.038)	(0.174)	(1.313)	(0.145)	(0.124)	(0.151)	(0.120)
Etiquette Score	0.006	-0.027	-0.014	-0.138	0.574	-0.127	0.089	0.229†	0.082
	(0.031)	(0.049)	(0.040)	(0.102)	(1.348)	(0.110)	(0.103)	(0.131)	(0.109)
English Score	0.018	0.082	0.046	0.108	0.081	0.022	-0.068	0.485	0.292
	(0.041)	(0.060)	(0.044)	(0.175)	(0.496)	(0.250)	(0.186)	(0.300)	(0.252)
Science Awards	0.017	0.020	0.030	0.046	0.318	0.036	0.022	0.065	0.047
	(0.017)	(0.025)	(0.020)	(0.031)	(0.368)	(0.031)	(0.028)	(0.047)	(0.044)
Fine Arts Awards	-0.009	-0.012	-0.010	-0.044*	-0.152	-0.036†	-0.018	-0.054	-0.060*
	(0.006)	(0.009)	(0.008)	(0.021)	(0.188)	(0.021)	(0.016)	(0.035)	(0.027)
Sports Awards	0.004	-0.004	-0.006	-0.044**	0.179	-0.004	0.008	-0.010	0.017
	(0.009)	(0.013)	(0.011)	(0.017)	(0.298)	(0.025)	(0.020)	(0.039)	(0.023)
Arithmetic Skills	-0.028	-0.087	-0.054†	-0.073	-0.805	-0.132	-0.126	-0.336	-0.069
	(0.028)	(0.059)	(0.033)	(0.133)	(0.816)	(0.127)	(0.104)	(0.217)	(0.139)
Reading Skills	0.014	-0.008	0.020	-0.024	0.635	0.100	0.171†	0.152	0.136
	(0.025)	(0.037)	(0.031)	(0.109)	(1.073)	(0.116)	(0.098)	(0.116)	(0.118)

Clerical Skills	0.004	0.024	0.013	-0.166	-0.241	0.012	0.070	-0.065	-0.124
	(0.026)	(0.037)	(0.032)	(0.122)	(0.620)	(0.103)	(0.106)	(0.128)	(0.120)
Identification Skills	0.013	0.029	-0.009	0.107	-0.233	-0.008	0.076	0.196	0.188
	(0.026)	(0.038)	(0.032)	(0.108)	(0.611)	(0.147)	(0.099)	(0.137)	(0.116)
Public Service Interest	0.029***	0.023†	0.041***	0.021	-0.016	0.035	0.090*	-0.038	-0.015
	(0.009)	(0.013)	(0.011)	(0.033)	(0.160)	(0.035)	(0.044)	(0.065)	(0.040)
Business Management Interest	0.002	-0.000	0.005	-0.018	-0.130	-0.063	-0.024	0.051	0.023
	(0.009)	(0.014)	(0.013)	(0.034)	(0.094)	(0.038)	(0.038)	(0.058)	(0.041)
School-fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
First stage (Kleibergen-Paap) F-Statistic	7.77	0.91	2.94	1.92	0.10	1.86	1.76	2.06	3.86
p-value (F-Statistic)	0.10	0.46	0.02	0.10	0.98	0.12	0.14	0.09	0.01
Cragg-Donald-Wald F-Statistic	7.44	3.51	7.01	3.76	0.16	3.98	3.95	2.32	6.15
Sargan-Hansen statistic	4.98	0.44	0.28	0.34	0.31	3.22	0.35	7.74	2.39
p-value (Sargan-Hansen statistic)	0.174	0.930	0.964	0.953	0.956	0.359	0.951	0.05	0.50
Observations	11,210	6,102	6,431	680	325	420	745	309	442

† $p<0.1$ * $p<0.05$, ** $p<.01$, *** $p<.001$

Note: Table 9F reports the second stage coefficients and standard deviations from IV estimations. The instrumented variables are whether individuals acted as captain and president, captain only, or president. Information whether each parent was a member of a club of a team represents the instrumental variables. The sample for the treatment analysis are white females, non-white males and non-white females respectively. For the last two columns of the non-white female sample, due to the large number of variables the Sargan-Hansen statistic cannot be directly estimated, as the covariance matrix of moment conditions is not of full rank. This table differs from Table 4.9 as it includes robust standard errors.

Table 10A: Extension – *Ordered-Logistic-Regression: Effects of High School Leadership Activities on 2011/2012 Household Income for White Males*

	Model 1 Household Income	Model 2 Household Income	Model 3 Household Income	Model 4 Household Income
Leader				
Both Captain and President	1.049**	1.020**	1.145***	0.961**
	(0.353)	(0.343)	(0.344)	(0.354)
Captain Only	0.027	-0.029	-0.096	-0.124
	(0.396)	(0.414)	(0.430)	(0.435)
President Only	0.252	0.097	0.095	0.035
	(0.328)	(0.344)	(0.331)	(0.349)
Member				
Both on Team and in Club	0.240	-0.137	-0.144	-0.149
	(0.457)	(0.501)	(0.572)	(0.573)
On Team only	-0.047	-0.450	-0.523	-0.718
	(0.694)	(0.648)	(0.669)	(0.701)
In Club Only	0.094	-0.177	-0.070	-0.130
	(0.479)	(0.528)	(0.599)	(0.594)
Controls				
Math Score		0.016***	0.015**	0.011*
		(0.005)	(0.005)	(0.005)
Parent´s Education				
High School			0.100	0.096
			(0.306)	(0.314)
College Degree			0.862*	0.717†
			(0.375)	(0.383)
Educational Attainment				
Some College				0.489
				(0.474)
College Degree or Higher				0.824*
				(0.406)
School-fixed Effects	No	No	No	No
Wald Chi^2	15.40	26.01	30.48	38.93
p > Chi^2	0.052	0.002	0.002	0.001
Pseudo R^2	0.022	0.041	0.051	0.059
Observations	251	251	251	251

† p<0.1 * p<0.05, ** p<.01, *** p<.001

Note: Table 10A contains the same variables and sample as Table 4.10 reported in the paper but employs ordered-logistic regressions instead of OLS regressions. As fixed effects frequently cause issues with ordered-logistic regressions Table 10A does not include school dummies. The first line for each variable corresponds to marginal effects with standard errors in parentheses. All models include (unreported) grade attainment control variables. Models 3 and 4 include an unreported dummy variable for lack of parent's education. Model 4 includes an unreported dummy variable for lack of educational attainment. The sample is based on the 2011-12 Pilot Study.

Table 10B: Extension – *OLS-Regression: Effects of High School Leadership Activities on 2011/2012 Household Income for White Females*

	Model 1 Household Income	**Model 2** Household Income	**Model 3** Household Income	**Model 4** Household Income
Leader				
Both Captain and President	0.042	0.013	0.014	-0.025
	(0.174)	(0.161)	(0.161)	(0.161)
Captain Only	-0.086	-0.098	-0.073	-0.098
	(0.181)	(0.161)	(0.160)	(0.156)
President Only	0.239	0.208	0.218	0.169
	(0.173)	(0.169)	(0.166)	(0.163)
Member				
Both on Team and in Club	0.797**	0.580*	0.429	0.348
	(0.241)	(0.267)	(0.277)	(0.290)
On Team only	-0.102	0.173	-0.399	-0.436
	(0.356)	(0.366)	(0.430)	(0.432)
In Club Only	0.631**	0.445†	0.292	0.200
	(0.226)	(0.257)	(0.268)	(0.281)
Controls				
Math Score		0.011***	0.011***	0.009***
		(0.002)	(0.002)	(0.002)
Parent´s Education				
High School			0.227	0.233†
			(0.144)	(0.139)
College Degree			0.327	0.275
			(0.199)	(0.192)
Educational Attainment				
Some College				0.384†
				(0.212)
College Degree or Higher				0.280
				(0.196)
School-fixed Effects	Yes	Yes	Yes	Yes
F	3.92	6.26	5.09	4.67
p > F	0.000	0.000	0.000	0.000
Adjusted R^2	0.044	0.148	0.171	0.195
Observations	302	302	302	302

† p<0.1 * p<0.05, ** p<.01, *** p<.001

Note: Table 10B differs from Table 4.10 reported in the paper by using only white females as the underlying sample. All models include (unreported) grade and school attainment control variables. Models 3 and 4 include an unreported dummy variable for lack of parent’s education. Model 4 includes an unreported dummy variable for lack of educational attainment. The sample is based on the 2011-12 Pilot Study.

Table 10C: Extension – *Ordered-Logistic-Regression: Effects of High School Leadership Activities on 2011/2012 Household Income for White Females*

	Model 1 Household Income	**Model 2** Household Income	**Model 3** Household Income	**Model 4** Household Income
Leader				
Both Captain and President	0.115	0.084	0.089	0.044
	(0.308)	(0.316)	(0.336)	(0.342)
Captain Only	-0.151	-0.169	-0.167	-0.214
	(0.348)	(0.347)	(0.361)	(0.354)
President Only	0.390	0.241	0.253	0.200
	(0.314)	(0.318)	(0.333)	(0.335)
Member				
Both on Team and in Club	1.933***	1.691***	1.362***	1.276**
	(0.259)	(0.335)	(0.387)	(0.397)
On Team only	0.015	0.543	-0.823	-0.574
	(0.405)	(0.478)	(0.657)	(0.649)
In Club Only	1.578***	1.314***	0.958*	0.844*
	(0.300)	(0.374)	(0.421)	(0.429)
Controls				
Math Score		0.018***	0.019***	0.019***
		(0.004)	(0.004)	(0.004)
Parent´s Education				
High School			0.522†	0.550†
			(0.309)	(0.307)
College Degree			0.817*	0.821*
			(0.400)	(0.412)
Educational Attainment				
Some College				0.620
				(0.410)
College Degree or Higher				0.119
				(0.404)
School-fixed Effects	No	No	No	No
Wald Chi^2	11.18	35.67	48.38	56.06
$p > Chi^2$	0.192	0.000	0.000	0.000
Pseudo R^2	0.019	0.047	0.065	0.077
Observations	302	302	302	302

† p<0.1 * p<0.05, ** p<.01, *** p<.001

Note: Table 10C differs from Table 4.10 reported in the paper by using only white females as the underlying sample. This table differs from Table 10B as coefficients derive from ordered-logistic regressions with standard errors in parentheses. As fixed effects frequently cause issues with ordered-logistic regressions Table 10C does not include school dummies. All models include (unreported) grade attainment control variables. Models 3 and 4 include an unreported dummy variable for lack of parent's education. Model 4 includes an unreported dummy variable for lack of educational attainment. The sample is based on the 2011-12 Pilot Study.

Table 10D: Extension – *OLS-Regression: Effects of High School Leadership Activities on 2011/2012 Household Income for White Males including Living Alone Dummy*

	Model 1 Household Income	**Model 2** Household Income	**Model 3** Household Income	**Model 4** Household Income
Leader				
Both Captain and President	0.436*	0.452*	0.460*	0.358†
	(0.191)	(0.187)	(0.188)	(0.199)
Captain Only	-0.173	-0.138	-0.162	-0.181
	(0.240)	(0.243)	(0.249)	(0.254)
President Only	0.031	-0.026	-0.041	-0.091
	(0.168)	(0.171)	(0.170)	(0.175)
Member				
Both on Team and in Club	0.219	0.006	0.030	0.000
	(0.249)	(0.279)	(0.289)	(0.293)
On Team only	0.178	0.011	-0.007	-0.160
	(0.308)	(0.330)	(0.306)	(0.310)
In Club Only	0.149	-0.012	0.003	-0.051
	(0.278)	(0.298)	(0.311)	(0.312)
Controls				
Math Score		0.006*	0.006*	0.004
		(0.003)	(0.003)	(0.003)
Parent´s Education				
High School			0.045	0.077
			(0.186)	(0.184)
College Degree			0.413†	0.382†
			(0.234)	(0.230)
Educational Attainment				
Some College				0.202
				(0.245)
College Degree or Higher				0.450†
				(0.239)
Demographics				
Living Alone	-0.697***	-0.601**	-0.596***	-0.532**
	(0.184)	(0.184)	(0.176)	(0.178)
School-fixed Effects	Yes	Yes	Yes	Yes
F	3.18	3.12	2.69	2.50
p > F	0.001	0.001	0.002	0.002
Adjusted R^2	0.137	0.156	0.162	0.166
Observations	249	249	249	249

† p<0.1 * p<0.05, ** p<.01, *** p<.001

Note: Table 10D extends the results from Table 4.10 reported in the paper by including the additional variable *Living Alone* which captures whether respondents lived alone at the time of the 2011-2012 Pilot Study interview. All models include (unreported) grade and school attainment control variables. Models 3 and 4 include an unreported dummy variable for lack of parent's education. Model 4 includes an unreported dummy variable for lack of educational attainment. The sample is based on white males from the 2011-12 Pilot Study.

Table 10E: Extension – *OLS-Regression: Effects of High School Leadership Activities on 2011/2012 Household Income for White Females including Living Alone Dummy*

	Model 1 Household Income	**Model 2** Household Income	**Model 3** Household Income	**Model 4** Household Income
Leader				
Both Captain and President	0.104	0.075	0.072	0.039
	(0.164)	(0.150)	(0.150)	(0.149)
Captain Only	-0.022	-0.035	-0.024	-0.061
	(0.168)	(0.149)	(0.150)	(0.147)
President Only	0.230	0.200	0.200	0.154
	(0.166)	(0.162)	(0.160)	(0.158)
Member				
Both on Team and in Club	0.805***	0.585*	0.458*	0.391
	(0.239)	(0.228)	(0.225)	(0.245)
On Team only	0.377	0.669*	0.219	0.181
	(0.347)	(0.329)	(0.394)	(0.391)
In Club Only	0.705**	0.514*	0.384†	0.300
	(0.225)	(0.215)	(0.208)	(0.230)
Controls				
Math Score		0.011***	0.011***	0.009***
		(0.002)	(0.002)	(0.002)
Parent´s Education				
High School			0.207	0.205
			(0.132)	(0.128)
College Degree			0.324†	0.249
			(0.184)	(0.178)
Educational Attainment				
Some College				0.270
				(0.203)
College Degree or Higher				0.382*
				(0.178)
Demographics				
Living Alone	-0.703***	-0.715***	-0.689***	-0.698***
	(0.109)	(0.103)	(0.102)	(0.104)
School-fixed Effects	Yes	Yes	Yes	Yes
F	7.37	10.12	8.56	7.93
p > F	0.000	0.000	0.000	0.000
Adjusted R^2	0.171	0.280	0.293	0.318
Observations	301	301	301	301

† p<0.1 * p<0.05, ** p<.01, *** p<.001

Note: Table 10E differs from Table 4.10 reported in the paper by using only white females as the underlying sample and the additional variable *Living Alone* as an additional explanatory variable which captures whether respondents lived alone at the time of the 2011-2012 Pilot Study interview. All models include (unreported) grade and school attainment control variables. Models 3 and 4 include an unreported dummy variable for lack of parent's education. Model 4 includes an unreported dummy variable for lack of educational attainment. The sample is based on white females from the 2011-12 Pilot Study.

Table 10F: Extension – *OLS-Regression: Effects of High School Leadership Activities on 2011/2012 Household Income for White Males (robust standard errors)*

	Model 1 Household Income	**Model 2** Household Income	**Model 3** Household Income	**Model 4** Household Income
Leader				
Both Captain and President	0.478*	0.487*	0.498*	0.371†
	(0.201)	(0.195)	(0.195)	(0.203)
Captain Only	-0.171	-0.139	-0.170	-0.192
	(0.255)	(0.256)	(0.266)	(0.270)
President Only	0.016	-0.053	-0.062	-0.114
	(0.176)	(0.177)	(0.178)	(0.181)
Member				
Both on Team and in Club	0.155	-0.098	-0.070	-0.095
	(0.260)	(0.285)	(0.287)	(0.286)
On Team only	0.033	-0.150	-0.153	-0.325
	(0.324)	(0.333)	(0.306)	(0.302)
In Club Only	0.067	-0.118	-0.101	-0.151
	(0.278)	(0.297)	(0.304)	(0.300)
Controls				
Math Score		0.008**	0.008**	0.005*
		(0.003)	(0.003)	(0.003)
Parent´s Education				
High School			0.025	0.070
			(0.187)	(0.185)
College Degree			0.350	0.313
			(0.237)	(0.234)
Educational Attainment				
Some College				0.172
				(0.251)
College Degree or Higher				0.514*
				(0.250)
School-fixed Effects	Yes	Yes	Yes	Yes
F	1.37	2.10	1.67	1.71
p > F	0.211	0.031	0.078	0.053
Adjusted R^2	0.083	0.117	0.119	0.132
Observations	251	251	251	251

† p<0.1 * p<0.05, ** p<.01, *** p<.001

Note: Table 10F differs from Table 4.10 reported in the paper by including robust standard errors. All models include (unreported) grade attainment control variables. Models 3 and 4 include an unreported dummy variable for lack of parent's education. Model 4 includes an unreported dummy variable for lack of educational attainment. The sample is based on the 2011-12 Pilot Study.

5 Essay 4 – Open data practices in innovation management research

5.1 Introduction

In 2015, Carlos Moedas (2015), the European Commissioner for Research, Science and Innovation at this time, introduced the three Os – open innovation, open science, and open to the world – as goals for research and innovation policies in the EU.[28] Open innovation focuses on firms collaborating with their environment to gather and supply new approaches and technologies (Chesbrough, 2003). Open science concentrates on improving the input and impact to and from research (Nosek et al., 2015). Open to the world highlights the need for partnerships across country or disciplinary borders (Moedas, 2015). In this article, our focus is on open science, a concept often connoted with transparency, accessibility, collaboration, and above all, sharing: "Open science is transparent and accessible knowledge that is shared and developed through collaborative networks" (Vicente-Saez, & Martinez-Fuentes, 2018: 428). Corresponding practices aim to expand the ethos of scientific inquiries by publicly, freely revealing results and their underlying empirical data (Vicente-Saez et al., 2020).

Regarding the sharing of results, ever since ancient Greece, scientists have worked to share their generated insights with colleagues, peers, and a broader audience. In contrast to the scale and scope of the sharing of results, the data that underlie academic research are shared to a much smaller extent. In fact, many scientists prefer to keep their data private (Tenopir et al., 2011), despite the advantages data sharing could offer for subsequent research and society (Molloy, 2011). If open science is the future of scientific research, then it is important to understand the drivers and obstacles of sharing research data publicly and to explain why we see so little open data policies in practice.

Being management scholars (and writing for management scholars), we analyse data sharing by studying management researchers' perceptions about the perceived trade-off between the costs and benefits of data sharing and how amenable they would be to institutional arrangements that balance the corresponding trade-offs to provide more incentives for data sharing. Within management, we picked a research field that should – by definition – be more

[28] I wish to thank Gloria Barczak, Jermain Kaminski, Christian Hopp and Frank Piller for the collaboration, their valuable comments and their inputs on this paper.

open: *innovation research*, where researchers have strongly advocated the importance of openness for all types of innovation outcomes (e.g. Bogers et al., 2017; von Hippel, 2017). Numerous studies in the innovation literature argue for the benefits of knowledge sharing and collaboration for more innovative outcomes (Randhawa et al., 2016; West et al., 2014). However, only recently have researchers in the innovation community started to make their research data accessible to a broader audience. Examples of open data are, for instance, Reynolds (2007), sharing panel data in Entrepreneurial Dynamics (PSED); Marx, & Fuegi, (2020a, 2020b), providing a dataset with citations from worldwide patents to scientific articles, or authors like Sorenson et al. (2016), and Yu et al. (2017) with data on household innovation or crowdfunding research, who share their data through the repository DataVerse (King, 2007). As an important case in point, DataVerse features more than 98,000 open analysable datasets (as of July 2020), of which 41,813 are attributed to the social sciences in general, but only 582 to Business and Management.

The last numbers indicate that the previously cited innovation scholars sharing their research data are a rare exception. This led us to our **research question**: *When and why do innovation management scholars engage in open data sharing and what factors increase or reduce the likelihood of data sharing?* Using the Resource Based View (RBV) of the firm, we argue that researchers' proprietary data are resources that can provide a competitive advantage to individuals or research teams as these data are valuable, rare, imperfectly imitable, and not substitutable (Barney, 2001; Wade, & Hulland, 2004). As a result, we investigate through an exploratory study, individual and institutional factors that can be perceived as barriers or enablers to the likelihood of one's publicly sharing their data. By doing so, this research makes two important contributions. First, it extends the RBV to the emerging field of open science with a focus on open data. In viewing data as a critical resource for researchers, we examine the trade-offs between the individual level barriers and the public benefit of open data sharing among (innovation) management scholars. Second, by investigating open data within the realm of innovation management scholars, the findings lead to suggestions that individuals and institutions can implement to encourage more data sharing and help to initiate a discussion on open (research) data in our own discipline.

In the following, we outline existing research on data sharing and open data and build a set of hypotheses. We then explore the perceptions and practices of data sharing in the management sciences with a survey among 173 innovation researchers. Subsequently, we present the results of our statistical analyses and discuss our findings. The article concludes by presenting a set of advice for journal editors, maintainers of bibliographic databases, and academic societies. We also indicate potential improvements for research policy and highlight future research opportunities.

5.2 Context: Evidence and practices of sharing research data in different disciplines

Researchers across different disciplines are restrictive in providing access to their data. As a case in point, Wicherts et al. (2006) contacted corresponding authors of 141 published articles in four major journals of the American Psychological Association (all of which require authors to share their data upon request), attempting to replicate their studies. After six months (and more than 400 e-mail correspondences), they received data from only 38 papers (~27%). These results are not singular but confirmed by other studies. Krawczyk and Reuben (2012) received data from 44% of the corresponding authors of two hundred economics articles. Only 25% of surveyed pharmaceutical researchers shared their data upon request (Kirwan, 1997), and only one out of 29 corresponding authors from articles in the British Medical Journal provided data upon request (Reidpath, & Allotey, 2001). Savage and Vickers (2009) received one data set out of ten for articles published in PLoS Medicine and PLoS Clinical Trials, journals with requirements to share data upon request. Evidently, many scholars from different scientific disciplines do not engage in open data practices, despite calls for data sharing and policies of funding organisations (in Europe at least) implying that scientists should make the empirical data behind their research publicly accessible (Muscio et al., 2013).

To corroborate evidence from these article-based studies, Tenopir et al. (2011) found that one third of respondents in a survey of 1,329 scientists skipped questions entirely regarding the amount of data shared, whereas 46% of those answering the questions stated that they have never shared any data and only less than 6% indicated that they shared all their data. Paradoxically, Fecher et al. (2017) found that 76% (out of 1,564 respondents) agreed that other researchers should share their data. Yet, only 35% of these very same respondents acknowledged that data sharing is common in their research field and only 13% had shared data publicly at least once. Social science researchers were the most averse to share data (Fecher et al., 2017). Alas, data sharing in management research brings us into surprisingly uncharted territory. There is sparse theoretical work that highlights factors that may explain data sharing in the management sciences, and there is even less empirical work on the antecedents to data sharing practices in our own discipline of innovation management (Kim, & Stanton, 2016; Kim, & Adler, 2015;

Kim, & Stanton, 2012). As Friesike et al. (2015: 581) conclude, “while academic studies on open innovation are burgeoning, most research on the topic focuses on the later phases of the innovation process. So far, the impact and implications of the general tendency towards more openness in academic and industrial science at the very front-end of the innovation process have been mostly neglected”.

In the following, we pick up this torch and develop a theoretical framework to better understand the potential benefits and caveats that come with open data. Based on this framework, we develop a set of hypotheses that link individual incentives, costs, and trade-offs with the institutional environment in which academic data collection efforts, research, and publishing, are embedded.

5.3 Theory and hypotheses development

In a famous quote, Thomas Jefferson described in 1813 a core tension of sharing information: “If nature has made any one thing less susceptible than all others of exclusive property, it is the action of the thinking power called an idea, which an individual may exclusively possess as long as he keeps it to himself; but the moment it is divulged, it forces itself into the possession of every one, and the receiver cannot dispossess himself of it” (Jefferson, 1813).

Researchers deliberating whether or not to reveal their research data are at the very junction that Thomas Jefferson so vividly described. Going back to Merton (1969), there are strong arguments that scientists should have a right to priority for an eventually made discovery so to provide them incentives to engage in the process of risky and uncertain discovery in the first place. In economics and management research, the “property” (in the words of Jefferson) that drives research is often data: “The goal of empirical economics is to learn from data” (Heckman, & Singer (2017: 299). Data, and technology to analyse this data, represent the key resources on which empirical researchers draw to advance science – and their individual careers. Researchers do not only decide about research questions, but also on research strategies. Evidently, different methods can answer the same question and consequently the same data can be employed to answer multiple research questions (Partha, & David, 1994).

High quality (proprietary) datasets allow scientists to address various research questions and thus publish a number of articles without facing the need to collect data for every single individual publication (Kirkman, & Chen, 2011). With publications being the “gold standard” (Altbach, 2015: 6) of academic productivity, non-public datasets give their owner a significant competitive advantage over their fellow researchers with whom they compete for publication spots, job and tenure positions as well as grant funding (Kwiek, 2015). This is especially important as prior studies have identified path and state-dependencies in research productivity with accumulative patterns of discovery and research status (Merton, 1969).

5.3.1 A resource-based view on the resistance to open data

To start with, we need to understand the extent to which researchers are amenable to collaborate and share their knowledge (and data). Empirical scientists are ambivalent between intermediate (data sharing) vs. final disclosure of their research (publications of all associated studies based on underlying data). Intermediate disclosure is the more piecemeal provision of knowledge, methods, data, or progress, whereas final disclosure involves a standardized (for example through peer reviewed publications) provision of uncovered final solutions (Boudreau, & Lakhani, 2015).

Hence, there is evidently a trade-off between providing incentives for researchers to make costly investments into the research effort, while at the same time encouraging follow-up knowledge reuse (through disclosure, among others). Problematically, mechanisms that enforce and govern the reuse of knowledge (and ideas) are hardly contractible and rewards for intermediate disclosure (before a researcher has concluded his research agenda with a given dataset) reduces incentives and limits the pioneer´s ability to control the reuse of his data collected (Boudreau, & Lakhani, 2015). A potential sharing of data may therefore be at odds with the academics' self-interest, the protection of the main resource for their scientific lead, leading to ambiguity as to when and how the norm should be followed (Defazio et al., 2020).

As a case in point, Fecher et al. (2015, 4) discussed the "fear of competitive misuse" that prevents individuals from sharing data. The discovery of the helix structure of the DNA represents one famous example of this. Watson and Cricks (1953) had unsuccessfully worked with different theoretical models suggesting various structures of DNA, without ever touching or looking at fibres of DNA. Yet, on January 30, 1953, Watson paid a visit to Cambridge's King's College, where Franklin had carried out an X-ray crystallography that showed the helical structure of DNA (Tobin, 2003). Eventually, Watson got access to Franklin´s X-ray photographs (through Wilkins) without Franklin's permission. The access to the data laid the groundwork for the model developed by Watson and Crick (1953), for which Watson, Crick and Wilkins, but not Franklin, were awarded the Nobel Prize in 1962.

