



Universiteit
Leiden
The Netherlands

A teacher like me: the role of teacher gender representation and gender stereotypes in education

Doornkamp, L.

Citation

Doornkamp, L. (2023, June 7). *A teacher like me: the role of teacher gender representation and gender stereotypes in education*.

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from:

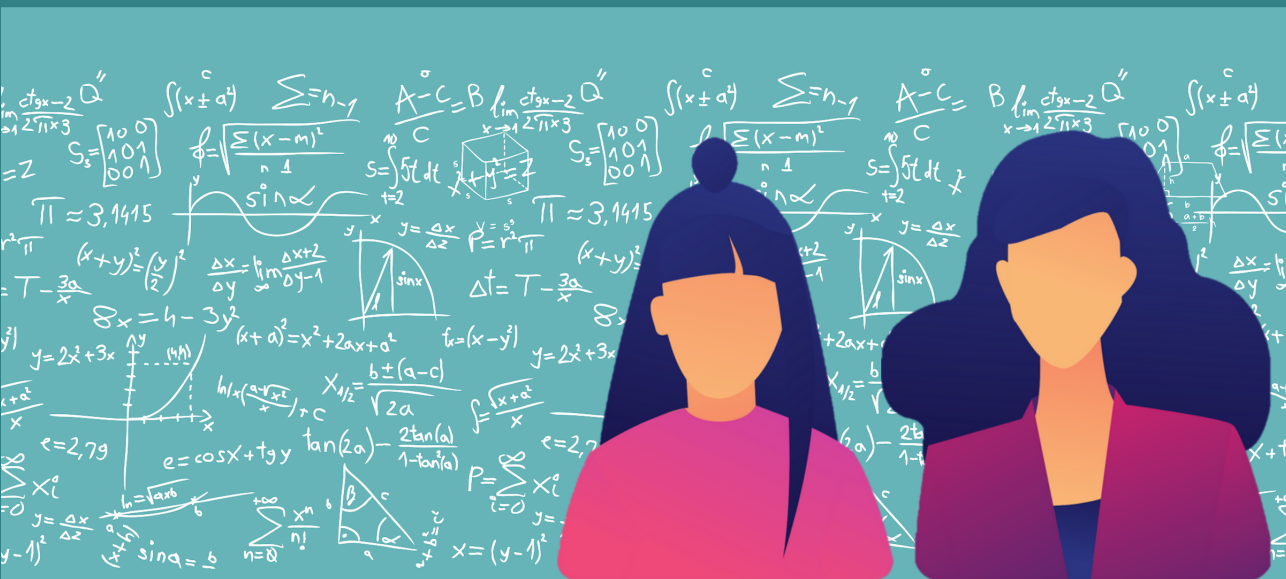
Note: To cite this publication please use the final published version (if applicable).



A TEACHER LIKE ME

The Role of Teacher Gender Representation and Gender Stereotypes in Education

Laura Doornkamp



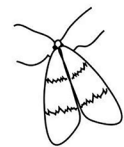


representation
is vital
otherwise the butterfly
surrounde by a group of moths
unable to see itself
will keep trying to become the moth

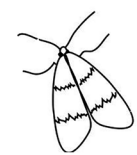
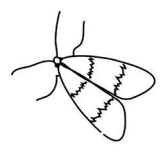
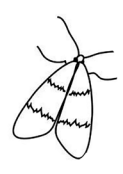
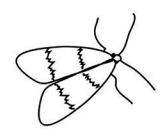
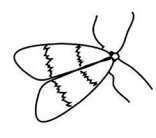
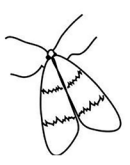
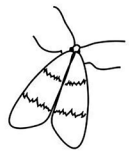
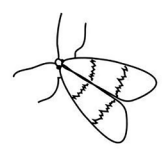
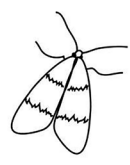
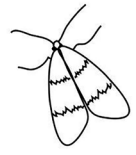
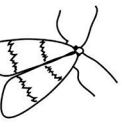


representation - rupi kaur

MO



representation - rupi kaur



A TEACHER LIKE ME

**The Role of Teacher Gender Representation
and Gender Stereotypes in Education**

Laura Doornkamp

Cover

Laura Doornkamp

Lay-out

Ilse Modder – www.ilsemodder.nl

Print

Gildeprint Enschede – www.gildeprint.nl

ISBN

978-94-6419-816-4

©2023 L. Doornkamp, The Netherlands. All rights reserved. No parts of this thesis may be reproduced, stored in a retrieval system or transmitted in any form or by any means without permission of the author. Alle rechten voorbehouden. Niets uit deze uitgave mag worden vermenigvuldigd, in enige vorm of op enige wijze, zonder voorafgaande schriftelijke toestemming van de auteur.