Consequently, we can very well think of research data as one of scholars' key resources that enables them to outperform their competitors (Barney, 2001), a premise of the resource-based view (RBV), which posits that firms should control valuable, rare, and inimitable resources to achieve a competitive advantage. Proprietary research data are exactly a resource of such character. Sharing gathered data publicly makes it accessible to colleagues who did not have to struggle their way through designing questionnaires, managing surveys, administrating ethical committees, and the like. As researchers compete for a limited number of publication spots (and tenured positions), "free riders" could submit (and hence publish) more papers in shorter periods than the scholars who also engage in data collection (van Raan, 2001; Longo, & Drazen, 2016). As a consequence, academics are “incentivised to withhold information as they are in a winner takes-all publishing competition” (Defazio et al., 2020: 5).

5.3.1.1 Perceived opportunity costs of data sharing

The incentives to keep research data proprietary echoes a prominent debate in the New England Journal of Medicine, where Longo and Drazen (2016) feared that having access to original authors' data might open the door to "research parasites", i.e. researchers gaining unearned benefits from the data collection efforts of others, which may, at worst, undermine the original publication. As the case of Watson and Cricks so vividly illustrates, this may at times even involve researchers that publish without approval from the original authors for using their data (Wilbanks, & Friend, 2016). The most extreme form is data thievery which occurs when scholars share their data while still writing their paper (da Silva, & Dobranszki, 2015). Other researchers take the data and publish an article with the same research question before the scholars who conducted the data collection can publish their article. In turn, the original authors lose out on the publication opportunity, but are still stuck with the direct costs of the data collection (da Silva, & Dobranzski, 2015).

Clearly, researchers perceive this as unfair and therefore tend to keep gathered data as their “trade secrets” (Pfenninger et al., 2017, 212), objecting the idea of data sharing, as the Stanford economist David already proposed in an early essay about the open science paradigm: “Priority creates a privately-owned asset from the very act of relinquishing exclusive possession of the new knowledge.” (David, 1995: 19). As such, data is a key resource for scientists and in line with the resource based view this provides strong incentives for academics to privatize their knowledge. Consequently, researchers might simply not share their data because they may be able to publish subsequent articles out of their initial data collection effort. This creates a situation where individual researchers who collected the data regards the data as proprietary because they bore the time and efforts to collect the data in the first place and do not want to give up on the opportunity to publish multiple papers based on the same dataset. We therefore formulate the following hypothesis:

> **Hypothesis 1a:** *The stronger researchers perceive that possessing non-public datasets provide them with an advantage over other researchers, the lower the likelihood of data sharing.*

5.3.1.2 Perceived direct costs of data sharing

Our previous argument can be seen as opportunity costs related to foregone potential publications. In addition to these opportunity costs associated with data sharing, there exist also direct costs associated with open data. As a case in point, existing research discussed that firms incurring high costs for resource sharing engage in strategic networks and other collaborative approaches less frequently (Das, & Teng, 2000; Silverman, 1999).

Engaging in the process of data collection, curation, and preparation activities relies on scarce resources (Defazio et al., 2020), that can be influenced by research, teaching and commercialization activities, or the individual family situation. Especially data curation and preparation involve great efforts and bind critical time and resources that are generally not awarded (though often asked for) by funding agencies. In fact, Kim and Stanton (2016) already pointed towards a negative association between the incurred costs and subsequent data sharing intentions. Hence, we expect a similar relationship between the direct costs associated with data collection, curation and publication and actual data sharing behavior. This leads us to the following hypothesis:

> **Hypothesis 1b:** *The larger researchers perceive the direct costs of data sharing, the lower the likelihood of data sharing.*

5.3.1.3 Perceived reputational costs of data sharing

At the same, data sharing could also harm scholars' reputations. If researchers share their data, they face the danger of revealing flaws in their research and may get unwanted public exposure. While 'disagreement likely facilitates the development of new ideas, contributing to creativity and innovation' (Wang, & Noe, 2010: 124), failed replication attempts using the shared data might defile scholars' reputation (Barry, & Bannister, 2014). Doctoral student Thomas Herndon found a spreadsheet mistake in Reinhart and Rogoff´s (2010) work on austerity (Herndon et al., 2014). The subsequent quagmire brought publicity and unwanted media scrutiny for all involved researchers (Cassidy, 2013).

Unintended typos or similar minor errors in a data set could render results insignificant and turn implications upside-down. This might even result in corrections or even retractions (Molckovsky et al., 2011). At worst, it might ruin researchers' careers because "[t]here is a fundamental misconception that retractions are 'bad' without pausing to ask why the retraction took place" (Barbour et al., 2017: 1964). As a case in point, Lu et al. (2013) showed that researchers receive 6.9% less citations for publications published after a retraction of one of their prior publications. These considerations may prevent scholars to share their data, perceiving that their voluntary activity creates risks and challenges that can be easily prevented by just not making the effort. This leads to the following hypothesis on reputational costs impeding data sharing:

> **Hypothesis 1c:** *The larger researchers perceive the reputational costs of data sharing related, the lower the likelihood of data sharing.*

5.3.2 Perceived (im-)balance of costs and rewards of open data

Existing research on the resource-based view highlights that individuals and organizations will only engage in resource sharing if their overall perceived benefits outweigh their overall perceived costs (Bogers, 2011a). Following this argument, researchers might view the opportunity cost, direct cost, and reputational costs associated with data sharing as misplaced energy, especially when sufficiently tailored individual incentives for them to encourage data sharing are not in place. As such, in the absence of additional incentives, it stands to reason that an unbalanced trade-off between incentives and costs lowers the likelihood of data sharing by individual researchers. Of course, one could argue that receiving citations for shared datasets might represent such additional incentives. Yet even if the datasets are shared, Wallis, Rolando, and Borgman (2013) showed that such publicly available datasets (that took many weeks to prepare) are requested infrequently and rarely reused. In management research, the dominant scientific approach is theory and literature-guided confirmatory research, requiring authors to use datasets specifically tailored to assess pre-determined research hypotheses (Tukey, 1980). This methodological paradigm might explain the low reuse of data as publications based on primary data possess citation advantages over articles based on secondary data (Piwowar, & Vision, 2013). As a consequence, "the effort to make data discoverable is difficult to justify, given the infrequency with which investigators are asked to release their data" (Wallis et al., 2013: 14).

If scholars not only perceive high costs for procuring external data but also perceive that they are rewarded disproportionately for their provided open data, they might consider to not share their data and thus save on time, risk and efforts. In summary, we formulate the following hypothesis.

Hypothesis 2: *The stronger researchers perceive that the costs associated with open data outweigh the rewards, the lower their likelihood of data sharing.*

5.3.3 An open innovation perspective for the support of open data

Our previous arguments addressed factors that reduce resource sharing. However, prior literature on open innovation in the context of the resource-based view also outlines various cases where resource sharing in the presence of proprietary resources actually works (e.g. Das, & Teng, 2000; Mowery et al., 1998).

5.3.3.1 Community benefits of data sharing

Arya and Lin (2007) showed that organizations sharing their own resources might gain reputation through this process. This could also apply to academia, as Fecher et al. (2017) conclude. For example, Piwowar et al. (2007) investigated 85 publications in the field of clinical cancer treatment and found that papers with shared data received 69% more citations than those without shared data. In management research, work by La Porta et al. (2002) or Bloom and van Reenen (2007) provides anecdotal evidence, as both author collectives made the underlying data accessible to a broader audience. These publications collected more than 10,000 and 2,000 citations, respectively, and are considered ground-breaking in their fields.

In addition to the expectation of positive individual reputational effects through the increase in transparency and reproducibility, resource sharing can also induce several other benefits (Arya, & Lynn, 2007). As a case in point, Vanhaverbeke and Cloodt (2014: 265) argue that in today's rapidly changing economy, "firms—even the largest one—cannot develop all required resources internally and have to team up with innovation partners enabling resource flows between firms." Similarly, in academic research, advances in information technology and the nearly worldwide dissemination of English as the language of sciences have fiercely increased competition for publication spots (Di Bitetti, & Ferreras, 2017). To withstand this pressure, more and more scientists (including the author collective of this article) have teamed-up to collectively conduct and publish research (Lee, & Bozeman, 2005). Task division and specialization have enabled those teams to produce more creative and more profound articles at higher rates (Ductor, 2015; Manton, & English, 2007). In this context, publishing research data allows other interested scientists (e.g. those working on similar topics) to receive a better picture of what research is currently underway (Ross, & Krumholz, 2013). This might lead to researchers from different areas with different skills and knowledge offering to join the research team, thus further amplifying its investigations and profoundness (Edwards et al., 2011). As a working example, Walport and Brest (2011) argue that open data strongly enhances collaborations between scientists.

Open data also reduces the need for individual data collection. Open innovation is centred on the idea of purposively managing in and outflows of knowledge to speed up both, external as well as management of innovation (Tucci et al., 2016; Vicente-Saez, Gustafsson, & Van den Brande, 2020). Similar to open innovation inducing "that companies need not to reinvent the wheel, since they can rely on external sources" (Chesborough, 2003: 49), open data induces that researchers can skip the often costly and resource intensive steps of data collection (Uhlir, & Schröder, 2007). Hence, scholars can publish related studies quicker and therefore spend more time on other discoveries (Fischer, & Zigmond, 2010). The accelerated research process is not only beneficial for scholars, as they can publish more papers in a shorter period, but also for society as problem solving and technology development take less time (Kaye et al., 2009). We propose:

> **Hypothesis 3a***: The more affirmative scientists are to the community-wide benefits of data sharing, the higher the likelihood of data sharing.*

5.3.3.2 Institutional policies enforcing data sharing

Following prior research, resource sharing collaborations can emerge if all participants agree on an appropriate governance structure that facilitates and oversees the resource exchanges by aiming at avoiding one-sided exploitations and opportunism (Madhok, & Tallman, 1998; Boudreau, & Lakhani, 2015). A similar approach could be applied to academia as scientists are embedded into the wider academic community and as such are amenable to institutional logics (Defazio et al., 2020), meaning that their individual behavior is also affected by the behavior of others around them and the expectations the community holds about them. Institutional norms are crucial ways in which a scientific community affects peers and individual academics. These norms may equally attest to appropriate behavior or sanction inappropriate behavior, thus, providing guidance (Azoulay et al., 2015).

As a case in point, Nature's author policy encourages the citation of datasets, thus increasing the citation counts of authors who share their datasets. In fact, more and more journal policies require authors to make their data available to the public (Federer et al., 2018; Andreoli-Versbach, & Mueller-Langer, 2014; Borgman, 2012), and there is a rather high prevalence of authors' statements indicating that data would be available upon request (Piwowar, & Chapman, 2008; Whitlock, 2011). Still, many journals do not have a data policy at all (Stodden et al., 2018). Stodden, Guo and Ma (2013) report that among 170 multidisciplinary journals, 62% did not have a data policy. Nevertheless, some scholars reported that they would like to see the introduction of more data policies and data citation opportunities (Savage, & Vickers, 2009). As the efforts to comply with such new policies are lower for scholars who already share their datasets, as compared to those who do not engage in open data, we propose that scholars, who are in favour of such new policies, are those who already share more data:

> **Hypothesis 3b:** *The more affirmative scientists are to institutional pressure to increase data sharing, the higher the likelihood of data sharing.*

5.3.3.3 Proliferation of replication studies to enforce data sharing

Beccera, Lunnan and Huemer (2008) highlight that firms tend to share their knowledge and resources with other firms only if they perceive them to be trustworthy. This way they can ensure that they and their partners can benefit from the fruits of cooperation, but at the same time they can reduce the risk that their valuable resources are diluted. Firms' perceived trustworthiness's depends on various factors, including past business relations (Möllering, 2002), social ties (Wang et al., 2006), and cultural backgrounds (Özer et al., 2014).

Similarly, scholars' trustworthiness depends on various factors, too (Dilger et al., 2015). Recently, various disciplines have been plagued by research scandals and questions about the reproducibility of research findings (e.g. Hopp, & Hoover, 2017; 2019). Consequently, various management journals have begun to institutionalize replication studies that rely on the presence of open data (Bettis et al., 2016; Clapp-Smith et al., 2017). In fact, data sharing is the only way to enable direct replications, because they require the availability of the primary dataset (Schmidt, 2009). Hence, data sharing increases the credibility of existing research (Freese, 2007; Gerwin, 2016).

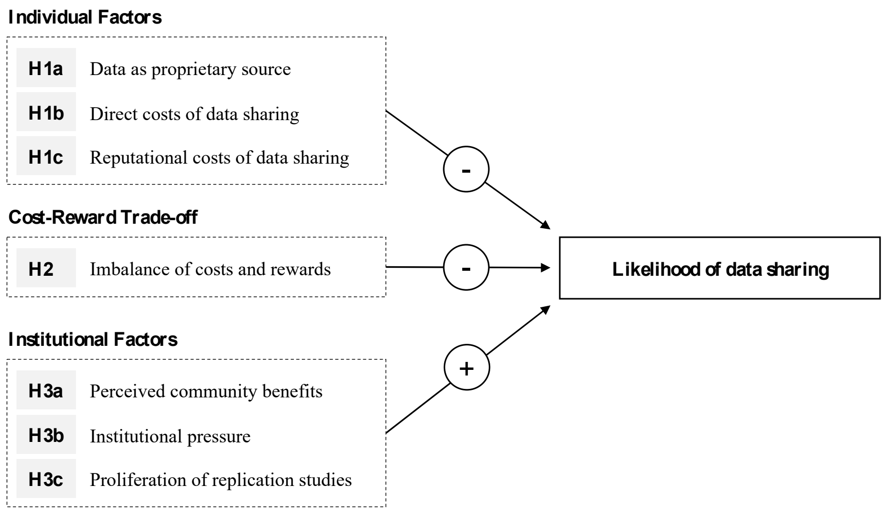
As a prime example of how these principles can be applied, researchers at the CERN and the Tevatron at Fermi Labs both undertook a simultaneous search for the Higgs Boson in 2012. Both were involved in replications of each other's findings using the very same data. Research that has an impact needs to be replicated to avoid misconceptions (Aaltonen, et al., 2012a; Aaltonen, et al. 2012b). Initial data from CERN was subsequently repeated and corroborated by other institutions and the consensus among involved scientists was, that only if signals obtained from each institution were statistically significant, evidence in favour of the missing piece in the particle physics standard model would be announced (Reich, 2012). This eventually led to the Nobel Prize for Francois Englert and Peter Higgs in 2013 and triggered further research as the data sharing enabled open experimentation (Adam-Bourdarios et al., 2015).

In management research, data access and exact replications might have worked to uncover the fraudulent asterisks that appeared for insignificant coefficients in Lichtenthaler and Ernst (2012), and it may have clarified the methodological irregularities in Avolio, Rotundo, and Walumbwa (2009). Data sharing, at the very least with reviewers, gives journals the chance to detect erroneous data or analyses prior to publishing (Morey, et al. 2016). Using data in peer review can act as the "the first line of defense" (Honig et al., 2014: 16) against potential academic misconduct. Hence, if scholars believe that replications increase the trustworthiness of existing research and ensure the continued credibility and sustainability of their academic field, they should also be more inclined to share data more openly. This leads us to formulate the following hypothesis:

> ***Hypothesis 3c:*** *The more affirmative scientists are to the proliferation of replication studies, the higher the likelihood of data sharing.*

Figure 1 graphically depicts our conceptual model and the corresponding hypotheses. Hypotheses 1a through 1c and 2 capture individual factors that reduce the likelihood of data sharing. Hypotheses 3a through 3c capture institutional factors that increase the likelihood of data sharing. We test this model empirically in the following chapters.

Figure 5.1: Conceptual model of Essay 4



5.4 Data and methodology

5.4.1 Survey distribution

To explore the factors driving and impeding data sharing among (innovation) management scholars, we conducted two empirical surveys using Qualtrics. To identify respondents, we used the participant lists of the World Open Innovation Conferences (WOIC) from 2013 to 2017 and the Druid conferences from 2011 to 2018. On the one hand, WOIC states at its website that at their conferences "world leading organizations share their open innovation challenges with expert academic colleagues to foster stronger connections between business practice and academia." The Druid website, on the other hand, refers to its conference series as "one of the world's premier academic conferences on innovation." Hence, we selected those two conferences to capture insights on innovation scholars with academic, but also with practitioners' backgrounds. As it follows from the contact data base deriving from those two conferences, participation in one of the two conference series represents a prerequisite of having a chance to participate in the survey. This might induce selection biases. However, both conferences attract not only eminent scholars of innovation management, but also give junior scholars chances for participation and presentation. It is therefore not surprising, that the share of PhD students that replied to the survey (23.12%) nearly mirrors the share of PhD students (26.07% on April 18, 2020) in the Technology and Innovation Management (TIM) division of the Academy of Management.

In 2018, we sent e-mails containing the survey link and a short description of the aim and scope of the research project to the 736 participants of the WOIC conferences between 2013 and 2017. After updating our contact directory from information contained in bounce-back emails that indicated changes of email addresses, we sent out a reminder email two weeks later. Overall, the first survey distribution wave generated 141 replies.

As the data received from this pre-test seemed plausible, we sent the same questionnaire also to the 2,429 participants of the Druid conferences between 2011 and 2018. We only included email addresses that were not already included in the WOIC contact directory. In addition, we prohibited ballot-boxing by allowing each IP-address to only take the survey once. We updated the email addresses based on bounce-backs and sent a reminder email two weeks later.

In total, we received 242 responses. The final sample consists of 173 respondents.[29] The numbers of observations vary slightly in the succeeding tables, due to respondents answering "N/A" at times or omitting answers entirely.

Comparing the answers from WOIC participants to Druid participants to assess whether they account for our relevant population of innovation management scholars, we find that the share of scientists with industry affiliations is higher for WOIC participants. This derives from the mere fact that WOIC focuses more on practitioners' work and problems than Druid. The only large other differences in characteristics occur for qualitative researchers. While about half of the WOIC participants conduct qualitative research, only about one-third of the Druid participants conduct qualitative research. The remainder of the characteristics (e.g. gender, reviewers for FT-50, …) differ only slightly between the two groups.

Furthermore, due to the anonymous responses, we cannot identify who participated and who did not participate in the study. Nevertheless, we assessed the potential implications of non-respondent analysis by comparing the characteristics of early respondents (those that replied before we sent the reminder e-mail) to late respondents (those that replied after we sent the reminder e-mail). The largest difference lies within non-European researchers, as they constitute about 20% of the early respondents, but only about 10% of the late respondents. Nevertheless, the actual models in the following section include controls for all characteristics.[30]

5.4.2 Variables

The survey asked for sociodemographic and job-related information as well as respondents' experiences with and attitudes towards open data. Furthermore, we asked respondents to indicate their opinions about costs and benefits associated with open data on five-point Likert-scales ranging from "strongly disagree" to "strongly agree". Separate questions focused on respondents' attitudes towards institutional pressure and replications.

5.4.2.1 Dependent variables

We conceptualize Data Shared as the dependent variable to capture the public data sharing behavior of innovation scholars. We derive this variable from respondents' answers on a scale

[29] We exclude 27 respondents due to them indicating their occupation as 'member or founder of a company ("industry")' or 'Other'. Furthermore, 42 respondents who did not answer all sociodemographic and job-related questions located at the first page of the questionnaire as well as those that neither conducted qualitative nor quantitative data were pruned from the sample.

[30] Full results of the WOIC vs. Druid analyses and the early vs. late respondents analyses are available from the corresponding author upon requests.

ranging from 0 to 100 to the statement “In my estimation, the following percentage of my data is openly available for everyone”. This extends previous research (Kim, & Stanton, 2015; Kim & Zhang, 2015) as prior work used scholars’ intentions to share data as the dependent variable. Yet already Ajzen (1985) highlighted that while, in general, intentions induce behavior, this is not always the case. This might especially apply for questions inducing socially desirable responding. As a case in point, Persoskie and Nelson (2013) showed that while two thirds of their respondents wanted to quit smoking, only half of them stated that they attempted to quit smoking. As the goal of this study is to highlight the state of the art of data sharing among innovation management scholars, we employ the actual amount of shared data. In addition, by asking respondents to indicate the percentage of scholars’ shared data we reduce common method bias as we employ two different scales for the dependent and the independent variables (Podsakoff et al., 2003). Following Podsakoff et al., (2003), we further aimed to reduce common method bias by implicitly and explicitly assuring respondents that all responses are anonymous. We sent out the same questionnaire link to all respondents individually instead of sending individual links to individual respondents which implicitly ensures anonymity. Also, we included the following statement in the first paragraph of the first page of the questionnaire which explicitly was aimed to ensure anonymity: ”We are aware that we touch upon a potentially sensitive area and will therefore ensure that all responses and participant information will be anonymized.”

5.4.2.2 Explanatory Variables

We operationalize our hypotheses as follows. We employ scholars’ standardized agreements with several statements anchored on a five-point Likert scale ranging from strongly disagree to strongly agree. For analysing Hypotheses 1a through 1c and Hypothesis 2, we use single item statements addressing respondents' opinions towards data as Proprietary Source, the Direct Costs of data sharing, the fear of Reputational Costs arising from data sharing as well as the Cost-Reward Imbalance as the independent variables.

For addressing the set of Hypotheses 3, we employ common factor analysis (principal axis factoring option in Stata 15) to generate three reflective latent composite variables, Community Benefits, Institutional Pressure and Replications, out of four statements each. In line with our theoretical reasoning for Hypothesis 3a, Community Benefits composite standardized statements address transparency in research, reducing fraud opportunities and increasing collaboration opportunities. Researchers’ opinions on whether journals should implement policies enforcing data sharing (for review, at publication or after a twelve months grace period) as well as whether publishers should establish licensing policies for the free reuse of data represent the underlying statements for Institutional Pressure (Savage, & Vickers, 2009). Replications capture innovation scholars' attitudes towards both, direct and conceptual replications as well as on replications of their own work and the work of others (Schmidt, 2009; Jasny et al., 2011; Hopp, & Hoover, 2019). The items used to measure each construct as well as reliability and validity statistics are provided in Table 5.2.

5.4.2.3 Control variables

We control for sociodemographic and job-related factors for each respondent. Leahay (2006) shows that gender differences exist in regard to academic productivity. Hence, we include respondents' gender (Female = 1, 0 otherwise). Furthermore, we control for the location of the university at which respondents work (Europe = 1, 0 otherwise) as cultural habits affect research and publishing processes (Salita, 2010). We also control for the professional level (Full Professor = 1, 0 otherwise) (Carayol, & Matt, 2006). In addition, we consider the number of peer-reviewed articles (0 = no articles, 1 = 1 to 5 articles, 2 = 6 to 10 articles, 3 = 10 to 20 articles, 4 = more than 20 articles) and the number of FT-50 articles (0 = no articles, 1 = 1 to 5 articles, 2 = 6 to 10 articles, 3 = 10 to 20 articles, 4 = more than 20 articles)[31] published from 2013 to 2018. In addition, we consider whether respondents acted as reviewer for an FT-50 journal in the year leading up to the survey (Reviewed for FT-50 = 1, 0 otherwise) and whether or not they have held editorships at an FT-50 journal since 2013 (Editor at FT-50 = 1, 0 otherwise). Last, existing literature highlights that sharing qualitative data might in fact be less common than sharing quantitative data (van den Berg, 2005; Aguinis, & Solarino, 2019). Therefore, we control for respondents' research approaches (Quantitative = 1, 0 otherwise; Qualitative = 1, 0 otherwise, Theoretical = 1, 0 otherwise). [32]

5.5. Results

5.5.1 Descriptive statistics

For the dependent variable, Figure 5.2 depicts the Data Shared by respondents. We find that scholars in our sample indicate to have shared on average 28.95% of their data (red line) with a standard deviation of 28.67%. This implies that an overwhelming majority of the respondents have shared only a quarter or even less of their data, with 24 respondents having never shared

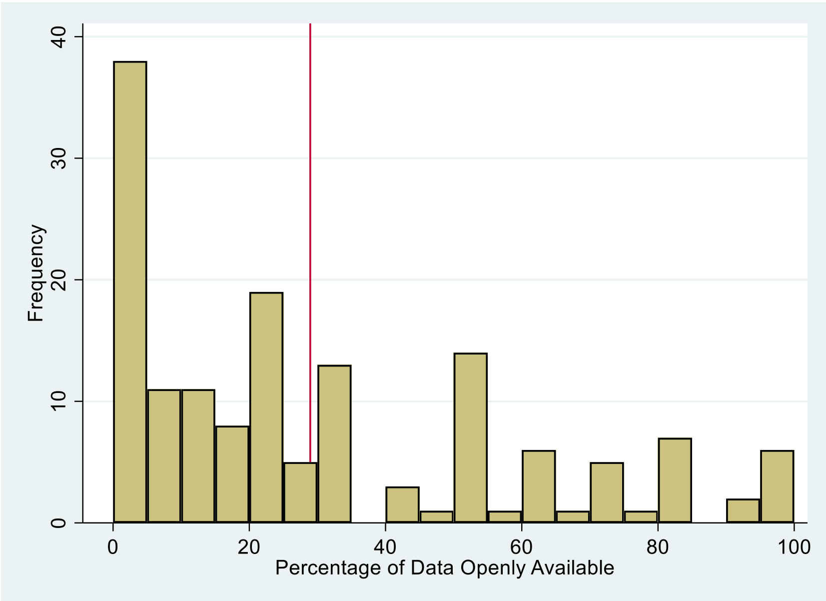
[31] To help respondents identify which journals are FT-50 journals, the survey included a link to the website of the Financial Times listing the FT-50 journals as of the 12th of September 2016.

[32] The variable coding is not exclusive. For example, scholars can engage in quantitative and theoretical research (*Quantitative* = 1; *Qualitative* = 0; *Theoretical* = 1).

any of their data. The spike in the middle of the histogram indicates 12 respondents who have shared half of their data. Interestingly, five respondents indicate to have shared all their data.

Table 5.1 depicts the descriptive statistics of the control variables. Nearly one-third of our respondents are *Female,* and more than 80% of them work at a *European* university. *Full Professor*s constitute approximately one quarter of the sample, and nearly fifty percent have *Reviewed for FT-50* journals at least once. 24 respondents have been *Editor*s *at FT-50* journals since 2013. Most of the respondents conduct *Quantitative* research (157) and have published between one to ten papers from 2013 to 2018, with a lower number of articles in FT-50 journals.

Figure 5.2: Histogram of the percentage data shared by scholars



Note: Bars show the frequency of respondents stating that the respective percentage of their data is openly available for everyone. The red line indicates the mean at 28.95%.

Table 5.1: Descriptive statistics of control variables

Total Respondents:				**173**	
Female				52	30.06%
Europe				143	82.66%
Full Professor				44	25.43%
Reviewed for FT-50				93	53.76%
Editor at FT-50				24	13.87%
Quantitative				157	90.75%
Qualitative				72	41.62%
Theoretical				60	34.64%
Number of…	**0**	**1-5**	**6-10**	**10-20**	**>20**
… Articles	31	63	38	30	11
… FT-50 Articles	80	76	9	6	2

Table 5.2 provides information on the explanatory variables. For the statements addressing the costs of data sharing, we find that innovation management scholars' opinions on whether data should remain a *Proprietary Source* differ, as about 45% of them agree with this statement while more than 25% of them disagree with it. The same applies to the *Direct Costs* of open data. Interestingly, more than two thirds of the survey participants agree that data sharing might bring along *Reputational Costs.* Assessing the *Cost-Reward Imbalance*, nearly every second scholar disagrees that the efforts outweigh the rewards. Addressing the *Community Benefits,* more than three quarters of the respondents (strongly) agree with each of the listed statements except for the positive effects of data sharing on collaboration and productivity, which are supported by less than two thirds of the respondents.

Concerning *Institutional Pressure*, respondents are undecided about whether journals should require public data sharing upon publication. Yet, two thirds of the survey participants are in favour of journal policies that force authors to share their data after a twelve-month grace period. Unsurprisingly, an even higher majority of scholars (strongly) agree that data and source code should be available at the time of submission to allow for a transparent review process. In addition, many respondents also would like to see licensing policies that allow the free reuse of author-submitted data. Last, respondents indicated their opinions on the benefits of *Replications* of the work of others and on replications of their own work. The majority of the innovation management scholars (strongly) agree on the importance of exact and conceptual replications for the work of others and their own work.[33]

5.5.2 Validity and reliability

Assessing empirically the validity of the explanatory variables, we find that all composite variables possess Eigenvalues greater than 1 and therefore fulfil the Kaiser criterion (Kaiser 1960). Furthermore, all component loadings are higher than 0.65, indicating the high relevance of the observed individual variables for the composite variables (DeCoster, 1998). Regarding the reliability of the composite variables, Cronbach's Alphas corresponds to 0.70 for Community Benefits, 0.83 for Institutional Pressure, and 0.72 for Replications.[34]

[33] We investigate whether opinions towards the replication of other´s work differ from the opinions towards the replication of own work using Wilcoxon signed-rank tests. We find that scholars' preferences do not differ neither for exact (z=0.495, p>0.1) nor for conceptual replications (z=-0.142, p>0.1).

[34] As our sample consists of innovation scholars only, and as the observed variables only contain about 150 observations, *Communal Benefits* and *Individual Costs and Risk* just exceed the ubiquitous thresholds for Cronbach's alpha of 0.7. Bernardi (1994) empirically showed that small homogenous samples often tend to have lower Cronbach's alphas. Following Nunnally and Bernstein (1994), we conclude to have reliable variables.

Table 5.2: Descriptive statistics of the explanatory variables

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Total
Proprietary Source						
Empirical models and datasets are large personal investments for researchers and should remain “trade secrets”.	25 17.61%	40 28.17%	39 27.46%	27 19.01%	11 7.75%	142
Direct Costs						
The time and effort it takes for the curation, documentation, standardization, normalization, and metadata association for research data are too large to make the data publicly available.	10 7.04%	43 30.28%	38 26.76%	38 26.76%	13 9.15%	142
Reputational Costs						
Releasing flawed code or data can discredit research results and cause embarrassment to the releasing authors.	2 1.41%	18 12.68%	18 12.68%	79 55.63%	25 17.61%	142
Imbalance of Costs and Rewards						
The effort of anonymizing data is too high, the reward too low.	18 12.68%	51 35.92%	27 19.01%	33 23.24%	13 9.15%	142

Community Benefits (μ=0; δ=0.8155; α=0.70)	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Total
Transparency, reproducibility and traceability in scientific research are almost impossible to implement without access to original data.	3 1.92%	19 12.18%	15 9.62%	66 42.31%	53 33.97%	156
Releasing data could help to expose cases of fraud and scientific misrepresentation.	2 1.28%	4 2.56%	10 6.41%	61 39.10%	79 50.64%	156
Public disclosure of research data helps to manage conflicts of interest and discourage misconduct.	1 0.64%	9 5.77%	21 13.46%	74 47.44%	51 32.69%	156
Releasing data could increase collaboration among scientists and increase academic productivity.	5 3.21%	19 12.18%	31 19.87%	55 35.26%	46 29.49%	156

Institutional Pressure (μ=0; δ=0.9219; α=0.83)	**Strongly Disagree**	**Disagree**	**Neutral**	**Agree**	**Strongly Agree**	Total
Journals should require authors to make their original materials and data publicly available (at publication).	14 9.72%	37 25.69%	34 23.61%	43 29.86%	16 11.11%	144
Journals should require authors to make their original materials and data publicly available (after a 12-months period, the latest).	12 8.33%	23 15.97%	28 19.44%	57 39.58%	24 16.67%	144
Data and source code should be available at the time of submission to allow for a transparent review process and the verifiability of data.	9 6.25%	22 15.28%	24 16.67%	55 38.19%	34 23.61%	144
Publishers should implement licensing policies that allow the free reuse of author-submitted data.	9 6.25%	22 15.28%	29 20.14%	51 35.42%	33 22.92%	144

Replications (μ=0; δ=0.8574; α=0.72)	**Strongly Disagree**	**Disagree**	**Neutral**	**Agree**	**Strongly Agree**	Total
It is important to exactly replicate (e.g. using the same material, manipulations, methods, dependent variable, etc. as the original study) research others have already conducted and published.	2 1.37%	17 11.64%	41 28.08%	52 35.62%	34 23.29%	146
It is important to conceptually replicate (e. g. using the same fundamental hypothesis, but different designs, and variables) research others have already conducted and published	1 0.68%	9 6.16%	15 10.27%	70 47.95%	51 34.93%	146
I would be glad if other researchers would exactly replicate (e.g. using the same materials, manipulations, methods, dependent variable, etc. as the original study) my published studies	4 2.74%	11 7.53%	38 26.03%	65 44.52%	28 19.18%	146
I would be glad if other researcher would conceptually replicate (e.g. using the same fundamental hypothesis, but different designs, and variables) my published studies	1 0.68%	4 2.74%	20 13.70%	71 48.63%	50 34.25%	146

Table 5.3 shows the correlation matrix for all variables. The absolute highest pairwise correlation exists between Direct Costs and Cost-Reward Imbalance with a correlation coefficient corresponding to 0.67. As this value is far below the problematic cases of 0.8 to 0.9, our analysis does not seem to suffer from multicollinearity (Mansfield and Helms, 1982). In addition, all variance inflation factors are smaller than two and hence we do not face any multicollinearity issues (Dormann et al., 2010).

We further employ the correlation matrix to check for common method bias by using the K*=1/r rule (Siemsen et al., 2010). The smallest bivariate correlation between our explanatory variables and the dependent variables amounts to 0.1080 between *Data Shared* and *Communal Benefits.* Hence, we need to include at least 10 independent variables (explanatory and control variables) in the regression to fulfil the K*=1/r rule. As our full model contains in total 14 explanatory variables, our regression estimates are not very likely to suffer from common method bias. As common-method variance equates to an omitted variable problem, we also address the implications of omitted variables and endogeneity in our limitation section.

5.5.2 Regression analysis

Table 5.4 presents the regression results of the effect of scholars' opinions on their own data sharing. We report coefficients from negative binominal regressions to account for the over-dispersed count variable distribution of the dependent variable. Model 1 solely contains the sociodemographic and job-related control variables. We find that those who have *Reviewed for FT-50* journals share less data, and *Quantitative* researchers share more data, though both coefficients are only marginally significant at the ten percent level. Model 2 includes data as *Proprietary Source.* As stipulated in Hypothesis 1a, the coefficient is significantly negative. Model 3 assesses the effects of the *Direct Costs* of data sharing. We find evidence in favour of Hypothesis 1b: Researchers believing that open data carries direct costs share significantly less data. Model 4 includes the *Reputational Costs,* but does not provide support for Hypothesis 1c; the coefficient is not significant at conventional levels.

Table 5.3: Correlation matrix

	Data Shared	Advantage	Direct Costs	Reputational Costs	Costs-Rewards Imbalance	Community Benefits	Institutional Pressure	Replications	Female
Data Shared	1.0000								
Proprietary Source	-0.3511***	1.0000							
Direct Costs	-0.3071***	0.3871***	1.0000						
Reputational Costs	-0.1779†	0.0186	0.1279	1.0000					
Cost-Reward Imbal.	-0.1170	0.2729**	0.6726***	0.1920*	1.0000				
Community Benefits	0.2225*	-0.4076***	-0.4837***	-0.0166	-0.3184***	1.0000			
Institutional Pressure	0.3012***	-0.4468***	-0.2382**	0.0076	-0.1022	0.4598***	1.0000		
Replications	0.3464***	-0.5109***	-0.3607***	-0.0245	-0.1128	0.5413***	0.5033***	1.0000	
Female	0.0078	0.0979	-0.1206	0.0428	-0.1723†	-0.0086	-0.1675†	-0.1295	1.0000
Europe	-0.1183	0.0551	0.0365	-0.0843	-0.0041	-0.0646	0.0031	0.0365	0.1231
Full Professor	-0.0641	0.0428	0.1002	0.0343	-0.0352	-0.1536†	-0.1486	-0.3031***	-0.0935
Articles	-0.0845	-0.0632	0.0709	-0.0212	0.0137	-0.0795	-0.1043	-0.1064	-0.0709
FT-50 Articles	-0.1289	0.1572†	0.1100	-0.1463	0.0505	-0.2982**	-0.2344*	-0.1819*	-0.0132
Reviewed for FT-50	-0.2025*	0.0287	0.0414	0.1839*	-0.0546	-0.2081*	-0.1657†	-0.1066	-0.0032
Editor for FT-50	-0.0578	0.1257	-0.0673	-0.0129	-0.0294	-0.0830	-0.0459	-0.0377	0.0368
Quantitative	0.1274	0.1758†	-0.0170	0.0184	0.0298	-0.0070	0.0226	0.0203	-0.0667
Qualitative	-0.0840	0.1394	-0.0005	-0.1688†	-0.0993	0.0039	-0.1463	-0.2779**	0.1327
Theory	0.0434	-0.1250	-0.1295	0.0366	-0.0691	-0.0205	-0.0244	-0.0589	0.0313

	Europe	Full Professor	Articles	FT-50 Articles	Reviewed for FT-50	Editor for FT-50	Quantitative	Qualitative	Theory
Europe	1.0000								
Full Professor	0.0044	1.0000							
Articles	-0.0846	0.3842***	1.0000						
FT-50 Articles	-0.0605	0.2630**	0.3129***	1.0000					
Reviewed for FT-50	-0.0146	0.1944*	0.2593**	0.3369***	1.0000				
Editor for FT-50	-0.3489***	0.2235*	0.1024	0.4647***	0.2708**	1.0000			
Quantitative	-0.0672	-0.0806	0.0250	0.1979*	0.2263*	0.1270	1.0000		
Qualitative	-0.0240	0.0901	0.0264	-0.0832	-0.2302*	0.1613†	-0.2559**	1.0000	
Theory	-0.1103	0.0625	-0.0960	-0.1993*	-0.1758†	0.0458	-0.2283*	0.3860***	1.0000

† $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Note: Pairwise correlation coefficients derived from Pearson-correlations.

Table 5.4: Regression antecedents to data shared

	Model 1 Data shared	Model 2 Data shared	Model 3 Data shared	Model 4 Data shared	Model 5 Data shared	Model 6 Data shared	Model 7 Data shared	Model 8 Data shared	Model 9 Data shared
Explanatory Variables									
Proprietary Source		-0.28***							-0.21†
		(0.08)							(0.11)
Direct Costs			-0.28**						-0.37*
			(0.09)						(0.16)
Reputational Costs				-0.13					-0.18†
				(0.11)					(0.11)
Cost-Reward Imbalance					-0.17†				0.22
					(0.09)				(0.14)
Community Benefits						0.22*			-0.38*
						(0.10)			(0.18)
Institutional Pressure							0.39***		0.36*
							(0.10)		(0.14)
Replications								0.32**	0.10
								(0.12)	(0.15)
Sociodemographics									
Female	-0.01	-0.07	0.01	0.09	-0.08	0.07	0.20	0.15	0.30
	(0.18)	(0.18)	(0.18)	(0.21)	(0.18)	(0.18)	(0.19)	(0.18)	(0.22)
Europe	-0.14	-0.11	-0.16	-0.27	-0.12	-0.07	-0.20	-0.07	-0.44
	(0.24)	(0.25)	(0.24)	(0.26)	(0.25)	(0.23)	(0.28)	(0.24)	(0.34)
Full Professor	-0.15	-0.15	-0.26	-0.06	-0.24	-0.05	0.11	-0.04	0.34
	(0.21)	(0.21)	(0.21)	(0.23)	(0.22)	(0.20)	(0.22)	(0.21)	(0.26)
Academic Commitment									
Articles	0.01	-0.01	0.00	0.01	-0.00	0.00	-0.00	0.00	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
FT-50 Articles	-0.02	-0.00	-0.00	-0.03	-0.01	-0.00	-0.02	-0.00	-0.03
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)
Reviewed for FT-50	-0.31†	-0.28	-0.26	-0.28	-0.37†	-0.30†	-0.18	-0.22	-0.15
	(0.18)	(0.18)	(0.19)	(0.19)	(0.20)	(0.18)	(0.19)	(0.18)	(0.24)
Editor at FT-50	-0.04	-0.11	-0.25	0.01	-0.10	-0.11	-0.25	-0.21	-0.30
	(0.30)	(0.30)	(0.30)	(0.30)	(0.29)	(0.28)	(0.34)	(0.27)	(0.34)

Approach									
Quantitative	0.59†	0.56	0.54	0.50	0.55†	0.48†	0.78*	0.63†	0.87*
	(0.30)	(0.34)	(0.34)	(0.34)	(0.32)	(0.28)	(0.31)	(0.33)	(0.37)
Qualitative	-0.10	0.05	-0.02	-0.16	-0.10	-0.22	0.07	-0.04	0.18
	(0.20)	(0.20)	(0.21)	(0.20)	(0.21)	(0.20)	(0.23)	(0.20)	(0.29)
Theoretical	0.27	0.16	0.20	0.30†	0.17	0.29†	0.20	0.17	-0.08
	(0.17)	(0.19)	(0.18)	(0.18)	(0.18)	(0.17)	(0.19)	(0.17)	(0.23)
Chi-Square	16.98	31.39	30.03	20.11	22.08	23.54	37.10	26.44	73.57
p > Chi-Square	0.075	0.001	0.002	0.044	0.024	0.015	0.000	0.006	0.000
Observations	142	139	136	138	133	139	134	136	118

† $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Note: Coefficients derived from negative binominal regressions with robust standard errors in parentheses.

Model 5 in Table 5.4 highlights that a perceived *Cost-Reward Imbalance* significantly reduces scholars' data sharing behavior, thus providing support for Hypothesis 2. Model 6 contains *Community Benefits* as the explanatory variable. We find support for Hypothesis 3a, as the coefficient for *Community Benefits* is significantly positive. Model 7 incorporates *Institutional Pressure* and finds support for Hypothesis 2b: Scholars envisioning *Institutional Pressure* share more of their data. Model 8 includes the perceived benefits of *Replications* and finds that this significantly positively relates to data sharing, thus supporting Hypothesis 2c.

The final Model 9 employs the full set of explanatory variables in one full model. While we still find support for Hypothesis 1a and 1b, the coefficient of the *Cost-Reward Imbalance* turns insignificant. Yet the coefficient of *Reputational Costs* is now significantly negative, thus providing support for Hypothesis 1c. Furthermore, the positive coefficient of *Community Benefits* turns negative. This exactly contradicts Hypothesis 3a. *Institutional Pressure* (H3b) still positively relates to data sharing, whereas the *Replications* coefficient is not significant anymore (H3c). Summing up, we find consistent support for Hypothesis 1a, 1b and 1c as well as Hypothesis 3b in the full models. In addition, we find evidence in the opposite direction of Hypothesis 3a when including all variables. We do not find consistent support for Hypotheses 2 and 3c.

5.3 Robustness tests

We conduct a series of robustness tests. First, we employ a different, yet similar dependent variable measuring respondents' willingness to share research data upon paper acceptance (refer to Table 5.6 in the Appendix). Table 5.7 in the Appendix shows the marginal effects stemming from ordered logistic regressions. The results from the models including each independent variable separately align with our original results. Yet, several significance levels change when we include all variables into the same regression in Model (18). For the sake of simplicity, Table 5.5 contains an overview including the hypotheses, the corresponding variables and whether we find support in the original models and/or in the robustness test models.

Throughout both full models, we can only corroborate the significantly negative effect of *Reputational Costs* (H1c) and the positive effect of *Institutional Pressure* (H3b), while support for all other effects is mixed. Overall, Table 5 shows that more variables possess significant coefficients when including past data sharing behavior than when including data sharing intention. We address and explore these differences in the discussion.

Table 5.5: Overview of hypotheses, variables and support

Hypothesis	**Variable**	**Support in Table 4** (Partial / Full Model)	**Support in Table 7** (Partial / Full Model)
Hypothesis 1a: *The stronger researchers believe that possessing non-public datasets provide them with an advantage over other researchers, the lower the likelihood of data sharing.*	Proprietary Source	Yes / Yes	Yes / No
Hypothesis 1b: *The larger researchers perceive the direct costs of data sharing, the lower the likelihood of data sharing.*	Direct Costs	Yes / Yes	Yes / No
Hypothesis 1c: *The larger researchers perceive the reputational costs of data sharing related, the lower the likelihood of data sharing.*	Reputational Costs	No / Yes	Yes / Yes
Hypothesis 2: *The stronger researchers believe that the costs associated with open data outweigh the rewards, the lower their likelihood of data sharing.*	Cost-Reward Imbalance	Yes / No	Yes / No
Hypothesis 3a: *The more affirmative scientists are to the community-wide benefits of data sharing, the higher the likelihood of data sharing.*	Community Benefits	Yes / No*	Yes / No
Hypothesis 3b: *The more affirmative scientists are to institutional pressure to increase data sharing, the higher the likelihood of data sharing.*	Institutional Pressure	Yes / Yes	Yes / Yes
Hypothesis 3c: *The more affirmative scientists are to the proliferation of replication studies, the higher the likelihood of data sharing.*	Replications	Yes / No	Yes / No

Note: * indicates significant effect into the opposite direction of the hypothesis

Second, estimating all models with non-robust standard errors does not affect any significance level. Third, running standard OLS regressions instead of negative binominal regressions does not qualitatively affect the results. Sixth, not standardizing the variables and taking the mean instead of employing principal component analysis to generate the four composite variables do not change any directions and implications. Seventh, we exclude the control variables and estimated separate models. Results are again invariant. Last, we estimated all models using only the 113 observations for which we have full data (no N/A answers). This, again, does not alter the findings.

5.6 Discussion

The importance of open data for science *and* society is unquestionable. Yet achieving communal benefits is conditional on how individual researchers perceive the advantages and disadvantages of open data sharing. As it relates to the costs of data sharing, we find that those researchers, who believe that their datasets are trade secrets less often shared their data publicly. This corresponds to existing research on resource sharing (Das, & Teng, 2000; Silverman, 1999) and implies that researchers, in fact, view data as a proprietary source. Our research therefore contributes to prior work with the empirical observation that it is important to overcome the potential loss in future publication opportunities to encourage follow-up knowledge reuse through data disclosure.

Along these lines, our results also attest to the important role of individual borne costs of data collection. In line with our hypotheses, we find that the larger the direct and reputational costs, the lower is the likelihood of data sharing. This extends prior research by not only emphasizing that data collection, curation, and preparation bind critical time and resources which hinder data sharing subsequently but also by showing that researchers pay attention to potential reputational costs associated with data sharing. We extend prior findings by showing that a fear of embarrassment and a loss of reputation from flawed code or data also hinders researchers from sharing their datasets publicly. This goes in line Barbour et al. (2017) and Lu et al. (2013), highlighting the strong negative impact of corrections and retractions on researchers' careers.

When it comes to potential ramifications that could ensure more future data sharing, we find evidence in favour of our hypotheses regarding the beneficial effects of institutional pressure. Institutional pressure with respect to open data might therefore be necessary to facilitate data sharing (Stodden et al., 2018). Our data reveals that innovation scholars that positively attest to journal policies for data sharing also made their data publicly accessible and also intend to engage in open data sharing more often. This supports the results from Savage and Vickers (2009) that highlight that scholars across disciplines are in favour of stricter journal guidelines. Essentially, this would increase the pressure to release data for everyone and would not single out researchers that need to weigh the costs and benefits individually.

Last, we find mixed evidence regarding the role of replication studies. Although many of our respondents are in favour of exact and conceptual replications, we find mixed evidence that an increased emphasis on replication studies leads to an increased willingness to share data. Given the fragility of our estimates, that could stem from the somewhat small sample studied, we would urge researchers to explore this notion further. As it results from our study, the findings regarding replications are fairly ironic: If researchers want others to replicate their research, they also need to share their data so that others can conduct the replication.

We conclude that the existing benefits and incentives do not provide sufficient motivation to encourage researchers to share their own data as only few innovation scholars, who might be even more prone to openness compared to other researchers, made their datasets publicly available. That is, while most of the innovation scholars responding to our survey strongly attest to the communal benefits of open data, they do not share their data subsequently. We conclude that this low prevalence (that is similar to other disciplines nonetheless; e.g. Fecher et al., 2017; Tenopir et al., 2015) highlights that individual incentives for academics in innovation management are not sufficiently developed to induce researchers to share data, despite good intentions.

5.6.1 Implications

Our analysis confirmed a paradoxical tension perceived by scholars in the (innovation) management sciences: While the communal benefits from open data are seen as manifold, the overall costs and risks and the limited prevailing individual benefits of data sharing demotivate individual researchers to open their data. Open data, hence, is a perfect illustration of the challenges of governance in the sciences, as outlined by Merton (1942), who reflected that science occurs in accordance with individualistic values and views, but is also always embedded into an organizational and institutional context. Evidently, the root causes for not sharing data lie in the academic incentive system. Considering data from a resource-based perspective provides a comprehensive explanation for the reluctance to share data upon publication. Researchers perceive of research data as a private strategic resource, which exclusive possession provides competitive advantage in the academic system.

Or, in other words, sharing this resource openly is perceived as inducing a comparative disadvantage for those who share data vis-a-vis those who exploit the data collection of others (Wilbanks, & Friend, 2016). Researchers might “free ride” in using an existing data set for answering further research questions, and by submitting and/or publishing similar papers faster than the collectors (Murray, 2016). Hence, data sharing is perceived by the originators of the data as a counterproductive strategy to enable other researchers to "win the publication race". Further, the effort associated with data preparation, curation, and publication is perceived as an additional opportunity cost, reducing the time and capacity available for publishing another paper.

This situation is unlikely to be overcome by an individual researcher alone, despite good intentions and knowledge of the public benefits for the scientific system of data sharing. Our research has revealed that even scholars in the innovation management discipline, whose core theories and academic discourses promote an open approach to science and innovation (Bogers et al., 2017; Randhawa et al., 2016; von Hippel, 2017; West et al., 2014), do not share data openly. Alas, researchers are, after all, amenable to weighing of costs and incentives.

As Bogers (2011b:110) concludes, “the use of a knowledge exchange strategy in general and licensing in particular is the way in which firms shape the dimensions in the tension field to balance the sharing and protection of knowledge.” We therefore envision two different approaches to overcome this open innovation paradox (Bogers, 2011b): *(i)* on the institutional level, changing the incentive system for researchers to stimulate data sharing; and *(ii)* on the individual level, educating researchers in practices of "strategic openness" (Alexy et al., 2018), which allows them to freely reveal their data and still profit from this behavior individually, even under unchanged incentive regimes on the institutional level. We will elaborate on both directions in more detail in the following.

On the institutional level, we need to amend the academic incentive system. To increase scientists’ willingness to share their data, we echo Wilbanks and Friend (2016) who call for an academic performance measure system that does not only reward the publication of an article, but also the publication of its data. As we all know, researchers are receptive to publication incentives, with all shortcomings. As such, they should react favourably to open data incentives. Realizing those requires several concrete measures:

First, datasets need to be "publishable and citable" (Reichman et al., 2011: 704) so that we can include them in scholars' publication and citation counts. This demands a clear standard and syntax for the reference of a data set, which also needs to be searchable in the same literature databases used to find research results (papers), e.g., GoogleScholar or ScienceDirect. Publishers (such as Wiley or Elsevier) and open source repositories such as DataVerse already provide researchers with the opportunity to publish a dataset with a unique digital object identifier (DOI). A general approach in this area is the FAIR Principles, a multinational initiative to provide guidelines for the publication of research datasets or code in a manner that makes them "Findable, Accessible, Interoperable, and Reusable (FAIR)" (Wilkinson et al., 2016). On the national level, currently large-scale schemes are taking place to build (national) research data infrastructures. In Germany, for example, the aim of its National Research Data Infrastructure (NFDI) consortium is to systematically manage scientific and research data, provide long-term data storage, backup and accessibility, and network the data both nationally and internationally (Grunzke et al., 2017). Similarly, in Australia, the Research Data Infrastructure initiative by the Australian Department of Health funds the creation of data infrastructure with a focus on data registries, biobanks, and data linkage platforms in the area of medical research. On a multinational scale, EUDAT, the European Data e-Infrastructure Initiative, is working to construct and realize a global research data infrastructure. In short, these infrastructures are important supporting institutions so that research data made open can be stored and accessed by other scientists (Mons et al., 2017). In the management sciences, these initiatives, however, are largely unknown or, at least, not high on the awareness level, as compared to the medical or natural sciences. Professional education and awareness building for these initiatives, for example by professional organizations like PDMA or ISPIM for the field of innovation management, or AACSB and AOM for management research in general, would be an important element to foster open data sharing by management scholars.

Second, obviously sufficient data needs to be supplied into these infrastructures. For this, data sharing by a scientist has to become visible. Impact indicators in databases like Scopus, GoogleScholar, or SSCI should not only measure how often a publication has been cited, but also could indicate how often data provided by a certain author has been reused and cited.

This would allow the users of academic impact measures (like recruitment or tenure committees, academic associations, or grant-giving institutions) to recognize not only how often a researcher has published in prestigious journals, but also whether her or his results have been confirmed and replicated by others or not. In addition to incentivizing researchers (who are, as our data indicates, by and large open to replication studies) to share data more openly when they themselves receive credit for replicated studies, the central outcome would be the overall benefit from replication studies for the scientific process. For the individual researcher, a system that honours high citation counts of revealed data sets would foster good "citizen behavior" of scientists, making their service of data sharing to the academic community measurable. Platforms like *Publons* follow a similar approach: They recognize and incentivize peer review engagements of scholars and make them visible for academic performance measures (Ortega, 2017; da Silva, & Al-Khatib, 2019).

As a potential ramification, grace periods could give the data collectors a short advantage over potential research parasites in case they want to answer multiple research questions in multiple papers with the same dataset (Kirkman, & Chen, 2011). Indeed, the innovation scholars surveyed in our study aspire journal policies enforcing data sharing only twelve months after publication.

Third, there are also short-term measures that can be implemented immediately by editors of management journals. Kidwell et al. (2016) show that low-cost nudging can already increase the level of data that is shared. The journal *Psychological Science*, for example, introduced badges that visually signalled that data and material were available for interested readers of said articles. Data sharing increased from less than three percent to almost 40 percent within a two-year timespan (Kidwell et al., 2016). In biomedical research, the existence of data sharing policies correlates positively with the impact factor of the journal (Vasilevsky et al, 2017). This mirrors prior findings in the information science community (Sturges et al., 2015). However, we are currently not aware of any similar practices in a management journal.