A Teacher Like Me

The Role of Teacher Gender Representation and Gender Stereotypes in Education

PROEFSCHRIFT

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof.dr.ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op 7 juni 2023
klokke 15:00 uur door

Laura Doornkamp
geboren te Groningen in 1994

Promotors:

Prof. dr. Judi Mesman (Leiden University)

Prof. dr. Sandra Groeneveld (Leiden University)

Copromotors:

Dr. Lotte D. van der Pol (Leiden University)

Dagelijks begeleider:

Dr. Marleen G. Groeneveld (Leiden University)

Promotiecommissie:

Prof. Mr. Dr. Erwin R. Muller (Leiden University)

Prof. Dr. Jojanneke van der Toorn (Leiden University)

Prof. Dr. Kenneth J. Meier (American University and Leiden University)

Prof. Dr. Jet Bussemaker (Leiden University)

Prof. Dr. Monique L.L. Volman (University of Amsterdam)

Prof. Dr. Wilfried F. Admiraal (Oslo Metropolitan University)

The research described in Chapters 2, 4, and 5 was supported by the European Research Council (ERC) grant awarded to Judi Mesman (project Girls In Science, 726141).

Paranimfen:

Frank Doornkamp

Joost Doornkamp

DANKWOORD (ACKNOWLEDGEMENTS)

Dit proefschrift is tot stand gekomen door de betrokkenheid van een groot aantal enthousiaste, behulpzame en inspirerende mensen. Die mensen wil ik allemaal bedanken, omdat het schrijven van dit proefschrift zonder hen nooit was gelukt en bovendien nooit zo leuk en leerzaam was geweest!

Allereerst wil ik mijn promotoren Sandra Groeneveld en Judi Mesman bedanken. Sandra, jij was het die opmerkte dat het doen van promotieonderzoek misschien iets voor mij zou zijn. Met die opmerkzaamheid werd het zaadje voor dit proefschrift geplant. De daaropvolgende (voortzetting van de) begeleiding, betrokkenheid en altijd scherpe en prikkelende vragen waren motiverend, inspirerend en vooral heel erg leerzaam. Judi, dankzij jou kon ik ook daadwerkelijk aan dit promotietraject beginnen. Dankjewel voor deze kans en het vertrouwen dat jij me gaf voor de opzet en uitvoering van de schoolstudies op jouw al lopende onderzoeksproject. De dagelijkse begeleiding van Marleen en Lotte, en de samenwerking met mijn medeAIO's op het onderzoeksproject hebben een nieuwe 'gender wereld' voor mij geopend. Marleen en Lotte, dankjulliewel dat jullie er ALTIJD waren voor inhoudelijke vragen, waardevolle brainstorm sessies of gewoon een luchtig gesprek. Antoinette, Astrid, Stefanie en Tessa dankjulliewel voor het warme onthaal in het team en de inspirerende gesprekken over alles wat met gender te maken heeft, ook daar heb ik ontzettend veel van geleerd!

Daarnaast wil ik alle lerarenopleiders, docenten in opleiding, docenten, schoolbestuurders, directeuren en decanen bedanken voor hun interesse in en medewerking aan mijn onderzoeken. In het bijzonder wil ik Arthur Pormes, Anneke Wurth, Peter Kop en Ellen Veenema bedanken. Jullie enthousiasme, interesse, feedback en hulp speelden een belangrijke rol in het volbrengen van mijn studies en het plezier dat ik aan het uitvoeren van onderzoek heb beleefd. Ook wil ik alle gezinnen die hebben deelgenomen bedanken voor hun gastvrijheid en openheid. Verder wil ik alle studenten bedanken voor hun enorme bijdrage aan de dataverzameling en verwerking van de verschillende onderzoeken. In het bijzonder Anne Floor Lubbers: dankjewel voor je enorme inzet en de gezelligheid die je daarbij meebracht.