Fourth, university and research institutions should support scientists in the data curation and publication process. As a case in point, some university library scientists believe that curation, preparation, and publication of data should be taken over by university librarians who could focus on this task (Heidorn, 2011; Tenopir et al., 2015; Koltay, 2019). Especially as the actual job of collecting books, etc. gets reduced substantially. This could help researchers to publish their datasets without requiring them to make large individual efforts and might also counter potential job losses on the administrative front.

All these measures, however, demand that other academics make use of the published data sets. After all, Peters et al. (2016) show that about 85% of the citable available datasets remain uncited. While one can argue that this ratio is not worse than the citation records of many journal articles (Judge, et al. 2007), we consider building more demand for open data as a core measure. This is a fruitful area for future research, investigating the adoption and usage drivers (and barriers) of freely revealed research data. Corresponding open data policies can target the creation of incentives for authors conducting replication studies, i.e. by dedicating a section in a journal to such studies or by creating publication awards in this category. Also, frequently asking for replication and meta studies in editor's comments may foster more consumption of open data – stimulating in turn its supply.

In addition, there are also several measures related to the individual level that could help to foster data sharing, addressing a proactive strategic behavior of scientists beyond their reaction to institutional incentives. Here, transferring the concept of strategic openness to the context of open data could provide a fresh perspective. Alexy et al. (2018) proposed *strategic openness* in the context of open innovation, suggesting that organizations should voluntarily forfeit the control over strategically relevant resources. While such behaviors intuitively would hurt the organization, Alexy et al. (2018) show that companies can still maximize profitability by such behaviors, if, for example, they open parts of their resource base or use openness to find/create complementary services. The concept of strategic openness can be transferred to researchers considering sharing their data.

According to this perspective, research data is just part of a larger bundle of (research) resources, like data acquisition instruments, code for their analysis, data storage and management, or reporting tools helping to navigate the data. Hence, researchers openly sharing their data (i.e. giving away their strategic resource according to the RBV) could actually increase the value of the entire bundle of research resources connected with this data, if they only offer controlled access to the other resources. Having access to data as the open resource part of a bundle of research infrastructure hence can still provide control (without exclusive ownership) to the researcher (Alexy et al., 2018). Yet, to be able to fully understand how researchers arrived at prior conclusions might involve access to data coding and analysis files as well (Campbell, & Mau, 2019; Hopp et al., 2018).

Researchers further can derive "competitive advantage" from open data if the data has idiosyncratic features, that is, if the original researcher has superior information about and/or superior complementarities with the open resource a priori: The open data source may be available to everyone, but because the original researcher created it prior to release, he or she should hold superior information about what can be done with the data and superior complementarities on how to leverage it with other proprietary resources. Consequently, the original researcher can publish much faster than anyone trying to “free-ride” on the data collection.

Further, users of open data may be willing to "pay" for complementary services (payments in terms of references, acknowledgements, co-authorships, but also perhaps monetary payments for service provides) such as data integration or analysis in a specific context. We currently observe such a shift of competition in digitized industries, where commoditization of once-valuable (hardware) technology rewards organizations that have strong integration and services capabilities. Some scholars, for example, build a successful career, for example, by offering seminars and consulting on statistical methods they developed and explored first with own data. In light of the increasing economic relevancy of algorithms, the demand for open data access for the evaluation of the impact of digital platforms such as Google, Uber or Lyft is increasing as well (Barber, 2019; Scheiber, 2020). We might witness similar behavior among academics.

Concluding, our previous arguments indicate that researchers can be both "pulled" into sharing their data openly by institutional actors (like journals or academic societies) setting new incentives motivating this behavior and actively "push" data sharing by developing strategic openness as their own strategy to strive in the academic system. We see many opportunities for further research in studying these approaches in more details, either in experimental studies or by observing behavior of researchers in the field.

5.6.2 Limitations and future research

No study is perfect, and ours is no exception. We contacted 2,716 innovation researchers with valid contact details. This number is only a bit smaller than the 3,468 active members of the Technology and Innovation Management (TIM) division of the Academy of Management listed on its website as of September 2019. Considering that our final sample consists of 173 respondents, the approximate response rate is 6.40%. This compares to other recently conducted online surveys among scientists investigating research practices and academic misconduct (Hopp, &

Hoover, 2017; Liao et al., 2018). Yet, respondents selected themselves voluntarily into the sample by clicking on the link and conducting the survey. With regard to non-responses, out of 241 total responses, 68 were deemed insufficient for our sample. About 80% of incomplete responses had a progress of 3%, 24%, and 48%, respectively. 3% progress corresponds to the very first question, 24% progress corresponds to a drop-out at the question "How many scientific papers did you (co-)publish in peer-reviewed scientific journals (2013-2018)?", whereas 48% progress corresponds to a drop-out at "Views on journal data policies". As those innovation scholars highly interested in open science might more often be intrinsically motivated to take their time and efforts to complete the questionnaire, our results might suffer from sample bias. However, those innovation scholars highly interested in open science are also those who might more often engage in open practices like data sharing. Hence, our results at least provide an upper boundary for the level of data sharing among all innovation scholars.

This article only provides insights into the topic of open data eliciting opinions of innovation scholars. While this certainly limits the generalizability of the findings and makes them context-specific, it is important to bear in mind that this community explicitly studies the merits of openness. It therefore stands to reason that other fields of management research might emphasize even more pronounced costs of data sharing and place a lower emphasis on the benefits that it might provide. As such, a follow-up study could extend the research question beyond innovation scholars. Hereby, the focus could lie on increasing the sample by investigating differences in data sharing between various management fields or even between various social science disciplines.

Our research only elicited personal viewpoints, and individuals may certainly state that they share research data when in fact they never do so in real-life. A substantial number of innovation scholars indicates that they share their research data upon request. However, as noted in the literature review, existing empirical research points out that many scientists do not provide data upon request even if they included such data sharing statements in their articles (Wicherts et al., 2006; Krawczyk, & Reuben, 2012; Reidpath, & Allotey, 2001; Savage, & Vickers, 2009). Evidently, this may introduce a common-method variance bias, an omitted variable problem that potential invites the risk of endogeneity. There could be other non-measured variables that prevent those that indicate their willingness to share their datasets from actually sharing them; or they indicate willingness to comply with journal requirements without actually having the intention to share. This in turn, may open up room for future analyses that explicitly take stock of potential omitted variables due to common-method variance. Prior research has pointed out that ubiquitous tests are insufficient here (Richardson et al., 2009; Antonakis et al., 2010). Future studies might therefore rely on instrumental variable techniques to recover true parameter estimates.

This is important, as in our survey, we did not ask respondents for evidence on their actual sharing behavior or qualitative reasoning of their answers to keep the survey as anonymous as possible. To overcome this problem, we included respondents' willingness to share their data in the robustness section. We find that the implications differ between the two models as several coefficients, especially those capturing the costs of data sharing, turn insignificant. Yet researchers who have not shared data publicly so far might under- or overestimate the costs of data sharing due to their lack of experience. At times, there might even be institutional restrictions (data protection or confidentiality agreements) that prevent data sharing, despite good intentions.

Furthermore, they might also overstate their willingness to share their data due to socially desirable responding. This represents a common problem of research on questionable research practices and academic misconduct (Spaulding, 2009). A follow-up study could compare the survey results to the actual amount of shared data available on journal websites and data repositories. This would allow a more comprehensive and detailed understanding and would allow the assessment of sample and response biases among innovation scholars. In addition, it would also be interesting to run a similar survey using techniques that would increase respondents' perceived anonymity and in turn decrease socially desirable responding.

Last, our study revealed that qualitative researchers struggle more with making their data public (Pratt et al., 2020). Especially in light of the growing interest in replicating qualitative research, this seems to be a promising area for further research (e.g. King, 1995; Aguinis, & Solarino, 2019). It would be very interesting to see whether new technologies and processes exist that could help researchers to share qualitative data – and assist others in using this data (Antes et al., 2018).

5.7 Conclusion

Our research was sparked by the observations that, despite the proclaimed benefits of openness, top ranked innovation publications do not openly share the research data behind their articles. Consequently, we analysed this conundrum and elicited innovation scholars' attitudes towards open data and assessed their data sharing behaviors. The responses indicate that most scholars would be open to share their data upon request. Yet, data sharing is generally not very prevalent. Despite the generally acclaimed societal benefits, researchers refrain from putting their data into the public space. We identify antecedents and more importantly, inhibitors to data sharing behavior that could potentially provide policy implications. First, researcher behavior is susceptible to potential costs and threats that open data might provide. Essentially, when considering the "social dilemma" of open data (Linek et al., 2017: 1), the identified personal incentives to open data sharing might not outweigh the burden open data places on individual researchers. In summary, if open data sharing is to catch up, the burden for data preparation cannot be put on the individual researcher. Rather, institutional mechanisms on the level of the academic community need to be put into place: That could either be incentives that give more credit to data sharing, or journal policies that make data sharing a mandatory requirement for all publication. Yet, even with increased incentives the effort for data preparation rests upon the individual researcher. As an institutional remedy one might consider more administrative help for individual researchers.

5.8 Appendix

Table 5.6: Data sharing upon acceptance

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	N
Upon acceptance of a research article, I voluntarily make my original research material and data publicly available.	19	51	36	32	6	144
	13.19%	35.42%	25.00%	22.22%	4.17%	

Table 5.7: Regression antecedents to the willingness of data sharing upon acceptance of a research article (Table 5.6)

	Model 1 Acceptance data sharing	**Model 2** Acceptance data sharing	**Model 3** Acceptance data sharing	**Model 4** Acceptance data sharing	**Model 5** Acceptance data sharing	**Model 6** Acceptance data sharing	**Model 7** Acceptance data sharing	**Model 8** Acceptance data sharing	**Model 9** Acceptance data sharing
Proprietary Source		-1.03***							-0.44
		(0.23)							(0.30)
Direct Costs			-0.68**						-0.15
			(0.23)						(0.32)
Reputational Costs				-0.45*					-0.61*
				(0.21)					(0.25)
Cost-Reward Imbalance					-0.51**				-0.04
					(0.20)				(0.25)
Community Benefits						0.77**			-0.38
						(0.26)			(0.39)
Institutional Pressure							1.50***		1.31***
							(0.27)		(0.38)
Replications								0.94***	0.39
								(0.22)	(0.29)
Sociodemographics									
Female	-0.34	-0.14	-0.41	-0.21	-0.58	-0.15	0.25	0.06	0.45
	(0.36)	(0.37)	(0.38)	(0.39)	(0.37)	(0.34)	(0.36)	(0.37)	(0.44)
Europe	0.11	0.16	0.18	-0.02	0.23	0.16	-0.02	-0.05	-0.58
	(0.46)	(0.45)	(0.48)	(0.49)	(0.51)	(0.45)	(0.48)	(0.47)	(0.61)
Full Professor	-0.75†	-0.50	-0.74†	-0.63	-0.96*	-0.66	0.07	-0.62	0.15
	(0.40)	(0.40)	(0.43)	(0.46)	(0.46)	(0.45)	(0.49)	(0.44)	(0.66)

Academic Commitment									
Articles	0.04	0.00	0.04	0.04	0.03	0.03	0.01	0.03	-0.01
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
FT-50 Articles	-0.13*	-0.10†	-0.10†	-0.15**	-0.11*	-0.08	-0.07	-0.06	-0.09†
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)	(0.07)	(0.05)
Reviewed for FT-50	-0.74†	-0.68	-0.73†	-0.65	-0.67	-0.71	-0.72	-0.51	-0.62
	(0.43)	(0.42)	(0.43)	(0.42)	(0.43)	(0.45)	(0.45)	(0.43)	(0.50)
Editor at FT-50	0.64	0.34	0.30	0.78	0.63	0.47	-0.59	0.02	-0.54
	(0.73)	(0.71)	(0.72)	(0.78)	(0.72)	(0.76)	(0.69)	(0.80)	(0.74)
Approach									
Quantitative	1.27*	1.94***	1.25*	1.10†	1.14*	1.21*	1.65**	0.87	1.48*
	(0.54)	(0.59)	(0.55)	(0.58)	(0.56)	(0.57)	(0.61)	(0.53)	(0.65)
Qualitative	-0.22	0.35	-0.17	-0.39	-0.19	-0.32	0.52	0.14	0.61
	(0.40)	(0.48)	(0.43)	(0.41)	(0.41)	(0.38)	(0.44)	(0.42)	(0.53)
Theoretical	0.28	-0.14	0.08	0.37	0.19	0.31	0.21	0.29	-0.18
	(0.36)	(0.43)	(0.40)	(0.39)	(0.37)	(0.38)	(0.43)	(0.38)	(0.52)
Chi-Square	15.75	49.60	24.52	21.00	24.02	29.97	55.86	41.58	67.97
P > Chi-Square	0.107	0.000	0.011	0.033	0.013	0.002	0.000	0.000	0.000
Observations	136	133	131	132	128	133	130	132	115

† $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Note: Marginal effects derived from ordered logistic regressions with robust standard errors in parentheses.

6 Conclusion

The disruptive advancements in information and communication technologies heavily affected the social sciences. Thanks to the implementation of emails, conference calls and project management platforms, researchers from many different places can easily collaborate with each other (van Raan, 2001). Thanks to the introduction of online journals and publication databases, scholars can quickly search, identify and cite relevant literature (Emrouznejad et al., 2008). Thanks to the development of statistics software and machine learning scientists can process and analyze enormous amounts of data nearly instantly (Müller, & Guido, 2016). Overall, the improved infrastructure has substantially sped up research processes and has raised social scientific productivity significantly. As a result, social scientists publish today more articles in shorter time than ever before (Ossenblok et al., 2014).

In addition, journal databases eased the assessments of journals' impact factors and scholars' total number of publications and citations (Callaham et al., 2002). This substiantially changed researchers' achievements and merits assessments. Some authors stated that search and tenure commissions rely solely on impact factors, publications and citations in their decision making processes (e.g. McGrail et al., 2006; Park, & Gordon, 1996; Sestak et al., 2018). While such stark statements are at least debatable, there exists little doubt that "publish or perish" has become the motto of most disciplines (Kendall, & Campanario, 2016; Miller et al., 2011; Rawat, & Meena, 2014). Of course, publishing at all costs puts quantitiy over quality. The infamous misconduct cases of psychologist Dideriek Staple fabricating experimental data as well as business researcher Ulrich Lichtenthaler altering statistical results represented prime examples of how these incentives ended up inducing unethical behavior. However, those two cases revealed only the tip of the iceberg. In fact, journal editors and reviewers reported that they exhibit academic misconduct quite frequently (Hopp, & Hoover, 2017; Hopp, & Hoover, 2019). This has substantially reduced trust in social scientific results, creating a credibility crisis.

To reinsitute public trust in the social sciences, this thesis addressed the most pressing forms of unethical behaviors. Based on the existing literature, we discussed the dangers arising from HARKing and plagiarism and pointed towards existing and already implemented solutions. We moved on by analyzing authorship distributions and malpractices and showed that authorship problems, especially honorary authorship, occur frequently in the social sciences.

Consequently, we called for introducing existing policies from life and nature science journals enhancing merit-based authorship assignments in social science journals. Moving beyond scholars' credibility, we addressed the reliability of social scientific research. We highlighted how replication studies can help in assuring the credibility of scientific work. Consequently, we gave hands-on guidance on how to conduct thorough, rigor and extensive replications by successfully replicating and extending Kuhn and Weinberger (2005), an impactful leadership study that showed that team captainship and club presidency in high school relate to higher wages in later working life. Last, we investigated data availability, the pre-requiste for replication. By applying management theory, we showed that only few innovation management scholars make their data publicly available and that this might well stem from a mismatch between societal benefits and individuals' disadvantages that come along with open data.

So far, this thesis discussed the results and conclusions of each essay independently. While the introduction laid out the overall topic and showed how the essays relate to each other, we also elaborate on the implications arising from this thesis as a whole. The next section therefore covers the theoretical implications. This is followed by the practical recommendations. Last, we discuss areas and opportunities for future research.

6.1 Theoretical implications

This thesis adds to the existing literature stream on ethical behavior in the social sciences by returning several new insights. Using data from the social sciences, we reveal novel mechanisms influencing multiple forms of academic (mal)practices. In the following, we show that our findings do not only advance research on academic misconduct, but also add to the understanding of existing economic and management theories and how they apply to scholars' behaviors.

First, our results indicate that scholarly behavior often mirrors employee behavior. As a case in point, (Holmstrom, & Milgrom, 1991) highlighted that if employees' wages depend upon their piece rates, the quantitiy of their output increases whereas the quality decreases. Social scientists find themselves in a very similar scenario: The strong incentive focus on publications induces scholars to write, submit and resubmit papers as fast as possible (Kendall, & Campanario, 2016; Miller et al., 2011; Rawat, & Meena, 2014). In short, researchers only receive academic rewards for their output, not their efforts. This compares to output-based piece rate salaries. Lazear (1995) discussed that piece rate regimes incentivize employees to choose quantity becomes more important than quality. To overcome this problem, personnel economists pointed out that the loss of quality can be prohibited if there exist (often randomized) quality checks of the produced goods (Guiteras, & Jack, 2018). If such checks detect inferior quality, the employee(s) in charge for producing that product usually get(s) fine(s) dededucted from their wages (Heywood et al., 2013). In scientific research, we employ a very similar method. Editors and reviewers act as gatekeepers to ensure quality standards (Hojat et al., 2003). Obviously, if they perceive submissions to be of inferior quality, chances are high that they will reject this submission. Journal rejections represent fines for scholars because they require scientists to spend more time and efforts on looking for another suitable journal and adapting the manuscript for the new outlet. In fact, most social science journals already employ plagiarism check software and many editors and reviewers in the field already enforce high levels of rigorness especially in the data analysis (Shashok, & Handjani, 2010). Yet this thesis showed that numerous published research articles still lack authorship and data transparency. To improve this situation, we suggest to look at existing personnel economics solutions to quality loss in pay-for-performance systems and introducing similar techniques in academia.

Second, the lack of authorship transparency and more specifically the high prevalence of honorary authorship imply that freeriding is a contested topic in the social sciences. In general, Kandel and Lazear (1992) discussed that partnerships frequently induces freeriding. Hereby, partnerships refers to situations in which multiple individuals work together on a project, a perfect description of multiple researchers collaboratively working on a research article. After the finalization of the project, the whole team gets a reward (e.g. publications and citations) and distributes the reward among its members (e.g. first author, last author, …) (Kandel, & Lazear, 1992). If the team decides to split the reward equally among its members, it incentivizes each member to reduce his or her individual input (Kandel, & Lazera, 1992). This reduction in input also reduces the overall output and thus also the reward. Yet the reward reduction is split equally among all team members. Consequently, whereas the member reducing his or her input suffers only from the shared reduction in reward, he or she benfits completly from the withheld input (Backes-Gellner et al., 2004). Such freeriding is beneficial for team members as long as there exists at least one other team member whose input level exceeds zero (Backes-Gellner et al., 2004). We found that most social scientific publications possess multiple authors. It is therefore not surprising that this thesis showed that social scientists quite often exhibit freeriding, for example in the form of honorary authorship. Existing management research highlighted that peer-pressure as well as unequal reward distributions can reduce freeriding (Kandel,

& Lazear, 1992). This might also explain why academic freeriding occurs to a much lower degree than freeriding in for-profit organizations (Scalzini, 2015). A reason for this is the fact that social scientists tend to collaborate frequently with close colleagues and good friends (Colussi, 2018). Towry (2003) pointed out that closer social ties among team members substantially increase peer pressure and in turn reduce freeriding. Moreover, Wong et al. (2017) discussed that higher chances of future collaborations also reduce the likeliness of freeriding. This also applies to social scientists because Cummings and Kiesler (2008) pointed out that the more collaborations individuals had in past projects, the more likely they were to collaborate also in future projects.

Third, freeriding does not only apply to teamwork on collaborative research articles. According to our results, freeriding represents also one of the main reasons why most scholars do not share their data publicly. In fact, open data constitutes a prime example of a prisoners' dilemma. If all researchers shared their data, society and all scholars would be better off because this would enable faster, more accurate, more transparent and more reliable results (Vicente-Saez, & Martinez-Fuentes, 2018). Yet if all scholars except one shared their data, this scholar would still profit from the availability of an unimageinable high number of datasets but would still possess some proprietary data that he or she could exclusively use for new publications. Consequently, the optimal individual solution in which all scholars except the individual him/herself share their data outperforms the optimal societal solution of everyone sharing their data. This situation corresponds to a prisoners' dilemma (Colman, 1995). While there exists no game theoretical solution to the prisoners' dilemma, previous research has identified several practical solutions (e.g. omerta, cartel agreements, break-up fees) (Brembs, 1996). Indeed, this thesis highlighted that changing journal policies can alter the game theoretical settings of data sharing and thus can help us to overcome the open data prisoners' dilemma. In fact, we found that several management scholars would like to see more data sharing and stricter journal policies.

6.2 Practical Implications

This thesis contains manifold practical implications for various stakeholders. We start by addressing scholars per se. Then we discuss the effects on editors, reviewers, and journal publishers. Further we elaborate on the potentials we see for research societies. Afterwards, we address universities and research instiutions. Last, we highlight implications for everyone's daily life.

First, while this thesis indicated that academic malpractices are quite common in the social sciences, we neither found evidence that all scholarly behavior is unethical, nor did we suggest that all social scientists try to trick the system by cheating or taking shortcuts. In fact, addressing authorship problems, we showed that the majority of the investigated papers assigned authorship and contributorship correctly. Furthermore, our replication study successfully returned the same results and implications as the original study. Even for data sharing we found evidence that some scholars made all their data publicly available. Therefore, we need to praise those scholars upholding the high standards of ethical and responsible behavior in applied empirical research. They act as role models for their colleagues and students. In return, they positively enhance current and future academic conduct by showing that academic success and ethical behavior can go hand in hand. In addition, we reach especially out to junior faculty members. They live under the constant pressure of having to publish to increase their job application and tenure chances. Taking shortcuts or even behaving irresponsibly might increase those chances in the short term. Yet in the long term there always exists the possibility that other scholars could detect these irregularities. Such a detection would then compromise scholars' reputations and could lead, in the worst case, to job losses and the revoking of doctoral and/or habilitational titles.

Second, we address editors, reviewers and journal publishers. Those are the main gatekeepers deciding which research papers get published. Therefore, we call upon journal publishers to provide editors and reviewers with trainings on what constitutes (un)ethical behavior and on how to detect academic fraud, questionable research practices and authorship misconduct. Overall, social science journals should introduce increased transparency requirements for submissions. For example, journal guidelines should enforce at least transparency statements for author contributions and data disclosure. Hereby, authors should be forced to disclose their task distribution and whether the dataset(s) underlying the research article is publicly available and if not state reasons why not. Moreover, we urge editors to request the original data at least for the review process. This would enable reviewers to detect fabrication or falsification attempts. Last, we suggest that journals should introduce transparency badges for papers with high levels of rigor and transparency.

Third, research societies can also take actions to overcome the credibility crisis in the social sciences. Large societies like the American Sociological Association, the American Psychological Association and the Academy of Management could reduce the occurences of questionable research practices by clearly defining the borders of responsible and ethical behavior in research. The current code of ethics of these societies possess too many vague and ambigous paragraphs (Academiy of Management, 2018; American Psychological Association, 1983; American Sociological Association, 2018). This thesis highlighted that we urgently need new guidelines in specifically two areas: Authorship and data sharing. In both cases the largest social scientific research societies should work together to come up with uniform definitions and requirements applicable to all social scientific research fields. Standardized authorship criteria and data sharing protocols would substantially increase transparency in social scientific research. Furthermore, the societies should also generate uniform templates for authorship and data disclosure statements that journals could require upon submission. Last, large societies might not only set-up pre-registration platforms like the AEA Registry (2020) but also repositories for authorship disclosure statements and datasets. This would ease scholars' efforts to disclose this information even if the respective journal does not provide such opportunities.

Fourth, this thesis also contains important implications for universities and research instititions, the main employers for social scientists. To receive jobs at such organizations, researchers must successfully master search procedures. Nowadays, search procedures heavily focus on publication and citation counts (McGrail et al., 2006; Park, & Gordon, 1996). The same also applies to tenure procedures (Sestak et al., 2018). As tenured positions are among the top goals of many social scientists, it is not surprising that scholars often aim at maximizing their publications, sometimes by choosing quantiy over quality (Corley, 2005). Consequently, to ensure that researchers behave ethically and responsibly, universities and research institutions need to adapt their hiring and tenure decision criteria. As a case in point, instead of primarly looking at publications and citations, universities and research institutions could foster scientific credibility by including criteria capturing research rigorness and transparency in the evaluation processes. More specifically, search and tenure commissions could assess higher importance to publications that include technical appendices explaining all analyses in detail, that clearly indicate each authors' contribution and that make their data publicly available. This would motivate social scientists to overcome academic malpractices and focus on transparent analyses and reporting. In addition, universities and research institutions should introduce ombudsmen and whistleblowing platforms that would allow their employees to report suspicious behaviors, results and/or publications anonymously. These installations would enable earlier detections of fraudulent, unethical or irresponsible behavior. Universities and research institutions themselves would profit from ombudsmen and whistleblowing platforms because the early identification of scientific wrong doings empowers them to react quickly, probably even before the submission or publication of the work in question. Consequently, the organizations could handle misconduct cases internally without facing retractions and/or unwanted news coverage. This way, universities and research institutions could counter academic malpractices without having to fear reputation losses.

Last, this thesis addresses the general public. The emergence of terms like *fake news* and *alternative facts* symptomatically highlights the growing concerns on which information we can rely on. Nowadays, people can access more information than ever before via social media, internet media coverage, TV news, newspapers, etc. (Peters, 2018). However, the rapid growth in available information has not eased the identification of what is correct and what is wrong but instead has made it even more difficult (De Keersmaecker, & Roets, 2017). To overcome this issue, Rose-Wiles (2018) showed that students perceived peer-reviewed journals to provide reliable facts and results. Yet this thesis showed that we cannot even trust all information published in scientific journals. We therefore urge the public to not simply rely on stories and reports based on academic publications but instead to read the original publications and make up their own mind of the reliability and transparency of the information provided. Especially in economics and management, practitioners and policy makers should not blindly follow scholars' recommendations but instead critically question and scrutinize them first. This is the only way we can assure that we make profound decisions based on reliable information.

6.3 Implications for Future Research

Extensive research has covered the credibility crisis in the social sciences. Yet this thesis revealed several areas that require further investigations and provides opportunities for future research. In the following we elaborate on these areas.