En dan de mensen die het dichtst bij mij staan. Ik kan niet anders dan beginnen bij mijn broer(tje)s Frank en Joost die mij bij zullen staan tijdens de verdediging van dit proefschrift zoals zij mij altijd bijstaan tijdens alles wat ik meemaak. Dankjulliewel dat jullie er altijd zijn, ik ben super trots op jullie. Een speciaal dankwoord gaat uit naar Frank die het aandurfde om mee te werken op een van de papers in dit proefschrift – het was fantastisch om dat samen te kunnen doen! Mijn ouders wil ik bedanken voor hun onvoorwaardelijke steun en alles wat zij mij meegeven. 'Kan Niet ligt op het kerkhof en Wil Niet ligt ernaast', 'Bek op 't stuur' en 'Nait soez'n, deurbroez'n' zijn een kleine greep uit hun wijsheden die mij door de perikelen van dit promotietraject hebben geholpen. Ook wil ik mijn grootouders, schoonouders en vriendinnen uit Groningen, Leiden/Den Haag en Rotterdam bedanken

voor hun luisterend oor, de benodigde afleiding en aanmoedigingen: ik voel mij gezegend met jullie allemaal om mij heen. En ten slotte Vincent: het is gek om te bedenken dat ik jou leerde kennen net na het begin van mijn promotietraject – we kennen elkaar dus nog niet eens heel lang. Toch hebben we in korte tijd de meest fantastische avonturen beleefd en samen ons thuis opgebouwd. Wat is het heerlijk om mijn leven met jou te delen. Ik kijk uit naar alles wat er nog voor ons in het verschiet ligt!

TABLE OF CONTENTS

PREFACE: A brief discussion of positionality and use of terminology	16
Positionality	16
Use of Terminology	16
CHAPTER ONE. General Introduction	19
1.1 The Role of Gender and Gender Stereotypes in Education	21
1.2 Student-Teacher Gender Congruence and Student performance	24
1.3 When and Where: The Role of Context	25
1.4 How: Individual-Level Mechanisms Including Gender Stereotypes	27
1.4.1 The Student Perspective	27
1.4.2 The Teacher Perspective	29
1.5 Method	30
1.6 Aim and Outline of the Dissertation	31
CHAPTER TWO. Student-Teacher Gender Congruence and Student Performance: the Role of Context	35
2.1 Introduction	37
2.2 Student-Teacher Gender Congruence and Students' Performance	38
2.2.1 The Role of Context	39
2.3 Method	42
2.3.1 Procedure	42
2.3.2 Sample	43
2.3.3 Measures	43
2.3.4 Data Analyses	44
2.4 Results	46
2.4.1 Student-Teacher Gender Congruence and Students' Performance	46
2.4.2 The Role of Context	46
2.5 Discussion and Conclusion	51
2.5.1 Gender Congruence and Performance	51
2.5.2 The Role of Context	52
2.5.3 Limitations	53
2.5.4 Conclusion	54
2.6 Appendix	56
2.6.1 Appendix A – Robustness Check	56

CHAPTER THREE. The individual level effect of symbolic representation: An experimental study on teacher-student gender congruence and students' perceived abilities in math	59
3.1 Introduction	61
3.2 Representative Bureaucracy and Performance in Education	62
3.2.1 Symbolic Representation, Stereotypes, and Academic Self-Concepts	62
3.3 Method	64
3.3.1 Measures	65
3.4 Results	67
3.5 Discussion and Conclusion	71
3.6 Notes	73
3.7 Appendix	74
3.7.1 Appendix A - Operationalization of the Variables	74
3.7.2 Appendix B - Survey	76
 CHAPTER FOUR. Understanding the Symbolic Effects of Gender Representation: A Multi-Source Study in Education	 81
4.1 Introduction	83
4.2 Representative Bureaucracy	84
4.2.1 Social Cognitive Processes as Mechanism Underlying Symbolic Representation Effects	86
4.2.1.1 Self-Perceived Abilities	87
4.2.1.2 Gender-Stereotypical Beliefs	88
4.3 Study Design and Method	89
4.3.1 Study Context	89
4.3.2 Data Collection	89
4.3.3 Measures	91
4.3.4 Analytical Strategy	93
4.4 Results	95
4.5 Discussion and Conclusion	100
4.6 Notes	104
4.7 Acknowledgements	105
 CHAPTER FIVE. Understanding Gender Bias in Teachers' Grading: The Role of Gender Stereotypical Beliefs	 107
5.1 Introduction	109
5.2 Gender Grading Bias	109
5.2.1 The Role of Teachers' Gender and Gender Stereotypes in Gender Grading Bias	111
5.2.2 Teachers' Gender, Gender Stereotypes, and Gender Grading Bias	114