First, our results show that we still have a long way to go to overcome the credibility crisis in the social science. Yet the assessment of academic malpractices in this thesis stems largely from self-reporting survey data and might therefore suffer from social desirabel responding. Consequently, this thesis pictures only a best-case scenario of how prevalent authorship misconduct and data sharing is. To reveal the actual current state of (un)ethical and (ir)responsible behavior in applied empirical research, future research should overcome social desirability in responding. This can be achieved in two ways: On the one hand, instead of directly asking respondents to indicate their amount of shared data, their contributions to a research article, etc. one could employ item-count and/or item-sum techniques to identify the actual occurences of certain academic malpractices. Hereby, respondents are randomly split into two groups and receive a set of question containing also a sensitive question. One group answers each question individually including the sensitive question whereas the other group returns a single answer for all questions together. The difference between the total count of non-sensitive questions in the first group and the answers in the second group then reveals the second groups' average response to the sensitive question (given that the composition does not differ between the two groups) (Trappmann et al., 2014). As a case in point, Hopp and Speil (2020) showed that students were much more likely to admit plagiarism in a survey using item-count techniques than in a survey asking them directly whether they engaged in plagiarism. Consequently, future research should be able to reveal higher (and more accurate) levels of academic malpractices using item-count and/or item-sum techniques. On the other hand, further investigations of (un)ethical and (ir)responsible scholarly behavior should work with secondary data. For example, researchers could combine data from research articles, Google Scholar, SCOPUS, WoS, journal websites and data repositories to investigate data sharing rates, authorship teams, transparency reporting standards and much more.

Second, this thesis points out that the credibility crisis in the social sciences does not solely derive from implausible research results but also from misconduct related to other aspects of scholarly work. We pointed out that, in addition to data related questionable research practices, the social sciences also face issues when it comes to research design, plagiarism and credit distribution. While identifying and overcoming academic malpractices occurring in the data collection, analysis and reporting constitutes valueable work, it is of utmost importance to also address and investigate less studied forms of (un)ethical and (ir)responsible behavior. As a case in point, we identified a surprisingly large share of honorary authors across the social sciences despite prior research on social science (mal)practices largely neglecting the topic of authorship. Consequently, future research on ethical and responsible behavior should look among others into topics like journal and reviewer coercion, open access policies, article processing charges, requirements arising from fundings and university-industry collaborations. Journal coercion occurs if journals request authors to cite work published in their journals to increase their impact factors (Wilhite, & Fong, 2012). Reviewer coercion occurs if reviewers suggest or require authors to cite work published by the reviewers to increase their citation counts (Baas, & Fennell, 2019). Both, journal and reviewer coercions alter citation counts but have so far not been investigated in the social sciences. Open access allows everyone to access publications without having to pay for publisher subscriptions (Suber, 2007). This becomes more and more important in the dissemination process of scientific results not only within academia but also to practitioners. To finance themselves, many open access journals introduced article processing charges (Morrison et al., 2015). Article processing charges usually represent fees that authors must pay in case their paper is accepted for publication (Marincola, 2003). Yet there exist several arguments against article processing charges. On the one hand, from an incentive perspective, article processing charges motivate journals to accept as many papers as possible as this generates more cash for them (Björk, & Solomon, 2015). On the other hand, article processing charges make it difficult for doctoral students and junior faculty members to publish in those journals because they often do not have access to extensive funding (Beall, 2013). Funding, in turn, might also represent a factor driving academic malpractices. Sismondo (2007) discussed that if clinical trials are funded by pharmaceutical companies, chances are higher that those clinical trials are successful. This leads to the last example, university-industry collaborations. Because companies usually want to keep their trade secrets to themselves, scholars collaborating with them on research projects often cannot transparently report on research processes and results (Rodriguez, 2005). Future research needs to address those academic (mal)practices also from a social science perspective as literature from the life and nature sciences showed that they substantially impact research rigorness and transparency.

Third, the vast majority of literature on scientific misconduct has focused on addressing and raising the problem. Now that we are aware that there exists a credibility crisis in the social sciences, we need to move beyond stating the problem and find adequate solutions. This thesis adds to this by providing hands on guidance on how to conduct a rigor and in-depth replication study. Future research should continue this stream by looking at opportunities and pathways helping us in overcoming the various forms of academic malpractices. Further investigations should therefore assess and outline the implementation of processes and tools that could help us in prohibiting scientific misconduct (e.g. requesting data for reviewers, introducing authorship statements, whistleblowing platforms). Moreover, future research articles and editorials should evaluate academic incentive systems that do not solely rely on publications and citations. If scholars do no longer have to publish more and more papers in less and less time, they can focus more on scientific rigorness, transparency and research ethics. The ultimate goal should be to dispose “publish or perish” as the motto of social scientists’ faculty lifes and substitute it with a phrase like the old but still valid expression “I know that I do not know”. This would better characterize the social sciences as we will never be able to unveil all interactions, mechanisms and drivers of social phenomena.

References

Aaltonen, T., Gonzalez, B. A., Amerio, S., Amidei, D., Anastassov, A., Annovi, A. et al. (2012a). Search for the standard model Higgs Boson decaying to a bb pair in events with two oppositely charged leptons using the full CDF data set. *Physical Review Letters 109*(11): 111803.

Aaltonen, T., Abazov, V. M., Abbott, B., Acharya, B. S., Adams, M., Adams, T. et al. (2012b). Evidence for a particle produced in association with weak Bosons and decaying to a bottom-antibottom Quark pair in Higgs Boson searches at the Tevatron. *Physical Review Letters 109*(7): 071804.

Aarsen, L. W., Tregenza, T., Budden, A. E., Lortie, C. J., Koricheva, J., & Leimu, R. (2008). Bang for your buck: Rejection rates and impact factors in ecological journals. *The Open Ecology Journal 1*(1): 14-19.

Abadie, A., Drukker, D., Herr, J. L., & Imbens, G. W. (2004). Implementing matching estimators for average treatment effects in Stata. *The Stata Journal 4*(3): 290-311.

Abbott, B. P., Abbott, R., Abbott, T. D., Abernathy, M. R., Acernese, F., Ackley, K. et al. (2016). Observation of gravitational waves from a binary black hole merger. *Physical Review Letters 116*(6): e061102.

Abbott, B. P., Abbott, R., Abbott, T. D., Acernese, f., Ackley, K., Adams, C. et al. (2017). GW170817: Observation of gravitational waves from a binary neutron star inspiral. *Physical Review Letters* 119(16): e161101.

Abramo, G., D'Angelo, C. A., & Di Costa, F. (2009). Research collaboration and productivity: Is there correlation?. *Higher Education 57*(2): 155-171.

Academy of Management Journal (2020). *Author Resources.* Academy of Management. Retrieved from https://aom.org/research/publishing-with-aom/author-resources [7th of September, 2020].

Academy of Management. (2018). *AOM Code of ethics*. Retrieved from http://aom.org/About-AOM/AOM-Code-of-Ethics.aspx. [18th of October, 2018].

Adam-Bourdarios, C., Cowan, G., Germain-Renaud, C., Guyon, I., Kégl, B., & Roussea, D. (2015). The Higgs machine learning challenge. *Journal of Physics: Conference Series 664*(7): 072015.

AEA Registry (2020). *AEA RCT Registry.* American Economic Association. Retrieved from https://www.socialscienceregistry.org [5th of September, 2020].

Aguinis, H., & Solarino, A. M. (2019). Transparency and replicability in qualitative research: The case of interview with elite informants. *Strategic Management Journal 40*(8): 1291-1315.

Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In Kuhl, J., & Beckman, J. (Eds.), *Action Control* (pp. 11-39). Heidelberg: Springer-Verlag.

Akerlind, G. S. (2005). Postdoctoral researchers: Roles, functions and career prospects. *Higher Education & Research Development 24*(1): 21-40.

Alauddin, M., & Nghiemb, H. S. (2010). Do instructional attributes pose multicollinearity problems?. An empirical exploration. *Economic Analysis & Policy 40*(3): 351-361.

Albert, T. & Wager, E. (2004). How to handle authorship disputes: A guide for new researchers. In White, C. (Ed.), *The COPE Report 2003* (pp. 32-34). London: BMJ Books.

Alexy, O., J. West, Klapper, H., & Reitzig. M. (2018). Surrendering control to gain advantage: Reconciling openness and the resource-based view of the firm. *Strategic Management Journal 39*(6): 1704-1727.

Ali, A. M. E. T., Abdulla, H. M. D., & Snasel, V. (2011). *Survey of Plagiarism Detection Methods.* Paper presented at the Fifth Asia Modelling Symposium, Kuala Lumpur.

Altbach, P. G. (2015). What counts for academic productivity in research universities?. *International Higher Education 79*(4): 6-7.

American Economic Review (2020). *AER: Submission Guidelines.* American Economic Association. Retrieved from https://www.aeaweb.org/journals/aer/submissions/guidelines [7th of September, 2020].

American Institutes for Research. (2016). *Project TALENT.* https://www.projecttalent.org/ [10th of November, 2019].

American Psychological Association. (1983). *Authorship guidelines for dissertation supervision.* Washington, DC: American Psychological Association.

American Psychological Association. (2018). Summary report of journal operations, 2017. *American Psychologist 73*(5): 683-684.

American Sociological Association. (2018). *Code of ethics*. Retrieved from http://www.asanet.org/sites/default/files/asa_code_of_ethics-june2018.pdf. [18th of October, 2018].

American Sociological Association. 2020. *Annual Editor's Reports.* Retrieved from https://www.asanet.org/research-and-publications/journal-resources/annual-editors-reports#ST [10th of August, 2020].

Amrhein, V., Korner-Nievergelt, F., & Roth, T. (2017). The earth is flat (p > 0.05): Significance thresholds and the crisis of unreplicable research. *PeerJ 5*: e3544.

Anand, J., & Delios, A. (1997). Location specificity and the transferability of downstream assets to foreign subsidiaries. *Journal of International Business Studies 28*(3): 579-603.

Andersen, M. L., & Lu, F. (2016). Learning to Manage and Managing to Learn: The Effects of Student Leadership Service. *Management Science 63*(10): 3246-3261.

Anderson, J. C., Funk, J. B., Elliott, R., & Smith, P. H. (2003). Parental support and pressure and children's extracurricular activities: Relationships with amount of involvement and affective experience of participation. *Journal of Applied Developmental Psychology 24*(2): 241-257.

Andreoli-Versbach, P., & Mueller-Langer, F. (2014). Open access to data: An ideal professed but not practised. *Research Policy 43*(9): 1621-1633.

Andrews, I., Stock. J. H., & Sun, L. (2019). Weak instruments in instrumental variables regression: Theory and practice. *Annual Reviews of Economics 11*(1): 727-753.

Andrews, J., & Higson, H. (2008). Graduate employability, 'soft skills' versus 'hard' business knowledge: A European study. *Higher Education in Europe 33*(4): 411-422.

Antes, A. L., Walsh, H. A., Strait, M., Hudson-Vitale, C. R., & DuBois, J. M. (2018). Examining data repository guidelines for qualitative data sharing. *Journal of Empirical Research on Human Research Ethics 13*(1): 61-73.

Antonakis, J., Bastardoz, N., Jacquart, P., & Shamir, B. (2016). Charisma: An ill-defined and ill-measured gift. *Annual Review of Organizational Psychology and Organizational Behavior 3*: 293-319.

Antonakis, J., Bendahan, S., Jacquart, P. & Lalive, R. (2014). Causality and endogeneity: problems and solutions. In Day, D. V. (Ed.): *The Oxford Handbook of Leadership and Organizations* (pp. 93–117). Oxford: Oxford University Press.

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *The Leadership Quarterly 21*(6): 1086-1120.

Arce, D. G., Enders, W., & Hoover, G. A. (2008). Plagiarism and its impact on the economics profession. *Bulletin of Economic Research 60*(3): 231-243.

Arnowitt, R., Deser, S., & Misner, C. W. (1962) The dynamics of general relativity. In Witten, L. (Ed.), *Gravitation: An Introduction to Current Research.* New York: Wiley.

Arya, B., & Linz, Z. (2007). Understanding collaboration outcomes from an extended resource-based view perspective: The roles of organizational characteristics, partner attributes, and network structures. *Journal of Management 33*(5): 697-723.

AsPredicted. (2020). *AsPredicted.* AsPredicted. Retrieved from https://aspredicted.org/ [5th of September, 2020].

Aubert, B. A., Rivard, S., & Patry, M. (2004). A transaction cost model of IT outsourcing. *Information & Management 41*(7): 921-932.

Austin, J. T., & Hanisch, K. A. (1990). Occupational attainment as a function of abilities and interests: A longitudinal analysis using Project TALENT data. *Journal of Applied Psychology 75*(1): 77.

Avolio, B. J., Rotundo, M., & Walumbwa, F. O. (2009). RETRACTED: Early life experiences as determinants of leadership role occupancy: The importance of parental influence and rule breaking behavior. *The Leadership Quarterly 20*(3): 329-342.

Azoulay, P. J., Bonatti, A., & Krieger, J. L. (2017). The career effects of scandal: Evidence from scientific retractions. *Research Policy 46*(9): 1552-1569.

Baas, J., & Fennel, C. (2019). *When Peer Reviews Go Rogue – Estimated Prevalence of Citation Manipulation by Reviewers Based on the Citation Patterns of 69,000 Reviewers.* Paper presented at the International Society Meeting 2019, Rome.

Bacolod, M., Blum, B. S., & Strange, W. C. (2010). Elements of skill: Traits, intelligences, education, and agglomeration. *Journal of Regional Science 50*(1): 245-280.

Backes-Gellner, U., Mohnen, A., & Werner, A. (2004). *Team Size and Effort in Start-Up Teams – Another Consequence of Free-Riding and Peer Pressure in Partnerships.* SSRN. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=518443 [6th of November, 2020].

Balkin, D. B., Trevino, L., Fitza, M. A., Gomez-Mejia, L. R., & Tadikonda, D. D. H. (2018). *Who's On First? Uncovering the Factors that Lead to First Authorship in Management Scholarship.* Paper presented at the Academy of Management 2018, Chicago.

Bandura, A. (1977). *Social Learning Theory.* Englewood Cliffs: Prentice Hall.

Banks, G. C., O'Boyle, E. H., Pollack. J., White, C. D., Batchelor, J. H., Whelpley, C. E. et al. (2016). Questions about questionable research practices in the field of management: A guest commentary. *Journal of Management 42*(1): 5-20.

Barber, G. (2019). *Artificial Intelligence Confronts a 'Reproducibility' Crisis.* Wired. Retrieved from https://www.wired.com/story/artificial-intelligence-confronts-reproducibility-crisis/ [27th of July, 2020].

Barbour, V., Bloom, T., Lin, J., & Moylan, E. (2017). Amending published articles: Time to rethink retractions and corrections?. *F1000Research 6*(1): 1960.

Barney, J. B. (2001). Resource-based theories of competitive advantage: A ten-year retrospective on the resource-based view. *Journal of Management 27*(6): 643-650.

Barron, J. M., Bradley, T. E. & Wadden, G. R. (2000). The effects of high school athletic participation on education and labor market outcomes. *The Review of Economics and Statistics 82*(3): 409-421.

Barry, E., & Bannister, F. (2014). Barriers to open data release: A view from the top. *Information Polity 19*(12): 129-152.

Bartel, A., Ichniowski, C., & Shaw, K. (2007). How does information technology affect productivity? Plant-level comparisons of product innovation, process improvement, and worker skills. *The Quarterly Journal of Economics 122*(4): 1721-1758.

Bartle, S. A., Fink, A. A., & Hayes, B. C. (2000). Psychology of the scientist: LXXX. Attitudes regarding authorship issues in psychological publications. *Psychological Report 86*: 771-788.

Bascle, G. (2008). Controlling for endogeneity with instrumental variables in strategic management research. *Strategic Organization 6*(3): 285-327.

Bastardoz, N., & Van Vugt, M. (2019). The nature of followership: Evolutionary analysis and review. *The Leadership Quarterly 30*(1): 81-95.

Bates, T., Anic, A., Marusic, M., & Marusic, A. (2004). Authorship criteria and disclosure of contributors – Comparison of 3 general medical journals with different author contribution forms. *JAMA 292*(1): 86-88.

Baum, C. F., & Lewbel, A. (2019). Advice on using heteroskedasticity-based identification. *The Stata Journal 19*(4): 757-767.

Beall, J. (2013). Scholarly publishing free for all. *College Quarterly 16*(2): 8.

Becerra, M., Lunnan, R., & Huember, L. (2008). Trustworthiness, risk, and the transfer of tacit and explicit knowledge between alliance partners. *Journal of Management Studies 45*(4): 691-713.

Beck, S., Bergenholtz, C., Bogers, M., Brasseur, T. M., Conradsen, M. L., Di Marco, D. et al. (2020). The open innovation in science research field: A collaborative conceptualization approach. *Industry and Innovation* Published online first 4th of August, 2020.

Becker, B., Cronqvist, H., & Fahlenbrach, R. (2011). Estimating the effects of large shareholders using a geographic instrument. *Journal of Financial and Quantitative Analysis 46*(4): 907-942.

Becker, C. (2007). *Giordano Bruno – Die Spuren des Ketzers. Ein Beitrag zur Literatur-, Wissenschafts- und Gelehrtengeschichte um 1600.* Stuttgart: ibidem-Verlag.

Becker, S. O., & Ichino, A. (2002). Estimation of average treatment effects based on propensity scores. *The Stata Journal 2*(4): 358-377.

Bekkers, H. (2012). *Risk Factors for Fraud and Academic Misconduct in the Social Sciences.* Rene Bekkers. Retrieved from https://renebekkers.wordpress.com/2012/11/29/risk-factors-for-fraud-and-academic-misconduct-in-the-social-sciences/ [5th of September, 2020].

Bendavid, E. (2020). *Eran Bendavid.* Stanford University. Retrieved from https://www.sfgate.com/news/editorspicks/article/Santa-Clara-antibody-study-revised-Stanford-death-15263047.php [7th of September, 2020].

Bendavid, E., Mulaney, B., Sood, N., Shah, S., Ling, E., Bromley-Dulfano, R., et al. (2020). *COVID-19 Antibody Seroprevalence in Santa Clara County, California.* medRxiv. Retrieved from https://www.medrxiv.org/content/10.1101/2020.04.14.20062463v2 [7th of September, 2020].

Bennet, A., & Derevensky, J. (1995). The medieval kingdom topology: Peer relations in kindergarten children. *Psychology in Schools 32*(2): 130-141.

Bennett, D. M., & Taylor, D. M. (2003). Unethical practices in authorship of scientific papers. *Emergency Medicine 15*(3): 263-270.

Bergh, D. D., Sharp, B. M., Aguinis, H., & Li, M. (2017). Is there a credibility crisis in strategic management research? Evidence on the reproducibility of study findings. *Strategic Organization 15*(3): 423-436.

Bergman, R. G., & Danheiser, R. L. (2016). Reproducibility in chemical research. *Angewandte Chemie International Edition 55*(41): 12548-12549.

Bernardi, R. A. (1994). Validating research results when Cronbach's alpha is below .70: A methodological procedure. *Educational and Psychological Measurement 54*(3): 766-775.

Bettis, R. A., Helfat, C. E., & Shaver, J. M. (2016). The necessity, logic, and forms of replication. *Strategic Management Journal 37*(11): 2193-2203.

Bhattacharjee, Y. (2013). *The Mind of a Con Man.* New York Times. Retrieved from https://www.nytimes.com/2013/04/28/magazine/diederik-stapels-audacious-academic-fraud.html [3rd of September, 2020].

Bhattacherjee, A. (2012). *Social Science Research: Principles, Methods and Practices* (2nd ed.). Tampa: University of South Florida Textbooks Collection.

Björk, B. C., & Solomon, D. (2015). Article processing charges in OA journals: Relationships between price and quality. *Scientometrics 103*(2): 373-385.

Blank, R. M. (1991). The effects of double-blind versus single-blind reviewing: Experimental Evidence from the American Economic Review. *American Economic Review 81*(5): 1041-1067.

Bloom, N. & van Reenen, J. (2007). Measuring and explaining management practices across firms and countries. *The Quarterly Journal of Economics 122*(4): 1351-1408.

Bogers, M. (2011a). Knowledge sharing in open innovation: An overview of theoretical perspectives on collaborative innovation. In de Pablos Heredero, C., & Lopez, D. (Eds.), *Open Innovation in Firms and Public Administrations* (pp. 1-14). Hershey: Information Science Reference.

Bogers, M. (2011b). The open innovation paradox: knowledge sharing and protection in R&D collaborations. *European Journal of Innovation Management 14*(1): 93-117.

Bogers, M., Zobel, A. K., Afuah, A., Almirall, E., Brunswicker, S., Dahlander, L. et al. (2017). The open innovation research landscape: Established perspectives and emerging themes across different levels of analysis. *Industry and Innovation 24*(1): 8-40.

Borghans, L., Ter Weel, B., & Weinberg, B. A. (2014). People skills and the labor-market outcomes of underrepresented groups. *ILR Review 67*(2): 287-334.

Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology 63*(6): 1059-1078.

Borrego, A., Barrios, M., Villarroya, A., Frias, A., & Olle, C. (2008). *Research output of Spanish postdoctoral scientists: Does gender matter?.* Paper presented at the Fourth International Conference on Webometrics, Informetrics and Scientometrics & Ninth COLLNET Meeting, Berlin.

Borsuk, R. M., Budden, A. E., Leimu, R., Aarssen, L. W., & Lortie, C. J. (2009). The influence of author gender, national language and number of authors on citation rate in ecology. *The Open Ecology Journal 2*(1): 25-28.

Bosch, X., & Ross, J. S. (2012). Ghostwriting: Research misconduct, plagiarism or fool's gold?. *The American Journal of Medicine 125*(4): 324-326.

Bosco, F. A., Aguinis, H., Field, J. G., Pierce, C. A., & Dalton, D. R. (2016). HARKing's threat to organizational research: Evidence from primary and meta-analytic sources. *Personnel Psychology 69*(3): 709-750.

Botner, O., Danielsson, U., Ingelman, g., Rose, J., Barnes, C., & Jarnestad, J. (2017). *Cosmic Chips.* Stockholm: The Royal Swedish Academy of Sciences.

Boudreau, K. J., & Lakhani, K. R. (2015). 'Open' disclosure of innovations, incentives and follow-on reuse: Theory on processes of cumulative innovation and a field experiment in computational biology. *Research Policy 44*(1): 4-19.

Bowler, M. (1999). Women's earnings: An overview. *Monthly Labor Review 122*(12): 13-21.

Brennan, T. A. (1994). Buying editorials. *The New England Journal of Medicine 331*(10): 673-675.

Bretad, T., & Mahmud, S. (2009). Self-plagiarism of appropriate textual re-use?. *Journal of Academic Ethics 7*(1): 193-205.

Breuer, K., & Kampkötter, P. (2013). Determinants and effects of intra-firm trainings: evidence from a large German company. *Journal of Business Economics 83*(2): 145-169.

Broad, W. J. (1981). The publishing game: Getting more for less. *Science 211*(4487): 1137-1139.

Brodeur, A., Le, M., Sangnier, M., & Zylberberg, Y. (2016). Star wars: The empirics strike back. *American Economic Journal: Applied Economics 8*(1): 1-32.

Brooks, M. (2018). *Exclusive: Grave Doubts over LIGO's Discovery of Gravitational Waves.* New Scientists. Retrieved from https://www.newscientist.com/article/mg24032022-600-exclusive-grave-doubts-over-ligos-discovery-of-gravitational-waves/ [23rd of August, 2020].

Brown, C. (1984): Black-white earnings ratio since the Civil Rights Act of 1964: The importance of labor market dropouts. *The Quarterly Journal of Economics 99*(1): 31-44.

Brüggen, A., Maas, V., & Van den Abbeele, A. (2020). *Special Issue: Experiments, Replication and the Advancement of Management Accounting and Control Research.* Journal of Management Control. Retrieved from https://www.springer.com/journal/187/updates/17197186 [10th of September, 2020].

Burlig, F. (2018). Improving transparency in observational social science research: A pre-analysis plan approach. *Economic Letters 168*(1): 56-60.

Burman, L. E., Reed, W. R., & Alm, J. A call for replication studies. *Public Finance Review 38*(6): 787-793.

Bush, G. P., & Hattery, L. H. (1956). Teamwork and creativity in research. *Administrative Science Quarterly 1*(3): 361-372.

Byington, E. K., & Felps, W. (2017). Solutions to the credibility crisis in management science. *Academy of Management Learning & Education 16*(1): 142-162.

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys 22*(1): 31-72.

Callaham, M., Wears, R. L., & Weber, E. (2002). Journal prestige, publication bias, and other characteristics associated with citation of published studies in peer-reviewed journals. *JAMA 287*(21): 2847-2850.

Callahan, J. L. (2018). The retrospective (im)moralization of self-plagiarism: Power interests in the social construction of new norms for publishing. *Organization 25*(3): 305-319.

Camerer, C. F., Dreber, A., Forsell, E., Ho, T. H., Huber, J., Johannesson, M. et al. (2016). Evaluating replicability of laboratory experiments in economics. *Science 351*(6280): 1433-1436.

Campbell, D. L., & Mau, K. (2019). *Trade Induced Technological Change: Did Chinese Competition Really Increase European Innovation?.* Working Paper No. 262. Moscow: New Economic School.

Carayol, N. (2007). Academic incentives and research organization at a large French university. *Economics of Innovation and New Technology 16*(2): 119-138.

Carayol, N., & Matt, M. (2006). Individual and collective determinants of academic scientists' productivity. *Information Economics and Policy 18*(1): 55-72.

Carpenter, C. R., Cone, D. C., & Sarli, C. C. (2014). Using publication metrics to highlight academic productivity and research impact. *Academic Emergency Medicine 21*(10): 1160-1172.

Cassidy, L. (2013). *The Reinhart and Rogoff Controversy: A Summing Up.* The New Yorker. Retrieved from https://www.newyorker.com/news/john-cassidy/the-reinhart-and-rogoff-controversy-a-summing-up [17th of October, 2019].

Castelvecchi, D., & Witze, A. (2016). *Einstein's Gravitational Waves Found at last.* Nature. Retrieved from: https://www.nature.com/news/einstein-s-gravitational-waves-found-at-last-1.19361 [20th of August, 2020].

Catalini, C., Fons-Rosen, C., & Gaule, P. (2016). *Did Cheaper Flights Change the Direction of Science?.* Discussion Paper 9897. Bonn: Institute for the Study of Labor (IZA).

Chan, K. C., Chen, C. R., & Cheng, L. T. W. (2009). Co-authorship patterns in accounting research. *International Review of Accounting, Banking and Finance 1*(2): 1-10.

Chao, C. C., & Tian, D. (2011). Culturally universal or culturally specific: A comparative study of anticipated female leadership styles in Taiwan and the United States. *Journal of Leadership & Organizational Studies 18*(1): 64-79.

Chase, J. M. (1970). Normative criteria for scientific publication. *The American Sociologist 5*(3): 262-265.

Cherkashin, I., Demidova, S., Imai, S., & Krishna, K. (2009). The inside scoop: Acceptance and rejection at the Journal of International Economics. *Journal of International Economics 77*(1): 120-132.

Chesbrough, H. (2003). *Open Innovation: The New Imperative for Creating and Profiting from Technology.* Cambridge, MA: Harvard Business School Publishing.

Cho, A. (2017). Trio surfs gravitational waves to Nobel glory. *Science 358*(6359): 17.