5.3	Method	115
5.3.1	Sample	115
5.3.2	Study Design	116
5.3.3	Measures	118
5.3.4	Data Analyses	119
5.4	Results	120
5.4.1	Gender Grading Bias on the Aggregate Level	120
5.4.2	Teachers' Gender and Gender Grading Bias	121
5.4.3	Teachers' Implicit Gender Stereotypes and Gender Grading Bias	121
5.4.4	The Indirect Relation between Gender and Gender Grading Bias	124
5.5	Conclusion and Discussion	126
5.6	Acknowledgements	131
5.7	Appendix	132
5.6.1	Appendix A – Manipulation	132
5.6.2	Appendix B – Bias Against Male or Female Students	133
 CHAPTER SIX. Conclusion		 137
6.1	The Role of Gender and Gender Stereotypes in Education	138
6.2	Summary of Findings	138
6.2.1	When and Where: The Role of Context	138
6.2.2	How: The Student Perspective	139
6.2.3	How: The Teacher Perspective	140
6.3	Lessons Learned	140
6.4	Theoretical and Methodological Implications	141
6.4.1	Representative Bureaucracy	141
6.4.2	The Study of Gender and Gender Stereotypes in Education	143
6.4.3	Methodological Implications	144
6.5	Limitations	145
6.6	Conclusion	146
 References		 149
 Dutch Summary: Nederlandse Samenvatting		 169
 About the Author: Curriculum Vitae		 177

PREFACE: A BRIEF DISCUSSION OF POSITIONALITY AND USE OF TERMINOLOGY

POSITIONALITY

I would like to reflect on my positionality as a researcher as it may have influenced the research questions that were formulated, the methods that were chosen, and the interpretation of the results. I graduated in Public Administration and I attained a teaching qualification for secondary education in the Netherlands. I have always had a fascination for the education system, because of its important role in the public sector, in society, and in every individual's life. During this PhD trajectory I immersed myself into social psychological theories to gain a better understanding of the role of gender stereotypes in education. The fact that I was part of an interdisciplinary research team helped me in this process.

Further, the focus on gender and gender stereotypes in this dissertation is no coincidence. I am a cisgender female and I was born in a village in the north of the Netherlands (rural area) where I grew up with a mother, father and two younger brothers. Gender equality was an important factor in my upbringing. My parents taught me that my brothers and I, and men and women in general, are capable and worthy of the same things in life. At the same time, I experienced differential treatment based on my gender in other contexts, for example at school. My upbringing and my experiences outside the family, as well as the different treatment of men and women in society in general, have led to me attaching great importance to gender equality and that I want to contribute to the reduction of social and psychological barriers that stem from gender stereotypes. As a result, the studies described in this dissertation were partially motivated by hopefulness to increase awareness, and possibly contribute to the reduction, of gender bias against both female and male students in Dutch secondary education, in Dutch society, and beyond. This motivation may have led to blind spots or bias in the design of the studies in this dissertation or the interpretation of the results. In order to prevent and address blind spots and bias as much as possible, I had discussions with the members of the interdisciplinary research team, but also with others, including research practitioners, statisticians, reviewers, and editors, on themes such as the consequences of the binarity and non-binarity use of the concept of gender, the validity of indicators of socioeconomic status, and the classification of rural areas in the Netherlands.

USE OF TERMINOLOGY

The concept *gender* and the role of gender stereotypes are central to this dissertation. In this dissertation I use a binary approach of the concept gender although I am aware that gender entails more than the solely male or female classification. Despite social developments towards inclusiveness and diversity when it comes to gender identities, societies are still largely organized in a binary way categorizing humans as either male or female. This binary categorization lies at the basis of gender stereotypes and the accompanying gender bias and discriminative behaviors in everyday life (Hyde et al., 2019). This study focusses on

these processes for students and teachers and therefore uses the gender binary approach.

Further, this dissertation consists of four studies that were conducted within an interdisciplinary collaborative research project. I came up with the research ideas and I led the recruitment, data collection, and data processing for three of the four studies (Chapter 2, 3 and 5). Furthermore, I was the leading author of each of the scientific papers presented in this dissertation. The research design, data analyses, and interpretation of the results were a collaborative effort with the interdisciplinary research team. In this dissertation, I therefore use the plural first-person pronouns (e.g., 'we') to refer to the empirical test of the hypotheses and the findings with regard to these hypotheses (e.g., 'we tested, we found'). However, when it concerns the experiences, choices, and ideas in this dissertation I use the singular first-person pronouns (e.g., 'I suggested').



CHAPTER **ONE**

General introduction

In many Western countries like the Netherlands children are obliged to attain primary and secondary education from the age of five until the age of sixteen (Rijksoverheid, 2022). As a result, people spend a significant part of their lives in schools interacting with teachers. Teachers can therefore play an important role in the development of an individual's interests, self-beliefs, and performances in school and further life (Fredriksen & Rhodes, 2004). Indeed, everybody may remember a teacher that played a decisive role in their development. For me, one of my teachers later turned out to be an important inspiration for this research. One incident in particular showed me the possible consequences of gender and gender stereotypes in education: In secondary education during an ordinary math class, I was discussing my plans for the weekend with two of my male peers instead of working on my math homework. Our math teacher corrected us and told us to pay attention to our homework. Then he explained why. He turned to my male peers and explained that they should pay attention so they could achieve higher grades in math. He turned to me and explained that I should pay attention so I would not fail the class. However, until that moment, I had achieved the same grades in math as my male peers. The gendered message of my math teacher contributed to the feeling that I could not excel in math, and I believed that for a long time, until I started with statistics in my masters under the supervision of a female professor.

1.1 THE ROLE OF GENDER AND GENDER STEREOTYPES IN EDUCATION

Like many Western societies, Dutch society is characterized by gender role segregation; men and women fulfil different roles within society (Daalmans et al., 2017). For instance, in families, men often fulfil the role of breadwinner, while women work part time and take on most of the care of the children (OESO, 2019). Further, the educational and professional pathways of men and women are segregated in which men are overrepresented in the Science, Technology, Engineering, and Math (STEM) sector while women are overrepresented in languages, early education, and the health care sector (Onderwijsraad, 2019).

This gender role segregation corresponds with and reinforces gender stereotypes that describe men as being naturally talented at leadership roles and STEM skills and women as being naturally talented at caretaking and languages/communication skills. Gender stereotypes are socially constructed ideas that describe what men and women are like and prescribe what men and women should be like regarding their appearances, capabilities, interests, and behaviors (Ellemers, 2018). Gender stereotypes are transmitted to individuals from a very young age and gender stereotypes evolve during one's lifetime (Bem, 1981). Following social cognitive theory, individuals' gender stereotypes develop through direct information, social consequences to gender-typed behaviors, and most importantly through observing the behaviors and roles of men and women in one's environment (Bussey & Bandura, 1999). Men and women that confirm stereotypical gender roles can strengthen individuals' gender stereotypes whereas men and women that counteract stereotypical gender roles can weaken individuals' gender stereotypes. Individuals' gender stereotypes can influence interests, beliefs, and behaviors of oneself and others (Bem, 1981).

The psychological processes of transmission of gender stereotypes and the impact of gender stereotypes also take place in schools (e.g., Maries et al., 2020; Muntoni & Retelsdorf, 2018; Thompson, 2003). Students' and teachers' gender, among others, may play an important role in these processes (Jungbluth, 1982). In particular in subjects with gender stereotypes connected to them, students' gender stereotypes may affect students' interests, self-beliefs, and performances (e.g., Brown & Stone, 2016). In these subjects, the mere presence of a male or female teacher may play an important role in students' gender stereotypes. For instance, the gender of a math teacher may confirm or contradict the gender stereotype that men are good at math and women are bad at math (Ten Dam & Volman, 1991). The mere presence of a female math teacher may weaken students' gender stereotype about math, while the mere presence of a male math teacher may strengthen students' gender stereotype. Further, the gender of a teacher may influence the strength of teachers' own gender stereotypes (Martin & Dinella, 2012; Smeding, 2012). A female math teacher herself may have weaker gender stereotypes because of her personal experience of becoming a math teacher as a woman.