Chrousos, G. P., Kalantaridou, S. N., Margioris, A. N., & Gravanis, A. (2012). The 'self-plagiarism' oxymoron : Can one steal from oneself ?. *Europen Journal of Clinical Investigation 42*(3): 231-232.

Cicero, M. T. (45 BC). *Hortensius. Lucullus. Academici Libri.* Retrieved from https://www.degruyter.com/view/title/496414 [18th of August, 2020].

Cicero, M. T., & Rackham, H. (1933). *De Natura Deorum; Academica; With an English Translation by H. Rackman.* Retrieved from https://archive.org/details/de-naturadeorumac00ciceuoft/page/424/mode/2up [18th of August, 2020].

Clapp-Smith, R., Carsten, M., Gooty, J., Connelly, S., Haslam, A., Bastardoz, N., & Spain, S. (2018). Special registered report issue on replication and rigorous retesting of leadership models. *The Leadership Quarterly 29*(2): V-VI.

Clapp-Smith, R., Carsten, M., Gooty, J., Connelly, S., Haslam, A., Bastardoz N., & Spain, S. (2017). Replication and rigorous retesting of leadership models. *The Leadership Quarterly 28*(5): 6-8.

Claxton, L. D. (2005). Scientific Authorship – Part 2. History, Recurring Issues, Practices, and Guidelines. *Mutation Research 589*(1): 31-45.

Coase, R. H. (1937). The nature of the firm. *Economica 4*(16): 386-405.

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist 49*(12): 997-1003.

Cohen, P. N., & Huffman, M. L. (2007). Working for the woman? Female managers and the gender wage gap. *American Sociological Review 72*(5): 681-704.

Cole, J., & Cole, S. (1973). *Social Stratification in Science.* Chicago: The University of Chicago Press.

Colman, A. M. (1995). Priosners' dilemma, chicken, and mixedstrategy evolutionary equilibria. *An International Journal of Current Research and Theory with Open Peer Commentary 18*(3): 550-551.

Colussi, T. (2018). Social ties in academia: A friend is a true treasure. *Review of Economics and Statistics 100*(1): 45-50.

Corley, E. A. (2005). How do career strategies, gender, and work environment affect faculty productivity levels in university-based science centers?. *Review of Policy Research 22*(5): 637-655.

Costa, M. M., & Gatz, M. (1992). Determination of authorship credit in published dissertations. *Psychological Science 3*(6): 354-357.

Cousineau, D. (2014). Restoring confidence in psychological science findings: A call for direct replication studies. *Quantitative Methods for Psychology 10*(2): 77-79.

Cox, A., Craig, R., & Tourish, D. (2018). Retraction statements and research malpractice in economics. *Research Policy 47*(5): 924-935.

Creswell, J., von Hausegger, S., Jackson, A. D., Liu, H., & Naselsky, P. (2017). On the time lags of the LIGO signals. *Journal of Cosmology and Astroparticle Physics 8*(13): 1-28.

Cumming, G. (2008). Replication an p intervals: p values predict the future only vaguely, but confidence intervals do much better. *Perspectives on Psychological Science 3*(4): 286-300.

Cummings, J. N., & Kiesler, S. (2007). Coordination costs and project outcomes in multi-university collaborations. *Research Policy 36*(10): 1620-1634.

Cummings, J. N., & Kiesler, S. (2008). *Who Collaborates Successfully? Prior Experience Reduces Collaboration Barriers in Distributed Interdisciplinary Research*. Paper presented at the Association for Computing Machinery conference on computer supported cooperative work, San Diego.

Cunha, M. P., Lewis, M., Rego, A., & Smith, W. K. (2017). 15. Biographical methods in leadership research. In Schyns, B., Hall, R. J. & Neves, P. (Eds.), *Handbook of Methods in Leadership Research* (pp. 372-402). Cheltenham: Edward Elgar Publishing.

da Silva, J. A. T., & Al-Khatib, A. (2019). The Clarivate Analytics acquisition of Publons – An evolution or commodification of peer review?. *Research Ethics 15*(3-4): 1-11.

da Silva, J. A. T., & Dobranszki, J. (2015). Potential dangers with open access data files in the expanding open data movement. *Publishing Research Quarterly 31*(4): 298-305.

da Silva, J. A. T., & Dobranszki, J. (2016). Multiple authorship in scientific manuscripts: Ethical challenges, ghost and guest/gift Authorship, and the cultural/disciplinary perspective. *Science and Engineering Ethics 22*(5): 1457-1472.

Dale, S. B. & Krueger, A. B. (2002). Estimating payoff to attending a more selective college: An application of selection on observables and unobservables. *Quarterly Journal of Economics 117*(4): 1491-1527.

Das, T. K., & Teng, B. S. (2000). A resource-based theory of strategic alliances. *Journal of Management 26*(1): 31-61.

David, P.A. (1995). *Reputation and Agency in the Historical Emergence of the Institutions of 'Open Science'.* Paper presented at the National Academy of Sciences Conference on Science, Technology and the Economy 1995, Irvine.

Davidoff, F. (2000). News from the International Committee of Medical Journal Editors. *Annals of Internal Medicine 133*(3): 229-231.

Davidson, R., & MacKinnon, J. G. (1993). *Estimation and Inference in Econometrics.* Oxford: Oxford University Press.

Day, V. D., Fleenor, J. W., Atwater, L. E., Sturm, R. E. & McKee, R. A. (2014). Advances in leader and leadership development: A review of 25 years of research and theory. *The Leadership Quarterly 25*(1): 63-82.

De Keersmaecker, J., & Roets, A. (2017). 'Fake news': Incorrect, but hard to correct. The role of cognitive ability on the impact of false information on social impressions. *Intelligence 65*(1): 107-110.

Deb, P. (2009). *MTREATREG: Stata Module to Fits Models with Multinomial Treatments and Continuous, Count and Binary Outcomes using Maximum Simulated Likelihood.* EconPapers. Retrieved from https://econpapers.repec.org/software/bocbocode/s457064.htm [14th of November, 2019].

DeCoster, J. (1998). *Overview of Factor Analysis.* Stat-Help. Retrieved from http://www.stat-help.com/factor.pdf [26th of September, 2019].

Defazio, D., Kolympiris, C., Salter, A., & Perkman, M. (2020). Busy academics share less: The impact of professional and family roles on academic withholding behaviour. *Studies in Higher Education* Published online first 17th of July, 2020.

Degen, R. (2015). *Diederik Stapel now has 58 Retractions.* Retraction Watch. Retrieved from https://retractionwatch.com/2015/12/08/diederik-stapel-now-has-58-retractions/ [3rd of September, 2020].

Dehejia, R. H., & Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and Statistics 84*(1): 151–161.

Deming, D. J. (2017). The growing importance of social skills in the labor market. *The Quarterly Journal of Economics 132*(4): 1593-1640.

Di Bitetti, M. S., & Ferreras, J. A. (2017). Publish (in English) or perish: The effect on citation rate of using languages other than English in scientific publications. *Ambio 46*(1): 121-127.

Dilger, A., Lütkenhöner, L., & Müller, H. (2015). Scholars‘ physical appearance, research performance, and feelings of happiness. *Scientometrics 104*(1): 555-583.

DiPrete, T. A., & Gangl, M. (2004). 7. Assessing bias in the estimation of causal effects: Rosenbaum bounds on matching estimators and instrumental variables estimation with imperfect instruments. *Sociological Methodology 34*(1): 271-310.

Dirnagl, U. (2020). *Der Fall Ioannidis.* LaborJournal. Retrieved from https://www.laborjournal.de/editorials/2009.php [7th of September, 2020].

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carre, G. et al. (2010). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography 36*(1): 27-46.

Dormody, T. J., & Seevers, B. S. (1994). Predicting youth leadership life skills development among FFA members in Arizona, Colorado, and New Mexico. *Journal of Agricultural Education 35*(2): 65-71.

Dougherty, T. W., Dreher, G. F., Arunachalam, V., & Wilbanks, J. E. (2013). Mentor status, occupational context, and protégé career outcomes: Differential returns for males and females. *Journal of Vocational Behavior 83*(3): 514-527.

Dowden, K. (2006). *Zeus.* Oxon: Routledge.

Drenth, J. P. H. (1998). Multiple authorship – The contribution of senior authors. *JAMA 280*(3): 219-221.

Duch, J., Zeng, X. H. T., Sales-Pardo, M., Radicchi, F., Otis, S., Woodruff, T. K. et al. (2012). The possible role of resource requirements and academic career-choice risk on gender differences in publication rate and impact. *PLoS ONE 7*(12): e51332.

Ductor, L. (2015). Does co-authorship lead to higher academic productivity?. *Oxford Bulletin of Economics and Statistics 77*(3): 385-407.

Dutton, W. H., & P. W. Jeffreys (Eds.). (2010). *World Wide Research: Reshaping the Sciences and Humanities.* Cambridge, MA: The MIT Press.

Duvendack, M., Palmer-Jones, R. W., & Reed, W. R. (2015). Replications in economics: A progress report. *Econ Journal Watch,* 12(2): 164-191.

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review 109*(3): 573-598.

Earp, B. d., & Trafimow, D. (2015). Replication, falsification, and the crisis of confidence in social psychology. *Frontiers in Psychology 6*(621): 1-11.

Ebersole, C. R., Atherton, O. E., Belanger, A. L., Skulborstad, H. M., Allen, J. M., Banks, J. B., ... & Brown, E. R. (2016). Many Labs 3: Evaluating participant pool quality across the academic semester via replication. *Journal of Experimental Social Psychology 67*(1): 68-82.

Edwards, P. N., Mayernik, M. S., Batcheller, A. L., Bowker, G. C., & Borgman, C. L. (2011). Science friction: Data, metadata, and collaboration. *Social Studies of Science 41*(5): 667-690.

EF Education First GmbH. (2019). *EF English Proficiency Index – Europe.* Retrieved from https://www.ef.com/epi/regions/europe/ [17th of December, 2019].

Eide, E. R., & Ronan, N. (2001). Is participation in high school athletics an investment or a consumption good? Evidence from high school and beyond. *Economics of Education Review 20*(5): 431-442.

Eime, R. M., Harvey, J. T., Craike, M. J., Symons, C. M., & Payne, W. R. (2013). Family support and ease of access link socio-economic status and sports club membership in adolescent girls: A mediation study. *International Journal of Behavioral Nutrition and Physical Activity 10*(1): 50.

Einstein, A. (1916). Näherungsweise Integration der Feldgleichungen der Gravitation. *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften 17*: 99-108.

Emrouznejad, A., Parker, B. R., & Tavares, G. (2008). Evaluation of research in efficiency and productivity: A survey and analysis of the first 30 years of scholarly literature in DEA. *Socio-Economic Planning Science 42*(3): 151-157.

Enders, W., & Hoover, G. A. (2006). Plagiarism in the economics profession: A survey. *Challenge 49*(5): 92-107.

Endersby, J. W. (1996). Collaborative research in the social sciences: Multiple authorship and publication credit. *Social Science Quarterly 77*(2): 375-392.

Engels, T., Ossenblok, T., & Spruyt, E. (2012). Changing publication patterns in the social sciences and humanities. *Scientometrics 93*(2): 373-390.

Engers, M., Gans, J. S., Grant, S., & King, S. P. (1999). First-author conditions. *Journal of Political Economy 107*(4): 859-883.

Enserink, (2012). *Final Report: Stapel Affair Points to Bigger Problems in Social Psychology.* Science. Retrieved from https://www.sciencemag.org/news/2012/11/final-report-stapel-affair-points-bigger-problems-social-psychology [3rd of September, 2020].

Esarey, J. (2016). *Acceptance Rates and the Aesthetics of Peer Review.* The Political Methodologist. Retrieved from https://thepoliticalmethodologist.com/2016/01/13/acceptance-rates-and-the-aesthetics-of-peer-review/ [18th of April, 2020].

Ethnologue. (2020). *What are the Top 200 Most Spoken Languages?* Retrieved from https://www.ethnologue.com/guides/ethnologue200 [15th of July, 2020].

Experimental Psychology. (2020). *Experimental Psychology.* Hogrefe. Retrieved from https://www.hogrefe.com/us/journal/experimental-psychology [10th of September, 2020].

Fanelli, D. (2012). Negative results are disappeared from most disciplines and countries. *Scientometrics 90*(3): 891-904.

Fecher, B., Firesike, S., Hebing, M., Linek, S., & Sauermann, A. (2015). *A Reputation Economy: Results from an Empirical Survey on Academic Data Sharing*. Discussion Paper No. 1454. Berlin: DIW.

Fecher, B., Friesike, S., Hebing, M., & Linek, S. (2017). A reputation economy: How individual reward considerations trump systemic arguments for open access to data. *Palgrave Communications 3*(17051): 1-10.

Federer, L. M., Belter, C. W., Joubert, D. J., Livinski, A., Lu, Y. L., Snyders, L. N., & Thompson, H. (2018). Data sharing in PLoS ONE: An analysis of data availability statements. *PLoS ONE 13*(5): e0194768.

Feeser, V. R., & Simon, J. R. (2008). The ethical assignment of authorship in scientific publications: Issues and guidelines. *Academic Emergency Medicine 15*(10): 963-969.

Figa-Talamanca, A. (2007). Are citations the currency of science?. *Journal of European Psychoanalysis,* 24(1).

Fine, G. (2008). Does Socrates claim to know that he knows nothing?. *Oxford Studies in Ancient Philosophy 35*(4): 49-88.

Firth, D., King, J. L., Koch, H., Looney, C. A., Pavlou, P. A., & Trauth, E. (2014). Addressing the credibility crisis in IS. *Communications of the Association for Information Systems 28*(1): 199-121.

Fischer, B. A., & Zigmond, M. J. (2010). The essential nature of sharing in science. *Science and Engineering Ethics 16*(4): 783-799.

Fisher, R. A. (1925). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society A 222*(1): 594-604.

Fitzsimmons, T. W., Callan, V. J., & Paulsen, N. (2014). Gender disparity in the C-suite: Do male and female CEOs differ in how they reached the top?. *The Leadership Quarterly 25*(2): 245-266.

Fleischman, R. K., & Schuele, K. (2009). Co-authorship in accounting history: Advantages and pitfalls. *Accounting, Business & Financial History 19*(3): 287-303.

Floyd, S. W., Schroeder, D. M., & Finn, D. M. (1994). "Only if I'm first author.": Conflict over credit in management scholarship. *Academy of Management Journal 37*(3): 734-747.

Fochler, M., & Sigl, L. (2018). Anticipatory uncertainty: How academic and industry researchers in the life sciences experience and manage the uncertainties of the research process differently. *Science as Culture 27*(3): 349-374.

Foster E. M., Porter, M. M., Ayers, T. S., Kaplan, D. L., & Sandler, I. (2007). Estimating the costs of preventive interventions. *Evaluation Review 31*(3): 261-286.

Frederiksen, A., & Kato, T. (2017). Human capital and career success: Evidence from linked employer-employee data. *The Economic Journal 128*(613): 1952-1982.

Freese, J. (2007). Replication standards for quantitative social science: Why not sociology?. *Sociological Methods & Research 36*(2): 153-172.

Friesike, S., Widenmayer, B., Gassmann, O., & Schildhauer, T. (2015). Opening science: Towards an agenda of open science in academia and industry. *The Journal of Technology Transfer 40*(4): 581-601.

Gall, T., Ioannidis, J. P. A., & Maniadis, Z. (2017). The credibility crisis in research: Can economics tools help? *PLoS Biology 15*(4): e2001846.

Gersik, C. J. G., Bartunek, J. M., & Dutton, J. E. (2000). Learning from academia: The importance of relationships in professional life. *Academy of Management Journal 43*(6): 1026-1044.

Gerwin, V. (2016). An open mind on open data. *Nature 529*(7584): 117-119.

Gigerenzer, G., & Murray, D. (1987). *Cognition as Intuitive Statistics.* Hillsdale: Erlbaum.

Glaenzel, W., & Schubert, A. (2001). Double effort = double impact? A critical view at international co-authorship in chemistry. *Scientometrics 50*(2): 199-214.

Glänzel, W. (2001). National characteristics in international scientific co-authorship relations. *Scientometrics 51*(1): 69-115.

Golden-Biddle, K., Locke, K., & Reay, T. (2006). Using knowledge in management studies: An investigation of how we cite prior work. *Journal of Management Inquiry 15*(3): 237-254.

Gonzales, J. E., & Cunningham, C. A. (2015). *The Promise of Pre-Registration in Psychological Research.* Psychological Science Agenda. Retrieved from https://osf.io/7pd5u/download [5th of September, 2020].

Graham, D. J., & Melo, P. C. (2011). Assessment of wider economic impacts of high-speed rail for Great Britain. *Transportation Research Record 2261*(1): 15-24.

Greenland, P., & Fontanarosa, P. B. (2012). Ending Honorary Authorship. *Science 337*(6098): 1019.

Grunzke, R., Adolph, T., Biardzki, C., Bode, A., Borst, T. Bugnartz, H. J. et al. (2017). Challenges in creating a sustainable generic research data infrastructure. *Softwaretechnik-Trends 37*(2): 74-77.

Guiteras, R. P., & Jack. B. K. (2018). Productivitiy in piece-rate labor markets: Evidence from rural Malawi. *Journal of Development Economics 131*(3): 42-61.

Guo, S., & Fraser, M. W. (2014). *Propensity Score analysis: Statistical Methods and Applications*. London: Sage Publications.

Haensly, P. J., Hodges, P. E., & Davenport, S. A. (2008). Acceptance rates and journal quality: An analysis of journals in economics and finance. *Journal of Business & Finance Librarianship 14*(1): 2-31.

Hahn, S. G., & Lindquist, R. W. (1964). The two-body problem in geometrodynamics. *Annals of Physics 29*(2): 304-331.

Hakim, C. (1987). *Research Design: Successful Designs for Social and Economic Research.* Oxon: Routledge.

Hallett, T. (2020). *Review 1: "The Infection Fatality Rate of COVID-19 Inferred from Sero-prevalence Data.* MIT Press. Retrieved from https://rapidreviewscovid19.mitpress.mit.edu/pub/6z09ucch/release/2 [7th of September, 2020].

Hambrick, D. C., & Wowak, A. J. (2012). Whom do we want as our business leaders? How changes in the corporate milieu have brought about a new breed of CEOs. In Ricart Costa, J. E. & Rosanas Marti, J. M. (Eds.), *Towards a New Theory of the Firm* (pp. 29-60). Bilbao: Fundacion BBVA.

Hamermesh, D. S. (2015). Age, cohort, and co-authorship. The statistics of collaboration. In Szenberg, M., & Ramrattan, L. B. (Eds.), *Collaborative Research in Economics: The Wisdom of Working Together* (pp. 65-93). Cham: Palgrave Macmillan.

Hamilton, B. H., & Nickerson, J. A. (2003). Correcting for endogeneity in strategic management research. *Strategic Organization 1*(1): 51-78.

Hamilton, J. B., Greco, A. J., & Tanner, J. R. (1997). Ethical questions regarding joint authorship: Business and nonbusiness faculty perceptions on noncontributing authorship. *Journal of Education for Business 72*(6): 325-330.

Harbring, C., & Irlenbusch, B. (2011). Sabotage in tournaments: Evidence from a laboratory experiment. *Management Science 57*(4): 611-627.

Harlow, L. L. (2016). Significance testing introduction and overview. In Harlow, L. L., Mulaik, S. A., & Steiger, J. H. (Eds.), *What if there were No Significance Tests?.* New York: Routledge.

Harper, A. M. (2018). *Experiences and Leadership Recommendations of African American CEOs at Fortune 500 Companies.* New York: Columbia University.

Harrison, J. S., Banks, G. C., Pollack, J. M., O'Boyle, E. H., & Short, J. (2017). Publication bias in strategic management research. *Journal of Management 43*(2): 400-425.

Hayes, S. (2008). *Navigating the Corporate Hierarchy: How Black Female Executives make Sense of their Experiences Ascending the Corporate Ladder and Sustaining Themselves in Senior-Level Positions*. New York: Columbia University.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica 47*(1): 153-161.

Heckman, J. J. (2001). Micro data, heterogeneity, and the evaluation of public policy: Nobel lecture. *Journal of Political Economy 109*(4): 673-748.

Heckman, J. J., & Singer, B. (2017). Abducting economics. *American Economic Review 107*(5): 298-302.

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.

Heeks, R. (2017). *Decent Work and the Digital Gig Economy: A Developing Country Perspective on Employment Impacts and Standards in Online Outsourcing, Crowdwork, Etc.* Development Informatics Working Paper no. 71. Manchester: University of Manchester.

Heidorn, P. B. (2011). The emerging role of libraries in data curation and e-science. *Journal of Library Administration 51*(7-8): 662-672.

Hekman, D. R., Johnson, S. K., Foo, M. D., & Yang, W. (2017). Does diversity-valuing behavior result in diminished performance ratings for non-white and female leaders?. *Academy of Management Journal 60*(2): 771-797.

Hendriks, F., Kienhues, D., & Bromme, R. (2016). Trust in science and the science of trust. In Blöbaum (Ed.), *Trust and Communication in a Digitized World*. Heidelberg: Springer.

Henriksen, D. (2016). The rise in co-authorship in the social sciences. *Scientometrics 107*(2): 455-476.

Henten, A., & Windekilde, I. (2015). *Transaction Costs and the Sharing Economy.* Paper presented at the 26th European Regional Conference of the International Telecommunications Society, Madrid.

Hermanns, P. (2014). *Urlich Lichtenthaler Retraction Count Rises to 16.* Retraction Watch. Retrieved from https://retractionwatch.com/2014/06/16/ulrich-lichtenthaler-retraction-count-rises-to-16/ [3rd of September, 2020].

Herndon, T., Ash, M., & Pollin, R. (2014). Does high public debt consistently stifle economic growth? A critique of Reinhart and Rogoff. *Cambridge Journal of Economics 38*(2): 257-279.

Hershey, D. A., Jacobs-Lawson, J. M., & Wilson, T. L. (2006). Research as a script. In Leong, F. T. L., & Austin, J. T. *The Psychology Research Handbook: A Guide for Graduate Students and Research Assistants* (2nd ed.). Thousand Oaks: Sage.

Heywood, J. S., Siebert, S., & Wei, X. (2013). *Unintended Consequences of a Piece Rate? Evidence from a Field Experiment.* Working Paper. Birmingham: University of Birmingham.

Hirschauer, S. (2004). Peer review research-reviewed: Sociological shortcomings of academic evaluation. *Zeitschrift für Soziologie 33*(1): 62-83.

Hitz, S. R. (1982). The impact of a computerized conferencing system on the productivity of scientific research communities. *Behaviour & Information Technology 1*(2): 185-195.

Hochreiter, R., & Waldhauser, C. (2014). Data mining cultural aspects of social media marketing. In Pernder, P. (Ed.), *Advances in Data Mining. Applications and Theoretical Aspects* (pp. 130-143). Cham, CH: Springer.

Hojat, M., Gonnella, J. S., & Caelleig, A. S. (2003). Impartial judgment by the "gatekeepers" of science: Fallibility and accountability in the peer review process. *Advances in Health and Science Education 8*(1): 75-96.

Hollenbeck, R. J., & Wright, P. M. (2017). Harking, sharking and tharking: Making the case for post hoc analysis of scientific data. *Journal of Management 43*(1): 5-18.

Holmstrom, B., & Milgrom, P. (1991). Multitask principal-agent analysis: Incentive contracts, asset ownership, and job design. *Economics & Organization 7*(1): 24-52.

Honig, B., & Bedi, A. (2012). The fox in the hen house: A critical examination of plagiarism among members of the Academy of Management. *Academy of Management Learning & Education 11*(1): 101-123.

Honig, B., Lampel, J., Siegel, D., & Drnevich, P. (2014). Ethics in the production and dissemination of management research: Institutional failure or individual fallibility?. *Journal of Management Studies 51*(1): 118–142.

Hopp, C., & Hoover, G. A. (2017). How prevalent is academic misconduct in management research?. *Journal of Business Research 80*(3): 73-81.

Hopp, C., & Hoover, G. A. (2019). What crisis? Management researchers' experiences with and views of scholarly misconduct. *Science and Engineering Ethics 25*(5): 1549-1588.

Hopp, C., & Speil, A. (2020). How prevalent is plagiarism among college students? Anonymity preserving evidence from Austrian undergraduates. *Accountability in Research* Published online first 11th of August, 2020.

Hopp, C., Greene, F. J., Honig, B., Karlsson, T., & Samuelsson, M. (2018). Revisiting the influence of institutional forces on the written business plan: a replication study. *Management Review Quarterly 68*(4): 361-398.

Horbach, S. P. J. M., & Halffmann, W. (2019). The extent and causes of academic text recycling of 'self-plagiarism'. *Research Policy 48*(2): 492-502.

Hossenfelder, S. (2017). *Was It All Just Noise? Independent Analysis Casts Doubt on LIGO's Detections.* Forbes. Retrieved from https://www.forbes.com/sites/startswithabang/2017/06/16/was-it-all-just-noise-independent-analysis-casts-doubt-on-ligos-detections/#3710fb345516 [23rd of August, 2020].

Hsiehchen, D., Espinoza, M., & Hsieh, A. 2015. Multinational teams and diseconomies of scale in collaborative research. *Science Advances 1*(8): e1500211.

Hudson, J. (1996). Trends in multi-authored papers in economics. *Journal of Economic Perspectives 10*(3): 153-158.

Hume, D. (1748). *An Enquiry Concerning Human Understanding.* London: A. Millar.

Hwang, S. S., Song, H. H., Baik, J. H., Jung, S. L., Park. S. E., Choi, K. H., & Park, Y. H. (2003) Researcher contributions and fulfillment of ICMJE authorship criteria: Analysis of author contribution lists in research articles with multiple authors published in *Radiology. Radiology 226*(1): 16-23.

Ioannidis, J. P. A. (2012). Why science is not necessarily self-correcting. *Perspectives on Psychological Science 7*(6): 645-654.

Ioannidis, J. P. A. (2020). *John P. A. Ioannidis.* Stanford University. Retrieved from https://med.stanford.edu/profiles/john-ioannidis?tab=publications [7th of September, 2020].

Ioannidis, J. P. A. (2020). *The Infection Fatality Rate of COVID-19 Inferred from Seroprevalence Data.* medRxiv. Retrieved from https://www.medrxiv.org/content/10.1101/2020.05.13.20101253v3.external-links.html [7th of September, 2020].

Ioannidis, J. P. A., Klavans, R., & Boyack, K. W. (2018). Thousands of scientists publish a paper every five days. *Nature 561*(7722): 167-169.

Jackson, P. A. (2006). Plagiarism instruction online: Assessing undergraduate student's ability to avoid plagiarism. *College & Research Libraries 67*(5): 418-428.

Jacso, P. (2005). As we may search – Comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases. *Current Science 89*(9): 1537-1547.

Jasny, B. R., Chin, G., Chong, L., & Vignerie, S. (2011). Again, and again, and again … . *Science 334*(6060): 1225.

Jefferson, T. (1813). Thomas Jefferson to Isaac McPherson. *The Founders' Constitution 3*(1): 8-8-12.