Negative gender stereotypes (i.e., men/women are bad at...) are generally easier to internalize and more resistant to change than positive gender stereotypes (i.e., men/

women are good at...) (Baumeister et al., 2001). The role of gender and gender stereotypes may- therefore be more important for students and teachers in subjects in which they are confronted with negative gender stereotypes. Hence, the role of teachers' gender and gender stereotypes can be important in STEM subjects for female students and teachers because these are the subjects in which women are stereotypically thought to have performance deficits. The role of gender and gender stereotypes can be important in language subjects for male students and teachers because these are the subjects in which men are stereotypically thought to have performance deficits. As a consequence, student-teacher gender congruence (i.e., student and teacher having the same gender) might play an important role in STEM for females, and in languages for males.

The consequences of students' and teachers' gender as well as STEM/languages gender stereotypes are demonstrated in previous research. A large body of research in different scientific disciplines demonstrated that student-teacher gender congruence can positively affect students' performance (e.g., Dee, 2005). In addition, students with relatively strong gender stereotypes were found to perform worse than students with relatively weak gender stereotypes (e.g., Maries et al., 2018). Further, teachers' gender stereotypes were found to translate into different expectations of male and female students which in turn were related to students' performance (Muntoni & Retelsdorf, 2018).

This body of evidence led previous research to suggest that effects of student-teacher gender congruence on students' performance can be explained through the role of gender stereotypes. However, empirical tests of these mechanisms are rare and solely situated in the 'females in math' context (Xu & Meier; 2021; Zhang, 2019). As a result, still little is known about *where and when* effects of student-teacher gender congruence occur, and there are still important gaps in our understanding of *how* effects of student and teacher gender emerge (Volman, Ten Dam, & Van Eck, 1995; Ten Dam, Van Eck, & Volman, 1997). In response, the main aim of this dissertation is to receive a better understanding of the role of students' and teachers' gender and gender stereotypes in student performance in order to contribute to the scientific, as well as the public, debate on the role of gender in education. The main research question is:

Does student and teacher gender affect students' performance and what is the role of gender stereotypes in this relation?

To increase the understanding of the role of gender and gender stereotypes in student performance, I first explore the relation between student-teacher gender congruence and students' performance in different subjects across educational contexts that are assumed to differ in the extent to which gender stereotypes play a role. In doing so, I try to specify *when and where* student-teacher gender congruence is associated with student performance. Subsequently, I zoom in on the individual level and integrate theories from different disciplines to propose mechanisms including gender stereotypes that may underly relations

between student and teacher gender and student performance. I do so by focusing on psychological processes in students and teachers. From the student perspective and the teacher perspective, I propose mechanisms that explain the theorized influence of the mere presence of a male or female teacher (student perspective) or student (teacher perspective) through the role of gender stereotypes in (un)conscious psychological processes. In this way, I aim to add to our understanding of *how* effects of the gender of a teacher emerge. This study's research angles are visualized in Figure 1.1.

The remainder of this chapter has five distinct sections. First, Section 1.2 provides an overview of studies on representative bureaucracy in public administration as a starting point for building a framework for understanding the relation between student-teacher gender congruence and student performance. In Section 1.3, I explore the role of educational context. I theorize in which educational contexts student-teacher gender congruence can affect students' performance. Thereafter, in Section 1.4, I shift to social psychological theories and propose mechanisms that explain effects of students' and teachers' gender on student performance from both the student and the teacher perspective. Section 1.5 describes the research setting and methods of this study. Finally, in Section 1.6 the aim and outline of this dissertation is discussed.

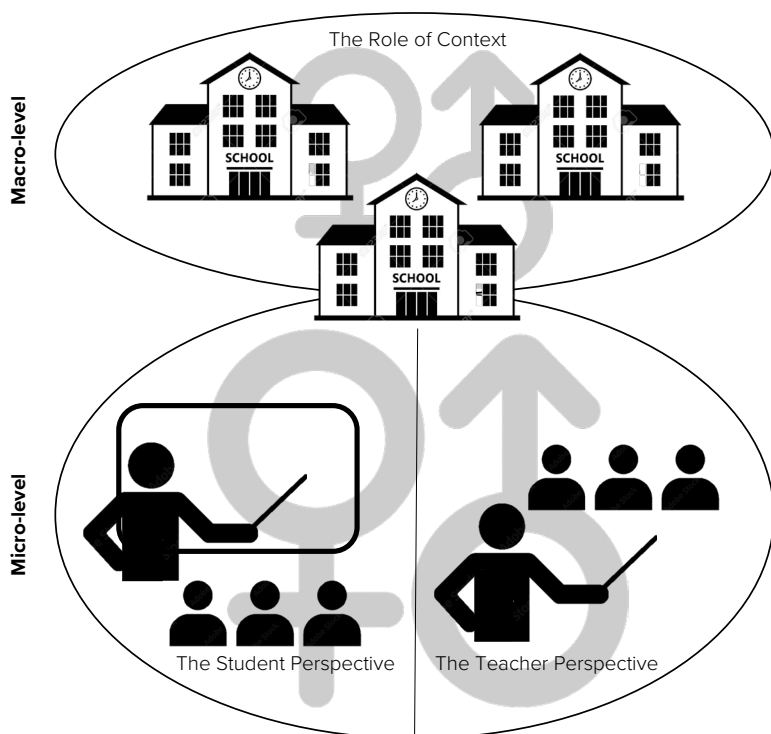


Figure 1.1. Research Angles

1.2 STUDENT-TEACHER GENDER CONGRUENCE AND STUDENT PERFORMANCE

The association between the gender of a teacher and male and female students' performances has been demonstrated by a large body of research in different scientific disciplines (e.g., Holmlund & Sund, 2008; Winters et al., 2013). In particular the match between teachers' gender and students' gender (i.e., student-teacher gender congruence) has been suggested and found to positively affect students' performance (e.g., Dee, 2005; Song 2018). Despite the large number of studies on the topic and gender theories on the general role of gender in education (e.g., Thompson, 2003), a gender theory that focuses on the effects of student-teacher gender congruence specifically does not exist. However, a substantial part of the studies on the relation between student-teacher gender congruence and student performance stem from research within the field of public administration that focus on representative bureaucracy. This stream in the literature focuses on effects of gender representation within the government and often situates empirical tests in the educational setting. Consequently, representative bureaucracy literature can be a suitable framework to start theorizing the influence of student and teacher gender on student performance.

Scholars within representative bureaucracy literature argue that the demographic composition of public organizations should reflect the demographic composition of the people they serve for democratic reasons. By representing different groups in the population, public organizations show equal access to administrative positions and create more connection between citizens and the government (Sowa & Selden, 2003). This results in more legitimacy and credibility for public organizations and their policies (Sowa & Selden, 2003). Apart from this normative argument, a large body of research found that public organizations with high levels of representativeness perform better than their less-representative counterparts (for an overview see Meier, 2019). The education system is a large and important part of the public sector (Meier & Stewart, 1992). This is one of the reasons why empirical tests of representative bureaucracy are often situated within the school context. A great deal of this research focuses on the role of gender representation in female students' math performance (e.g., Keiser et al., 2002). Studies in western countries (e.g., Keiser et al., 2002), Asian countries (e.g., Dhillon & Meier, 2020), and African countries (Agyapong, 2018) found that female students' math performance improved when the share of female teachers in the subject increased.

These positive associations between gender representation and female students performance are attributed to effects of passive representation, active representation, or symbolic representation. Passive representation refers to the composition of a workforce and suggests that the presence of representatives improves awareness among all members of the workforce on issues that affect outcomes for the represented. As a result of this awareness, behaviors of all members of the workforce are suggested to change and to translate into beneficial outcomes for the represented (Groeneveld & Meier, 2022). Following

this reasoning, the presence of female math teachers improves awareness and changes all (female and male) math teachers' behaviors that affect female students' performance (e.g., teaching approaches that benefit female students' performance). Active representation concerns the behaviors of bureaucrats, i.e., teachers, that directly translate into benefits for represented students (Grissom et al., 2015). For instance, female math teachers may spend more time on instructing female students or female teachers may give more positive feedback to female students. These active representative behaviors may directly translate into better performance of female students. Symbolic representation refers to the changes in attitudes, beliefs, and behaviors in the student as a result of the mere presence of a representative teacher. Female students may feel more confident in their ability in math as a result of being taught by female math teacher. This in turn may improve female students' math performance (An et al., 2021; Beilock et al., 2010; Xu & Meier, 2021; Zhang, 2019).