Jensen, K. B. (2008). Deduction vs Induction vs Abduction. In International Communication Association (Ed.), *The International Encyclopedia of Communication*. New York: Wiley.

Jez, S. J. (2014). The differential impact of wealth versus income in the college-going process. *Research in Higher Education 55*(7): 710-734.

John, K. L., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science 23*(5): 524-532.

Johnson, C. D., & Eby, L. T. (2011). Evaluating career success of African American males: It's what you know and who you are that matters. *Journal of Vocational Behavior 79*(3): 699-709.

Journal of Development Economics (2020). *New: Registered Reports.* Elsevier. Retrieved from https://www.journals.elsevier.com/journal-of-development-economics/announcements/new%20registered%20reports [5th of September, 2020].

Judge, T. A., Bono, J. E., Ilies, R., & Gerhardt, M. W. (2002). Personality and leadership: a qualitative and quantitative review. *Journal of Applied Psychology 87*(4): 765.

Judge, T. A., Cable, D. M., Colbert, A. E., & Rynes, S. L. (2007). What causes a management article to be cited?. *Academy of Management Journal 50*(3): 491-506.

Kandel, E., & Lazera, E. P. (1992). Peer pressure and partnerships. *Journal of Political Economy 100*(4): 801-817.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement 20*(1): 141-151.

Kaiser, R. B. & deVries, D. L. (2000). Leadership training. In W. E. Craighead & C. B. Nemeroff (Eds.), *The Corsini Encyclopedia of Psychology and Behavioral Science* (3rd ed.). New York: Wiley & Sons.

Karabag, S. F., & Berggren, C. (2012). Retraction, dishonesty and plagiarism: Analysis of a crucial issue for academic publishing, and the inadequate responses from leading journals in economics and management disciplines. *Journal of Applied Economics and Business Research 2*(3): 172-183.

Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research Policy 26*(1): 1-18.

Kaye, J., Heeney, C., Hawkins, N., de Vries, J., & Boddington, P. (2009). Data sharing in genomics – Re-shaping scientific practice. *Nature Reviews Genetics 10*(5): 331-335.

Kendall, D. L., & Campanario, S. C. (2016). Honoring god through scientific research: Navigating the ethics of publishing with our students. *International Journal of Christianity & Education 20*(2): 133-148.

Kennedy, M. S., Barnsteiner, J., & Daly, J. (2014). Honorary and ghost authorship in nursing publications. *Journal of Nursing Scholarship 46*(6): 416-422.

Kenny, L. W., Lee, L. F., Maddala, G. S., & Trost, R. P. (1979). Returns to college education: An investigation of self-selection bias based on the project talent data. *International Economic Review 20*(3): 775-789.

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review 2*(3): 196-217.

Kidwell, M. C., Lazarevic, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L. S. et al. (2016). Badges to acknowledge open practices: A simple, low-cost effective method for increasing transparency. *PLoS Biology 14*(5): e1002456.

Killingsworth, M. R., & Heckman, J. J. (1986). Female labor supply: A survey. *Handbook of Labor Economics 1*: 103-204.

Kim, H. J. (2001). The transition from paper to electronic journals. *The Serials Librarian 41*(1): 31-64.

Kim, Y., & Adler, M. (2015). Social scientists' data sharing behaviors: Investigating the roles of individual motivations, institutional pressures, and data repositories. *International Journal of Information Management 35*(4): 408-418.

Kim, Y., & Stanton, J. M. (2012). Institutional and individual influence on scientists' data sharing practices. *Journal of Computational Science Education 3*(1): 47-56.

Kim, Y., & Stanton, J. M. (2016). Institutional and individual factors affecting scientists' data-sharing behaviors: A multilevel analysis. *Journal of the Association for Information Science and Technology 67*(4): 776-799.

Kim, Y., & Zhang, P. (2015). Understanding data sharing behaviors of STEM researchers: The roles of attitudes, norms and data repositories. *Library and Information Science Research 37*(3): 189-200.

King, G. (1995). Replication, replication. *Political Science and Politics 28*(3): 444-452.

King, G. (2007). An introduction to the Dataverse network as an infrastructure for data sharing. *Sociological Methods & Research 36*(2): 173-199.

Kirkman, B. L., & Chen, G. (2011). Maximizing your data or data slicing? Recommendations for managing multiple submissions from the same dataset. *Management and Organization Review 7*(3): 433-446.

Kirwan, J. R. (1997). Making original data from clinical studies available for alternative analysis. *The Journal of Rheumatology 24*(5): 822-625.

Klein, C. J., & Moser-Veillon, P. B. (1999). Authorship: Can You Claim a Byline? *Journal of the American Dietetic Association 99*(1): 77-79.

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams Jr, R. B., Alper, S., ... & Batra, R. (2018). Many Labs 2: Investigating variation in replicability across samples and settings. *Advances in Methods and Practices in Psychological Science 1*(4): 443-490.

Koltay, T. (2019). Accepted and emerging roles of academic libraries in supporting research 2.0. *The Journal of Academic Librarianship 45*(2): 75-80.

Krawczyk, M., & Reuben, E. (2012). (Un)Available upon request: Field experiment on researchers' willingness to share supplementary materials. *Accountability in Research 19*(3): 175-186.

Krueger, T., Shorter, J., & Huff, K. (2012). International differences in business journal acceptance rates across business disciplines. *International Journal of Business and Social Science 3*(3): 1-16.

Krumpal, (2013). Determinants of social desirability bias in sensitive surveys: A literature review. *Quality & Quantity 47*(4): 2025-2047.

Krychkova, T. (2001). English as a language of science in Russia. In Ammon, U. (Ed.), *The Dominance of English as a Language of Science* (pp. 405-424). Berlin: Walter de Gruyter.

Kuhn, P. & Weinberger, C. (2005). Leadership skills and wages. *Journal of Labor Economics 23*(3): 395-436.

Kumar, S. (2018). Ethical concerns in the rise of co-authorship and its role as a proxy of research collaborations. *Publications 6*(37): 1-9.

Kwiek, M. (2015). Academic generations and academic work: Patterns of attitudes, behaviors, and research productivity of polish academics after 1989. *Studies in Higher Education 40*(8): 1354-1376.

La Porta, R., Lopez-De-Silanes, F., Shleifer, A., & Vishny, R. (2002). Investor protection and corporate valuation. *The Journal of Finance 57*(3): 1147-1170.

Laband, D. N. (2002). Contribution, attribution and the allocation of Intellectual Property rights: Economics versus agricultural economics. *Labour Economics 9*(1): 125-131..

Lake, D. A. (2010). Who's on first? Listing authors by relative contribution trumps the alphabet. *Political Science and Politics 43*(1): 43-47.

Lancaster, G. (2005). *Research Methods in Management.* Burlington: Elsevier Butterworth-Heinemann.

Langlois, R. N. (1998). Transaction costs, production costs, and the passage of time. In Medema S. G. (Ed.), *Coasean Economics Law and Economics and the New Institutional Economics* (pp. 1-21). Dordrecht: Springer.

Lariviere, V., Gingras, Y., & Archambault, E. (2006). Canadian collaboration networks: A comparative analysis of the natural sciences, social sciences and the humanities. *Scientometrics 68*(3): 519-533.

Lariviere, V., Gingras, Y., Sugimoto, C. R., & Tsou, A. (2015). Team size matters: Collaboration and scientific impact since 1900. *Journal of the Association for Information Science and Technology 66*(7): 1323-1332.

Lazear, E. P. (1995). *Personnel Economics.* Cambridge: MIT Press.

Lazear, E. P. (2012). Leadership: A personnel economics approach. *Labour Economics 19*(1): 92-101.

Leadership Quarterly (2020). *Guide for Authors.* Elsevier. Retrieved from https://www.elsevier.com/journals/the-leadership-quarterly/1048-9843/guide-for-authors [5th of September, 2020].

Leahey, E. (2006). Gender differences in productivity: Research specializations as a missing link. *Gender & Society 20*(6): 754-780.

Leahey, E., & Reikowsky, R. C. (2008). Research specialization and collaboration patterns in sociology. *Social Studies of Science 38*(3): 425-440.

Lechner, M. (2002). Program heterogeneity and propensity score matching: An application to the evaluation of active labor market policies. *Review of Economics and Statistics 84*(2): 205-220.

Lee, S., & Bozeman, B. (2005). The impact of research collaboration on scientific productivity. *Social Studies of Science 35*(5): 673-702.

Lehman, D., & Ramanujam, R. (2009). Selectivity in organizational rule violations. *Academy of Management Review 34*(4): 643-657.

Leung, C. K. S. (2011). Mining uncertain data. *WIREs Data Mining and Knowledge Discovery 1*(4): 316-329.

Leuven, E. & Sianesi, B. (2018). *PSMATCH2: Stata Module to Perform Full Mahalanobis and Propensity Score Matching, Common Support Graphing, and Covariate Imbalance Testing.* EconPapers. Retrieved from https://econpapers.repec.org/software/bocbocode/S432001.htm [12th of October, 2018].

Lewbel, A. (2012). Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business & Economic Statistics 30*(1): 67-80.

Li, M. (2013). Using the propensity score method to estimate causal effects: A review and practical guide. *Organizational Research Methods 16*(2): 188-226.

Liao, Q., Zhang, Y., Fan, Y., Zheng, M., Bai, Y., Eslick, G. D. et al. (2018). Perceptions of Chinese biomedical researchers towards academic misconduct: a comparison between 2015 and 2010. *Science and Engineering Ethics 24*(2): 629-645.

Lichtenthaler, U., & Ernst, H. (2009). RETRACTED: Technology licensing strategies: The interaction of process and content characteristic. *Strategic Organization 7*(2): 183-221.

Lichtenthaler, U., & Ernst, H. (2012). RETRACTED: Integrated knowledge exploitation: The complementarity of product development and technology licensing. *Strategic Management Journal 33*(5): 513-534.

Liddell, J. (2003). A Comprehensive definition of plagiarism. *Community & Junior College Libraries 11*(3): 43-52.

Lindsey, D. (1980). Production and citation measures in the sociology of science: The problem of multiple authorship. *Social Studies of Science 10*(2): 145-162.

Linek, S. B., Fecher, B., Friesike, S., & Hebing, M. (2017). Data sharing as social dilemma: Influence of the researcher's personality. *PloS ONE 12*(8): e0183216.

Lipton, P. (2005). Testing hypotheses: Prediction and prejudice. *Science 307*(5707): 219-221.

Long, J. S. (1978). Productivity and academic position in the scientific career. *American Sociological Review 43*(6): 889-908.

Long, J. S., & McGinnis, R. (1982). On adjusting productivity measures for multiple authorship. *Scientometrics 4*(5): 379-387.

Longo, D. L., & Drazen, J. M. (2016). Data sharing. *The New England Journal of Medicine 374*(3): 276-277.

Lozano, F. A. (2008). Language, high school leadership and the postsecondary outcome of Hispanic students. *Economics of Education Review 27*(3): 342-353.

Lu, S. F., Jin, G. Z., Uzzi B., & Jones, B. (2013). The retraction penalty: Evidence from the Web of Science. *Scientific Reports 3*(3146): 1-5.

Luparenko, L. (2014). Plagiarism detection tools for scientific e-journals. In Ermolayev, V., Mayr, H. C., Nikitchenko, M., & Spivakovsky, A. (Eds.), *Information and Communication Technologies in Education, Research and Industrial Applications*. Heidelberg: Springer.

Lych, P. B. (1966). *Conduct of Field Experiments.* Wellington: R. E. Owen.

Lykkesfeldt, J. (2016). Strategies for using plagiarism software in the screening of incoming journal manuscripts: Recommendations based on a recent literature survey. *Basic & Clinical Pharmacology & Toxicology 119*(2): 161-164.

Macfarlane, B. (2017). The ethics of multiple authorship: Power, performativity and the gift economy. *Studies in Higher Education 42*(7): 1194-1210.

Madhok, A., & Tallman, S. B. (1998). Resources, transactions and rents: Managing value through interfirm collaborative relationships. *Organization Science 9*(3): 326-339.

Mankiw, N. G. (2015). *Principles of Economics* (7th ed.). Stamford: Cengage Learning.

Manolis, J. C., Chan, K. M., Finkelstein, M. E., Stephens, S., Nelson, C. R., Grant, J. B., & Dombeck, M. P. (2009). Leadership: a new frontier in conservation science. *Conservation Biology 23*(4): 879-886.

Mansfield, E. R., & Helms, B. P. (1982). Detecting multicollinearity. *The American Statistician 36*(3a): 158-160.

Manton, E. J., & English, D. E. (2006). Reasons for co-authorship in business journals and the extent of guest or gift authorship. *Delta Pi Epsilon Journal 48*(2): 86-95.

Manton, E. J., & English, D. E. (2007). The trend toward multiple authorship in business journals. *Journal of Education for Business 82*(3): 164-168.

Manton, E. J., & English, D. E. (2008). An empirical study of gift authorship in business journals. *Journal of Education for Business 83*(5): 283-287.

Marcen, C., Gimeno, F., Gomez, C., Saenz, A., & Gutierreza, H. (2013). Socioeconomic status, parental support, motivation and self-confidence in youth competitive sport. *Procedia-Social and Behavioral Sciences 82*: 750-754.

Marincola, F. M. (2003). Introduction of article-processing charges (APCs) for articles accepted for publication in the Journal of Translational Medicine. *Journal of Translational Medicine 1*(11): 1-2.

Marshakova-Shaikevich, I. (2006). Scientific collaboration of new 10 EU countries in the field of social sciences. *Information Processing and Management 42*(6): 1592-1598.

Martin, B. R. (2013). Whither research integrity? Plagiarism, self-plagiarism and coercive citation in an age of research assessment. *Research Policy 42*(5): 1005-1014.

Martin, D. W. (2008). *Doing Psychology Experiments* (7th ed.). Belmont, CA: Thomson Higher Education.

Martinson, B. C., Anderson, M. S., & de Vries, R. (2005). Scientists behaving badly. *Nature 435*(7043): 737-738.

Marusic, A., Bosnjak, L., & Jeroncic, A. (2011). A systematic review of research on the meaning, ethics and practices of authorship across scholarly disciplines. *PLoS ONE 6*(9): e23477.

Marx, M., & Fuegi, A. (2020a). Reliance on science: Worldwide front-page patent citations to scientific articles. *Strategic Management Journal* Published online first 16th of April, 2020.

Marx, M. and A. Fuegi. (2020b). *Reliance on Science in Patenting (Version v23).* Zenodo. Retrieved from http://doi.org/10.5281/zenodo.3755799 [27th of July, 2020].

Matsa, D. A., & Miller, A. R. (2011). Chipping away at the glass ceiling: Gender spillovers in corporate leadership. *American Economic Review 101*(3): 635-39.

Mazzola, J. J., & Deuling, J. K. (2013). Forgetting what we learned as graduate students: HARKing and selective outcome reporting in I-O journal articles. *Industrial and Organizational Psychology 6*(3): 279-284.

McDonald, R. J., Neff, K. L., Rethlefsen, M. L., & Kallmes, D. F. (2010). Effects of author contribution disclosures and numeric limitations on authorship trends. *Mayo Clinic Proceedings 82*(10): 920-927.

McGrail, M., Rickard, C. M., & Jones, R. (2006). Publish or perish: A systematic review of interventions to increase academic publication rates. *Higher Education Research & Development 25*(1): 19-35.

McKee, S. C., Rider T. J., Thames, F. C., Wallace, S. J., Williams, D. L., & Ragauskas, R. (2019). *Political Research Quarterly 2019 Activity Report.* Paper presented at the WPSA Annual Meeting 2020, Los Angeles.

Merton, R. K. (1942). A note on science and democracy. *Journal of Legal and Political Sociology 1*(1): 115-126.

Merton, R. K. (1968). The Matthew effect in science. *Science 159*(3810): 56-63.

Merton, R. K. (1969). Behavior patterns of scientists. *The American Scholar 57*(1): 1-23.

Milan, J. (1997). Comparison of the quality of rail and air networks in West, Central and Eastern Europe. *Transport Policy 4*(2): 85-93.

Miller, A. N., Taylor, S. G., & Bedeian, A. G. (2011). Publish or perish: Academic life as management faculty live it. *Career Development International 16*(5): 422-445.

Misner, C. W. (1960). Wormhole initial conditions. *Physical Review 118*(4): 1110-1111.

Mitchell, M. B. (1961). Trends toward multiple authorship in scientific publications. *The Journal of Psychology 52*(1): 125-131.

Moedas, C. (2015). *Open Innovation, Open Science, Open to the World.* European Commission. Retrieved from https://europa.eu/rapid/press-release_SPEECH-15-5243_en.htm [22nd of June, 2019].

Moffatt, B. (2011). Responsible authorship: Why researchers must forgo honorary authorship. *Accountability in Research 18*(2): 76-90.

Moffatt, B., & Elliott, C. (2007). Ghost marketing: Pharmaceutical companies and ghostwritten journal articles. *Perspectives in Biology and Medicine 50*(1): 18-31.

Moizer, P. (2009). Publishing in accounting journals: A fair game?. *Accounting, Organizations and Society 34*(2): 285-304.

Molckovsky, A., Vickers, M. M., & Tang, P. A. (2011). Characterization of published errors in high-impact oncology journals. *Current Oncology 18*(1): 26-32.

Möllering, G. (2002). Perceived trustworthiness and inter-firm governance: Empirical evidence from the UK printing industry. *Cambridge Journal of Economics 26*(2): 139-160.

Molloy, J. C. (2011). The open knowledge foundation: Open data means better science. *PLoS Biology 9*(12): e100195.

Mons, B., Neylon, C., Velterop, J., Dumontier, M., da Silva Santos, L. O. B., & Wilkinson, M. D. (2017). Cloudy, increasingly FAIR: Revisiting the FAIR Data guiding principles for the European Open Science Cloud. *Information Services & Use 37*(1): 49-56.

Moreira, M. J. (2003). A conditional likelihood ratio test for structural models. *Econometrica 71*(4): 1027-1048.

Morey, R. D., Chambers, C. D., Etchells, P. J., Harris, C. R., Hoekstra, R., Lakens, D. et al. (2016). The peer reviewers' openness initiative: Incentivizing open research practices through peer review. *Royal Society Open Science 3*(1): 150547.

Morrison, H., Salhab, J., Calve-Genest, a., & Horava, T. (2015). Open access article processing charges: DOAJ survey May 2014. *Publications 3*(1): 1-16.

Morse, J. (2000). *In or Out: Inside College Admission.* TIME. Retrieved from http://content.time.com/time/nation/article/0,8599,57724,00.html [12th of October, 2018].

Mowatt, G., Shirran, L., Grimshaw, J. M., Rennie, D., Flanagin, A., Yank, V. et al. (2002). Prevalence of honorary and ghost authorship in Cochrane Reviews. *Journal of the American Medical Association 287*(21): 2769-2771.

Mowery, D. C., Oxley, J. E., & Silverman, B. S. (1998). Technological overlap and interfirm cooperation: Implications for the resource-based view of the firm. *Research Policy 27*(5): 507-523.

Müller, A. C., & Guido, S. (2016). *Introduction to Machine Learning with Python: A Guide for Data Scientists.* Sebastopol: O'Reilly.

Murphy, K. R., & Aguinis, H. (2019). HARKing: How badly can cherry-picking and question trolling produce bias in published results?. *Journal of Business and Psychology 34*(1): 1-17.

Murphy, S. E., & Johnson, S. K. (2011). The benefits of a long-lens approach to leader development: Understanding the seeds of leadership. *The Leadership Quarterly 22*(3): 459-470.

Murray, M. J. (2016). Thanks for sharing: The bumpy road towards truly open data. *BMJ 374*(3): 276-277.

Muscio, A., Quaglione, D., & Vallanti, G. (2013). Does government funding complement or substitute private research funding to universities? *Research Policy 42*(1): 63-75.

Nagy, A. M. (2016). *International Scientific Collaboration Links of Central Eastern European Countries Measured through Publications.* Paper presented at the Enterprise Research Innovation Conference, Rovinj.

Nature Human Behaviour (2017). Promoting reproducibility with registered reports. *Nature Human Behavior 1*(1): 34.

Neher, A. (1967). Probability pyramiding, research error and the need for independent replication. *The Psychological Record 17*(2): 257-262.

Niles, H. E. (1922). Correlation, causation and Wright's theory of "path coefficients". *Genetics 7*(3): 258-273.

Northouse, P. G. (1997). *Leadership: Theory and Research.* Thousand Oaks: Sage Publications.

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J. et al. (2015). Promoting an open research culture. *Science 348*(6242): 1422-1425.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory* (3rd ed.). New York: McGraw-Hill.

O'Brien, R.M. (2007). A caution regarding rule of thumb for variance inflation factors. *Quality and Quantity 41*(5): 673-90.

O'Brien, T. L. (2011). Change in academic coauthorship. 1953-2003. *Science, Technology & Human Values 37*(3): 210-234.

Oberlander, S. E., & Spencer, R. J. (2006). Graduate students and the culture of authorship. *Ethics & Behavior 16*(3): 217-232.

Offord, C. (2020). *How (Not) to Do an Antibody Survey for SARS-CoV-2.* The Scientist. Retrieved from https://www.the-scientist.com/news-opinion/how-not-to-do-an-antibody-survey-for-sars-cov-2-67488 [7th of September, 2020].

Olea, J. L. M., & Pflueger, C. (2013). A robust test for weak instruments. *Journal of Business & Economic Statistics 31*(3): 358-369.

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science 349(*6251): 943.

Open Science Framework Preregistration (2020). *OSF Preregistration.* Open Science Framework. Retrieved from https://osf.io/prereg/ [5th of September, 2020].

Ortega, J. L. (2017). Are peer-review activities related to reviewer bibliometric performance? A scientometric analysis of Publons. *Scientometrics 112*(2): 947-962.

Osborne, J. W., & Holland, A. (2009). What is authorship and what should it be? A survey of prominent guidelines for determining authorship in scientific publications. *Practical Assessment 14*(15): 1-19.

Osherovich, L. (2011). Hedging against academic risk. *Science-Business eXchange 4*(15): 1-2).

Ospina, S., & Foldy, E. (2009). A critical review of race and ethnicity in the leadership literature: Surfacing context, power and the collective dimensions of leadership. *The Leadership Quarterly 20*(6): 876-896.

Ossenblok, T. L. B., Engels, T. C. E., & Sivertsen, G. (2012). The representation of the social sciences and humanities in the Web of Science. A comparison of publication patterns and incentive structures in Flanders and Norway (2005-2009). *Research Evaluation 21*(4): 280-290.

Ossenblok, T. L. B., Verleysen, F. T., & Engels, T. C. E. (2014). Coauthorship of journal articles and book chapters in the social sciences and humanities (2000-2010). *Journal of the Association for Information Science and Technology 65*(5): 882-897.

Özer, Ö., Zheng, Y., & Ren, Y. (2014). Trust, trustworthiness, and information sharing in supply chains bridging China and the United States. *Management Science 60*(10): 2435-2460.

Padmanabhan, S. (2015). Who is an author? Rights and righteousness! *Journal of Indian Orthodontic Society 49*(3): 127-128.

Palmer, B., Walls, M., Burgess, Z., & Stough, C. (2001). Emotional intelligence and effective leadership. *Leadership & Organization Development Journal 22*(1): 5-10.

Park, S. H., & Gordon, M. E. (1996). Publication records and tenure decisions in the field of strategic management. *Strategic Management Journal 17*(2): 109-128.

Partha, D., &, David, P. A. 1994. Toward a new economics of science. *Research Policy 23*(5): 487-521.

Pedersen, O. (1983). Galileo and the council of Trent: The Galileo affair revisited. *Journal for the History of Astronomy 14*(1): 1-29.

Peel, M. J., & Makepeace, G. H. (2012). Differential audit quality, propensity score matching and Rosenbaum bounds for confounding variables. *Journal of Business Finance & Accounting 39*(5-6): 606-648.

Persoskie, A., & Nelson, W. L. (2013). Just blowing smoke? Social desirability and reporting of intentions to quit smoking. *Nicotine & Tobacco Research 15*(12): 2088-2093.

Peters, C. (2018). Journalism needs a better argument: Aligning public goals with the realities of the digital news and information landscape. *Journalism 20*(1): 73-76.

Peters, I., Kraker, P., Lex, E., Gumpenberger, C., & Gorraiz, J. (2016). Research data explored: An extended analysis of citations and altmetrics. *Scientometrics 107*(2): 723-744.

Pfeffer, J. (1993). Barriers to the advance of organizational science; Paradigm development as a dependent variable. *Academy of Management Review 18*(4): 599-620.

Pfenninger, S., DeCarolis, J., Hirth, L., Quoilin, S., & Staffell, I. (2017). The importance of open data and software: Is energy research lagging behind?. *Energy Policy 101*(3): 211-215.

Pierro, A., Cicero, L., Bonaiuto, M., van Knippenberg, D. & Kruglanski, A. W. (2005). Leader group prototypicality and leadership effectiveness: The moderating role of need for cognitive closure. *The Leadership Quarterly 16*(4): 503-516.

Piocuda, J. E., Smyers, J. O., Knyshev, E., & Harris, R. J. (2015). Trends of internationalization and collaboration in U.S. psychology journals 1950-2010. *Archives of Scientific Psychology 3*(1): 82-92.

Piro, F. N., Aksnes, D. W., & Rorstad, K. (2013). A macro analysis of productivity differences across fields: Challenges in the measurement of scientific publishing. *Journal of the American Society for Information Science and Technology 64*(2): 307-320.

Pitman, R. (2009). *Who Are You Calling Irrational? Marginal Costs, Variable Costs, and the Pricing Practices of Firms.* EAG Discussion Paper No. 09-3. Washington DC: Economic Analysis Group.

Piwowar, H. A., & Chapman, W. (2008). *A Review of Journal Policies for Sharing Research Data.* Paper presented at the International Conference on Electronic Publishing, Toronto.

Piwowar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. *PeerJ 1*: e175.

Piwowar, H. A., Day, R. S., & Fridsma, D. B. (2007). Sharing detailed research data is associated with increased citation rate. *PLoS ONE 2*(3): e308.

Plato (399 BC). *Apology.* Retrieved from https://www.gutenberg.org/ebooks/1656 [18th of August, 2020].

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology 88*(5): 879.

Polachek, S. W., & Jun, X. (2009). *The Gender Pay Gap Across Countries: A Human Capital Approach.* SOEPpaper on Multidiciplinary Panel Data Research No. 227, Berlin: DIW.

Popper, K. (1935). Logik der Forschung – Zur Erkenntnistheorie der modernen Naturwissenschaft. *Schriften zur Wissenschaftlichen Weltauffassung 9*(1): 1-248.

Pratt, M. G., Kaplan, S., & Whittington, R. (2020). Editorial essay: The tumult over transparency: Decoupling transparency from replication in establishing trustworthy qualitative research. *Administrative Science Quarterly 65*(1): 1-19.

Pruschak, G., & Hopp, C. (2019). *And the Credit Goes to … - Ghost and Honorary Authorship among Social Scientists.* Paper Presented at the Academy of Management 2019, Boston.

Pupovac, V., Bilic-Zulle, L., & Petrovecki, M. (2008). On academic plagiarism in Europe. An analytical approach based on four studies. *Digithum 2008*(10): 14-18.

Ragins, B. R., Townsend, B., & Mattis, M. (1998). Gender gap in the executive suite: CEOs and female executives report on breaking the glass ceiling. *Academy of Management Perspectives 12*(1): 28-42.

Randhawa, K., Wilden, R., & Hohberger, J. (2016). A bibliometric review of open innovation: Setting a research agenda. *Journal of Product Innovation Management 33*(6): 750-772.

Rawat, S., & Meena, S. (2014). Publish or perish: Where are we heading? *Journal of Research in Medical Sciences 19*(2): 87-89.

Reed, D. A., Enders, F., Lindor, R., McClees, M., & Lindor, K. D. (2011). Gender differences in academic productivity and leadership appointments of physicians throughout academic careers. *Academic Medicine 86*(1): 43.47.

Reich, E. S. (2012). *Boost for Higgs from Tevatron Data. Nature.* Retrieved from https://www.nature.com/news/boost-for-higgs-from-tevatron-data-1.10167 [27th of September, 2019].

Reichman, O. J., Jones, M. B., & Schildhauer, M. P. (2011). Challenges and opportunities of open data in ecology. *Science 331*(6018): 703-705.

Reidpath, D. D., & Allotey, P. A. (2001). Data sharing in medical research: An empirical investigation. *Bioethics 15*(2): 125-134.

Reinhart, C. M., & Rogoff, K. S. (2010). Growth in a time of debt. *American Economic Review 100*(2): 573-78.

Reitan, T., & Stenberg, S. A. (2019). From classroom to conscription. Leadership emergence in childhood and early adulthood. *The Leadership Quarterly 30*(3): 298-319.

Retraction Watch (2014). *After 16 Retractions, Management Professor Lichtenthaler Resigns Post.* Retraction Watch. Retrieved from https://retractionwatch.com/2014/10/10/after-16-retractions-management-professor-lichtenthaler-resigns-post/ [3rd of September, 2020].

Reynolds, P. D. (2007). *New Firm Creation in the United States: Initial Explorations with the PSED I Data Set.* Hanover, MA: Now Publishers.

Ricardo, D. (1817). *On the Principles of Political Economy and Taxation.* London: John Murray.

Richardson, H. A., Simmering, M. J., & Sturman, M. C. (2009). A tale of three perspectives: Examining post hoc statistical techniques for detection and correction of common method variance. *Organizational Research Methods 12*(4): 762-800

Robles, M. M. (2012). Executive perceptions of the top 10 soft skills needed in today's workplace. *Business and Professional Communication Quarterly 75*(4): 453-465.

Rodriguez, V. (2005). Material transfer agreements: Open science vs. proprietary claims. *Nature Biotechnology 23*(4): 489-491.

Rohrer, J., Tierney, W., Uhlmann, E. L., De Bruine, L., Heyman, T., Jones, B. et al. (2020). Putting the self in self-correction: Findings from the loss-of confidence project. *Perspectives on Psychological Science* Published online first 20th of July, 2020.

Rose-Wiles, L. M. (2018). Reflections on fake news, librarians, and undergraduate research. *Reference & User Services Quarterly 57*(3): 200-204.

Rosenbaum, P. R. & Rubin, D. B. (1983). The central role of propensity score in observational studies for causal effects. *Biometrika 70*(1): 41-55.

Rosenbaum, P. R. (2002). *Observational Studies* (2nd ed.). New York: Springer.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin 86*(3): 638-641.

Ross, J. S., & Krumholz, H. M. (2013). Ushering in a new era of open science through data sharing. *Journal of the American Medical Association 309*(13): 1355-1356.

Rothgeb, J. M. Jr., & Burger, B. (2009). Tenure standards in political science departments: Results from a survey of department chairs. *Political Science and Politics 42*(3): 513-519.

Rouse, K. E. (2012). The impact of high school leadership on subsequent educational attainment. *Social Science Quarterly 93*(1): 110-129.

Rousseau, R. (2009). The most influential editorials. In Astrom, F., R. Danell, B. Larsen & J. W. Schneider (Eds.), *Celebrating Scholarly Communication Studies* (pp. 47-54). Leuven: International Society for Scientometrics and Informetrics.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology 66*(5): 688-701.

Rubin, M. (2019). The costs of HARKing. *The British Journal for the Philosophy of Science* Published online first 3rd of December, 2019.

Salita, J. T. (2010). Authorship Practices in Asian Cultures. *The Write Stuff 19*(1): 36-38.

Sandler, J. C., & Russel, B. L. (2005). Faculty-student collaborations: Ethics and satisfaction in authorship credit. *Ethics & Behavior 15* (1): 65-80.

Sauermann, H., & Haeussler, C. (2017). Authorship and contribution disclosure. *Science advances 3*(11): e1700404.

Saunders, M., Lewis, P., & Thornhill, A. (2016). *Research Methods for Business Students.* Harlow: Pearson Education.

Savage, C. J., & Vickers, A. J. (2009). Empirical study of data sharing by authors publishing in PLoS journals. *PLoS ONE 4*(9): e7078.

Scalzini, S. (2015). Is there free-riding? A comparative analysis of the problem of protecting publishing materials online in Europe. *Journal of Intellecutal Property Law & Practice 10*(6): 454-464.

Schaffer, M. E., Baum, C., Finlay, K., Kleibergen, F., Magnusson, L., & Stillman, S. (2013). *Stata Software for Econometric Estimation and Testing; avar, weakiv, actest, ivreg2h, ranktest, ivreg2.* Retrieved from https://researchportal.hw.ac.uk/files/4818130/Schaffer_Stata_software_components_for_Documents_section_REVISED.pdf [1st of August, 2019].

Scheiber, N. (2020). *When Scholars Collaborate With Tech Companies, How Reliable Are the Findings?.* The New York Times. Retrieved from https://www.nytimes.com/2020/07/12/business/economy/uber-lyft-drivers-wages.html [27th of July, 2020].

Schmidt, S. 2009. Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology 13*(2): 90-100.

Schminke, M. (2009). Editor's comment: The better angels of our nature-ethics and integrity in the publishing process. *Academy of Management Review 34*(4): 586-591.

Schminke, M., & Ambrose, M. L. (2011). RETRACTED: Ethics and integrity in the publishing process: Myths, facts and a roadmap. *Management and Organization Review 7*(3): 397-406.

Schwenk, C. (1993). Management tenure and explanations for success and failure. *Omega 21*(4): 449-456.

Selmier, W. T., & Oh, C. H. (2012). The power of major trade languages in trade and foreign direct investment. *Review of International Political Economy 20*(3): 486-514.

Selwyn, N. (2008). 'Not necessarily a bad thing …': A study of online plagiarism amongst undergraduate students. *Assessment & Evaluation in Higher Education 33*(5): 465-479.

Seppänen, P. (1982). Sport clubs and parents as socializing agents in sports. *International Review for the Sociology of Sport 17*(1): 79-90.

Serenko, A., Bontis, N., Booker, L., Saddedin, K., & Hardie, T. (2010). A scientometric analysis of knowledge management and intellectual capital academic literature (1994-2008). *Journal of Knowledge Management 14*(1): 3-23.

Sestak, J., Fiala, J., & Gavrichev, K. S. (2018). Evaluation of the professional worth of scientific papers, their citation responding and the publication authority. *Journal of Thermal Analysis and Calorimetry 131*(1): 463-471.

Shahabuddin, S. (2009). Plagiarism in academia. *International Journal of Teaching and Learning in Higher Education 21*(3): 353-359.

Shamir, B. (2007). From passive recipients to active coproducers: followers' roles in the leadership process. In Shamir, B., Pillai, R., Bligh, M. C. & Uhl-Bien, M. (Eds.), *Follower-Centered Perspectives on Leadership: A Tribute to the Memory of James R. Meindl* (pp. 9-34). Greenwich: Information Age.

Shashok, K., & Handjani, F. (2010). Enhancing the quality of research publication: AuthorAID in the Eastern Mediterranean. *Journal of Tehran University Heart Center 5*(4): 169-171.

Shaver, J. M. (1998). Accounting for endogeneity when assessing strategy performance: Does entry mode choice affect FDI survival? *Management Science 44*(4): 571-585.

Shaw, B. T. (1967). *The Use of Quality and Quantity of Publication as Criteria for Evaluating Scientists.* Washington DC: United States Department of Agriculture.

Shondrick, S. J., Dinh, J. E., & Lord, R. G. (2010). Developments in implicit leadership theory and cognitive science: Applications to improving measurement and understanding alternatives to hierarchical leadership. *The Leadership Quarterly 21*(6): 959-978.

Siemsen, E., Roth, A., & Oliveira, P. (2010). Common method bias in regression models with linear, quadratic, and interaction effects. *Organizational research methods 13*(3): 456-476.

Sigurdsson, E. L. (2018). #metoo – A concern for general practice?. *Scandinavian Journal of Primary Health Care 36*(1): 1-2.

Silberzahn, R., & Uhlmann, E. L. (2013). It pays to be Herr Kaiser: Germans with noble-sounding surnames more often work as managers than as employees. *Psychological Science 24*(12): 2437-2444.

Silberzahn, R., Simonsohn, U., & Uhlmann, E. L. (2014). Matched-names analysis reveals no evidence of name-meaning effects: A collaborative commentary on Silberzahn and Uhlmann (2013). *Psychological Science 25*(7): 1504-1505.

Silverman, B. S. (1999). Technological resources and the direction of corporate diversification: Toward an integration of the resource-based view and transaction cost economics. *Management Science 45*(8): 1109-1124.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science 22*(11): 1359-1366.

Simons, D. J., Holcombe, A. O., & Spellman, B. A. (2014). An introduction to registered replication reports at *Perspectives on Sociological Science*. *Perspectives on Psychological Science 9*(5): 552-555.

Sismondo, S. (2008). Pharmaceutical company funding and its consequences: A qualitative systematic review. *Contemporary Clinical Trials 29*(2): 109-113.

Sloman, S. A., & Lagnado, D. A. (2005). The problem of induction. In Hoyoak, K. J., & Morrison, R. G. (Eds.), *The Cambridge Handbook of Thinking and Reasoning.* Cambridge: Cambridge University Press.

Smarr, L. L. (1979). Sources of gravitational radiation. *Proceedings of the Battelle Seattle Workshop* 245-274.

Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations.* London: W. Strahan & T. Cadell.

Smith, M. (1958). The trend toward multiple authorship in psychology. *American Psychologist 13*(10): 596-599.

Smith, P. B., Misumi, J., Tayeb, M., Peterson, M., Bond, M. (1989). On the generality of leadership style measures across cultures. *Journal of Occupational Psychology 62*(2): 97-109.

Sorenson, O., Assenova, V., Li, G. C., Boada, J., & Fleming, L. (2016). Expand innovation finance via crowdfunding. *Science 354*(6319): 1526-1528.

Spaulding, M. (2009). Perceptions of academic honesty in online vs. face-to-face classrooms. *Journal of Interactive Online Learning 8*(3): 183-198.

Suber, P. (2007). *Open Access Overview.* Philpapers. Retrieved from: https://philpapers.org/rec/SUBOAO [6th of November, 2020].

Stallings, W. M., & Singhai, S. (1970). Some observations on the relationship between research productivity and student evaluations of courses and teaching. *American Sociologist 5*(2): 141-143.

Stapel, D. A., & Lindenberg, S. (2011). RETRACTED: Coping with chaos: How disordered contexts promote stereotyping and discrimination. *Science 332*(6026): 251-253.

Stelder, D. (2016). Regional accessibility trends in Europe: Road infrastructure, 1957-2012. *Regional Studies 50*(6): 983-995.

Sterling, T. D. (1959). Publication decisions and their possible effects on interferences drawn from tests of significance – or vice versa. *Journal of the American Statistical Association 54*(285): 30-34.

Stertz, A. M., Grether, T., & Wiese, B. S. (2017). Gender-role attitudes and parental work decisions after childbirth: A longitudinal dyadic perspective with dual-earner couples. *Journal of Vocational Behavior 101*(1): 104-118.

Stevenson, B. (2010). Beyond the classroom: Using Title IX to measure the return to high school sports. *The Review of Economics and Statistics 92*(2): 284-301.

Stitzel, B., Hoover, G. A., & Clark, W. (2018). More on plagiarism in the social sciences. *Social Science Quarterly 99*(3): 1075-1088.

Stock, J. H., & Yogo, M. (2005). Asymptotic distributions of instrumental variables statistics with many instruments. In Andrews, D. W. K., & Stock, J. H. (Eds.), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg* (pp. 109-120). Cambridge: Cambridge University Press.

Stodden, V., Guo, P., & Ma, Z. (2013). Toward reproducible computational research: An empirical analysis of data and code policy adoption by journals. *PLoS ONE 8*(6): e67111.

Stodden, V., Seiler, J., & Ma, Z. (2018). An empirical analysis of journal policy effectiveness for computational reproducibility. *Proceedings of the National Academy of Sciences 115*(11): 2584-2589.

Stone, C., Scott, L., Battle, D. & Maher, P. (2014). Locating longitudinal respondents after a 50-year hiatus. *Journal of Official Statistics 30*(2): 311-334.

Storbeck, O. (2012). *Mannheimer BWL-Professor im Zwielicht.* Handelsblatt. Retrieved from https://www.handelsblatt.com/politik/deutschland/ungereimtheiten-mannheimer-bwl-professor-im-zwielicht/6895802.html?ticket=ST-1323830-ettj4yezMNkzdgc6LXYI-ap3 [3rd of September, 2020].

Sturges, P., Bamkin, M., Anders, J. H., Hubbard, B., Hussain, A., & Heeley, M. (2015). Research data sharing: Developing a stakeholder-driven model for journal policies. *Journal of the Association for Information Science and Technology 66*(12): 2445-2455.

Stürmer, S., Oeberst, A., Trötschel, F., & Decker, O. (2017). Early-career researchers‘ perceptions of the prevalence of questionable research practices, potential causes, and open science. *Social Psychology 48*(6): 365-371.

Sugimoto, C. R., Lariviere, V., Ni, C., & Cronin, B. (2013). Journal acceptance rates: A cross-disciplinary analysis of variability and relationships with journal metrics. *Journal of Informetrics 7*(4): 897-906.

Sun, Y. C. (2013). Do journal authors plagiarize? Using plagiarism detection software to uncover matching text across disciplines. *Journal of English for Academic Purpose 12*(4): 264-272.

Tardy, C. (2004). The role of English in scientific communication: lingua franca or Tyrannosaurus rex?. *Journal of English for Academic Purposes 3*(3): 247-269.

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. E., Wu, L., Read, E. et al. (2011). Data sharing by scientists: Practices and perceptions. *PLoS ONE 6*(6): e21101.

Tenopir, C., Hughes, D., Allard, S., Frame, M., Birch, B., Sandusky, R. et al. (2015). Research data services in academic libraries: Data intensive roles for the future?. *Journal of eScience Librarianship 4*(2): e1085.

Teute, F. J. (2001). To publish or perish: Who are the dinosaurs in scholarly publishing?. *Journal of Scholarly Publishing 32*(2): 102-112.

Thanuskodi, S. (2010). Journal of Social Sciences: A bibliometric study. *Journal of Social Sciences 24*(2): 77-80.

Tharenou, P., Donohue, R., & Cooper, B. (2012). *Management Research Methods.* Cambridge: Cambridge University Press.

Thatcher, A. (2006). Building and maintaining an online academic conference series. *International Journal of Industrial Ergonomics 36*(12): 1081-1088.

The American Sogiologist (2020). *Submission Guidelines.* Springer. Retrieved from https://www.springer.com/journal/12108/submission-guidelines [7th of September, 2020].

Thomas, D. A. (2001). The truth about mentoring minorities: Race matters. *Harvard Business Review 79*(4): 98-107.

Thorne, K. S. (2017). LIGO and gravitational waves, III: Nobel lecture, December 8, 2017. *Annalen der Physik 531*(1): e1800350.

Thurman, R. H., Chervenak, F., McCullough, L. B., Halwani, S., & Farine, D. (2016). *American Journal of Obstetrics and Gynecology 214*(1): 91-93.

Tichy, W. F. (1997). Should computer scientists experiment more?. *Computer 31*(5): 32-40.

Ting, E. (2020). *Experts Question Results of Startling Santa Clara Coronavirus Antibody Study.* SFGate. Retrieved from https://www.sfgate.com/bayarea/article/Santa-Clara-antibody-test-study-Stanford-COVID-19-15210149.php [7th of September, 2020].

Ting, E. (2020). *Researchers Adjust Results of Startling Santa Clara Antibody Study.* SFGate. Retrieved from https://www.sfgate.com/news/editorspicks/article/Santa-Clara-antibody-study-revised-Stanford-death-15263047.php [7th of September, 2020].

Tobin, M. (2003). April 25, 1953: Three papers, three lessons. *American Journal of Respiratory and Critical Care Medicine 167*(8): 1047-1049.

Towry, K. L. (2003). Control in a teamwork environment – The impact of social ties on the effectiveness of mutual monitoring contracts. *The Accounting Review 78*(4): 1069-1095.

Trappmann, M., Krumpal, I., Kirchner, A., Jann, B. (2014). Item sum: A new technique for asking quantitative sensitive questions. *Journal of Survey Statistics and Methodology 2*(1): 58-77.

Tryon, G. S., Bishop, J. L., & Hatfield, T. A. (2007). Doctoral students' beliefs about authorship credit for dissertations. *Training and Education in Professional Psychology 1*: 184-192.

Tsui, A. S., Lewin, A. Y., Schminke, M., & Ambrose, M. L. (2014). Retraction statement for 'Ethics and integrity of the publishing process: Myths, facts, and a roadmap' by Marhall Schminke and Maureen L. Ambrose. *Management and Organization Review 10*(1): 157-162.

Tucci, C.L., Chesbrough, H., Piller, F., & West, J. (2016). When do firms undertake open, collaborative activities? *Industrial and Corporate Change 25*(2): 283-288.

Tukey, J. W. (1980). We need both exploratory and confirmatory. *The American Statistician 34*(1): 23-25.

TurnItIN (2020). *Turnitin.* Turnitin. Retrieved form https://www.turnitin.com/de [7th of September, 2020].

Tymoczko, M. (2001). Will the traditional humanities survive the 21st century?. *Organizations 8*(2): 285-297.

Uhlir, P. F., & Schröder, P. (2007). Open data for global science. *Data Science Journal 6(OD)*: 36-53.

UniCheck (2020). *UniCheck.* UniCheck. Retrieved from https://unicheck.com [7th of September, 2020].

Valderas, J. M., R. Bentley, A., Buckley, R., Wray, K. B., Wuchty, S., Jones, B. F., & Uzzi, B. (2007). Why do team-authored papers get cited more?. *Science 317*(5844): 1496-1498.

Van den Berg, H. (2005). Reanalyzing qualitative interviews from different angles: The risk of decontextualization and other problems of sharing qualitative data. *Forum: Qualitative Social Research 6*(1): 30.

van Raan, A. F. J. (2001). Competition amongst scientists for publication status: Toward a model of scientific publication and citation distributions. *Scientometrics 51*(1): 347-357.

Vancouver, J. B. (2018). In defense of HARKing. *Industrial and Organizational Psychology 11*(1): 73-80.

Vanhaverbeke, W., & Cloodt, M. (2014). Theories of the firm and open innovation. In Chesbrough, H. W., Vanhaberbeke, W., & West, J. (Eds.), *New Frontiers in Open Innovation* (pp. 256-278). Oxford: Oxford University Press.

Vasilevsky, N. A., Minnier, J., Haendel, M. A., & Champieux, R. E. (2017). Reproducible and reusable research: Are journal data sharing policies meeting the mark? *PeerJ 5*(4): e3208.

Vermuelen, F. (2012). *I Am an Academic Fraud.* Forbes Magazine. Retrieved from http://www.forbes.com/sites/freekvermeulen/2012/09/24/i-am-an-academic-fraud.html [5th of September, 2020].

Vicente-Sáez, R., & Martínez-Fuente, C. (2018). Open science now: A systematic literature review for an integrated definition. *Journal of Business Research 88*(1): 428-436.

Vicente-Saez, R., Gustafsson, R., & Van den Brande, L. (2020). The dawn of an open exploration era: Emergent principles and practices of open science and innovation of university research teams in a digital world. *Technological Forecasting and Social Change 156*(1): 120037.

Vlastos, G. (1985). Socrates' disavowal of knowledge. *The Philosophical Quarterly 35*(138): 1-31.

von Hippel, E. (2017). *Free Innovation.* Cambridge, MA: MIT Press.

Wade, M., & Hulland, J. (2004). The resource-based view and information systems research: Review, extension and suggestions for future research. *MIS Quarterly 28*(1): 107–42.

Wager, E. (2012). The Committee on Publication Ethics (COPE): Objectives and Achievements 1997-2012. *La Presse Medicale 41*(9): 861-866.

Wall, R., & Carey, S. (2017). *How budget carriers transformed the airline industry – in 14 charts.* Wall Street Journal. Retrieved from https://www.wsj.com/articles/how-budget-carriers-transformed-the-airline-industryin-14-charts-1503501624 [12th of December, 2019].

Wallis, J. C., Rolando, E., & Borgman, C. L. (2013). If we share data, will anyone use them? Data sharing and the reuse in the long tail of science and technology. *PLoS ONE 8*(7): e67332.

Walport, M., & Brest, P. (2011). Sharing research data to improve public health. *The Lancet 377*(9765): 537-539.

Walsh, J. P., & Malone, N. G. (2003). *Problems in scientific collaboration: Does email hinder or help?* Working Paper. Tokyo: University of Tokio.

Wang, J. K., Ashleigh, M., & Meyer, E. (2006). Knowledge sharing and team trustworthiness: It's all about social ties!. *Knowledge Management Research & Practice 4*(3): 175-186.

Wang, S., & Noe, R. A. (2010). Knowledge sharing: A review and directions for future research. *Human Resource Management Review 20*(2): 115-131.

Wardle, D. A. (2012). On plummeting manuscript acceptance rates by the main ecological journals and the progress of ecology. *Ideas in Ecology and Evolution 5*(1): 13-15.

Watson, J. D., & Crick, F. H. C. (1953). Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature 171*(4356): 737-738.

Weber-Wulff, D. (2014). *False Feathers – A Perspective on Academic Plagiarism.* Berlin: Springer.

Weinberger, C. J. (2014a). The increasing complementarity between cognitive and social skills. *The Review of Economics and Statistics 96*(5): 849-861.

Weinberger, C. J. (2014b). Are there racial gaps in high school leadership opportunities? Do academics matter more? *The Review of Black Political Economy 41*(4): 393-409.

West, J. (2013). *The De-Habilitation of a Serial Salami Slicer.* Open Innovation Blog. Retrieved from http://blog.openinnovation.net/2013/09/the-de-habilitation-of-serial-salami.html [3rd of September, 2020].

West, J. D., Jacquet, J., King, M. M., Correll, S. J., Bergstrom, C. T. (2013). The role of gender in scholarly authorship. *PLoS ONE 8*(7): e66212.

West, J., Salter, A., Vanhaverbeke, W., & Chesbrough, H. (2014). Open innovation: The next decade. *Research Policy 43*(5): 805-811.

Whitlock, M. C. (2011). Data archiving in ecology and evolution: Best practices. *Trends in Ecology and Evolution 26*(2): 61-65.

Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist 61*(7): 726-728.

Wiese, B. S., & Freund, A. M. (2011). Parents as role models: Parental behavior affects adolescents' plans for work involvement. *International Journal of Behavioral Development 35*(3): 218-224.

Wiese, B. S., & Ritter, J. O. (2012). Timing matters: Length of leave and working mothers' daily reentry regrets. *Developmental Psychology 48*(6): 1797-1807

Wilbanks, J., & Friend, S. H. (2016). First, design for data sharing. *Nature Biotechnology 34*(4): 377-379.

Wilhite, A. W., & Fong, E. A. (2012). Coercive citation in academic publishing. *Science 353*(6068): 542-543.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A. et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data 3*(1): 1-9.

Williams, Richard. (2018). *Analyzing Rare Events with Logistic Regression.* University of Notre Dame Working paper. Retrieved from https://www3.nd.edu/~rwilliam/stats3/RareEvents.pdf [19th of September, 2019].

Williamson, O. E. (1979). Transaction-cost economics: The governance of contractual relations. *The Journal of Law & Economics 22*(2): 233-261.

Wise, L. L. (1985). Project TALENT: Mathematics course participation in the 1960s and its career consequences. In Chipman, S. F., Brush, L. R. & Wilson D. M. (Eds.), *Women and Mathematics: Balancing the Equation* (pp. 25-58). New York: Lawrence Erlbaum Associates.

Wong, A., Wei, L., Yang, J., & Tjosvold, D. (2017). Productivity and participation values for cooperative goals to limit free riding and promote performance in international joint ventures. *Journal of World Business 52*(6): 819-830.

Wood, G. A., & Parr, J. B. (2005). Transaction costs, agglomeration economies, and industrial location. *Growth and Change 36*(1): 1-15.

Wren, J. D., Kozak, K. Z., Johnson, K. R., Deakyne, S. J., Schilling, L. M., & Dellavalle, R. P. (2007). The write position: A survey of perceived contributions to papers based on byline position and number of authors. *EMBO Reports 8*(11): 988-991.

Yamada, Y. (2018). How to crack pre-registration: Toward transparent and open science. *Frontiers in Psychology* Published online first 26th of September, 2018.

Yiamouyiannis, A. & Osborne, B. (2012). Addressing gender inequities in collegiate sport: Examining female leadership representation within NCAA sport governance. *SAGE Open 2*(2): 1-13.

Yitzhaki, M. (1994). Relation of title length of journal articles to number of authors. *Scientometrics 30*(1): 321-332.

Yu, S., Johnson, S., Lai, C., Cricelli, A., & Fleming, L. (2017). Crowdfunding and regional entrepreneurial investment: An application of the CrowdBerkeley database. *Research Policy 46*(10): 1723-1737.

Yukawa, Y., Ktinaka, C., & Yokoyama, M. (2014). Authorship practices in multi-authored papers in the natural sciences at Japanese universities. *International Journal of Japanese Sociology 23*(1): 80-91.

Yukl, G., & Van Fleet, D. D. (1992). Theory and research on leadership in organizations. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of Industrial and Organizational Psychology* (pp. 147-197). Palo Alto: Consulting Psychologists Press.

Yu-Wei, C. (2019). Definition of authorship in social science journals. *Scientometrics 118*: 563-585.

Zeng, X. H. T., Duch, J., Sales-Pardo, M. Moreira, J. A. G., Radicchi, F., Riberio, H. V. et al. (2016). Differences in collaboration patterns across discipline, career stage, and gender. *PLoS Biology 14*(11): e1002573.

Zwart, H. (2017). The catwalk and the mousetrap: Reading Diederik Stapel's Derailment as a misconduct novel. In Zwart, H. (Ed.), *Tales of Research Misconduct.* Cham: Springer.

Declaration

I I hereby declare in lieu of an oath that I have completed the present thesis independently and without illegitimate assistance from third parties. I have used no other than the specified sources.

Vienna, November 8 2020

Pruschak

Gernot Pruschak