Despite the increasing number of studies that focus on these individual level effects of representation in education, the suggested mechanisms underlying active and symbolic representation are rarely theorized and empirically examined (Xu & Meier, 2021; Zhang, 2019). Consequently, still little is known about the explanatory mechanisms underlying gender representation including the working of gender stereotypes. Although scholars recognized the role of gender stereotypes (e.g., Riccucci & Van Ryzin, 2017), representative bureaucracy scholars made limited use of social psychological theories on this topic to unravel the working of gender representation through the role of gender stereotypes. Further, despite research on organizational (e.g., Song, 2018) and national (An et al., 2021) contexts that may facilitate or impede effects of gender representation, still little is known about the role of contextual factors in the vicinity of a public organization (Ding et al., 2021). In particular for schools, besides national contextual factors, contextual factors in the vicinity of the organization may play an important role in representation.

In the following paragraphs, these gaps in the literature are assessed using insights from social psychological theories on gender stereotypes. First, the role of context in the association between student-teacher gender congruence and student performance will be discussed. Thereafter, possible mechanisms underlying effects of representation are explicated.

1.3 WHEN AND WHERE: THE ROLE OF CONTEXT

An important condition for effects of gender representation to occur is gender salience (e.g., Keiser et al., 2002; Ding et al., 2022). Gender salience refers to the extent to which gender is perceived as an important characteristic. When gender is perceived as an important characteristic, a gender-related self-concept is activated which allows gender to have an influence (Palan, 2001). The rigidity of gender stereotypes in given contexts might contribute to the extent to which gender is perceived as an important characteristic. When gender stereotypes in a context are rigid, gender becomes an important characteristic

for individuals' identity, interactions, and performances (Hilliard & Liben, 2010; Kiefer & Sekaquaptewa, 2007; Shih et al., 1999).

As described above, the capabilities of men and women are gender stereotyped as fundamentally different in STEM and language subjects (i.e., men are good at STEM and bad at languages while women are good at languages and bad at STEM) (Arnot et al., 1998; Guimond & Roussel, 2000). Because negative gender stereotypes (i.e., bad at) are generally easier to internalize and more resistant to change than positive gender stereotypes (Baumeister et al., 2001), gender may become a salient characteristic for the group that is stereotypically thought to have performance deficits in the subject (Beilock et al., 2010; Spencer et al., 2016). Hence, I expect student-teacher gender congruence to be associated with female students' performance in STEM and for male students' performance in languages. In the first study, we test this hypothesis led by the research question:

To what extent is student-teacher gender congruence associated with male and female students' performance in math, physics, Dutch language and French language?

In addition to the type of subject, the extent to which gender stereotypes connected to subjects play a role in effects of gender in education may differ across different educational contexts (Arnot et al., 1998). Previous research indicates that gender stereotypes might be more rigid among students in lower educational levels (Turner et al., 2019), in religious schools (Schulze & Tomal, 2005), and in schools located in less-populated areas (Garcia-Retamero et al., 2011). Research on socioeconomic status showed that gender stereotypes were stronger among lower socioeconomic groups (Turner et al., 2019). As educational level and socioeconomic status are closely related concepts, the same might be true for students in lower educational levels. Further, gender stereotypes might be stronger in religious environments and less-populated areas because these are environments in which people tend to have more conservative and traditional views on roles of men and women (De Vries et al., 2022; Istenič, 2007). In these contexts, in which gender stereotypes are assumed to be relatively strong, the associations between student-teacher gender congruence and student performance might be strengthened. The second research question in the first study is therefore:

What is the role of students' educational level, schools' religiousness, and schools' location in the relation between student-teacher gender congruence and students' performance?

1.4 HOW: INDIVIDUAL-LEVEL MECHANISMS INCLUDING GENDER STEREOTYPES

As mentioned before, gender stereotypes are socially constructed ideas about what is typical and appropriate for men and women (Ellemers, 2018). Gender stereotypes affect individuals' attitudes, beliefs, interests, and behaviors of oneself and others through attempts to adhere to these ideas about men and women (Bem, 1981; Bussey & Bandura, 1999). The attempts to adhere to gender stereotypes (i.e., the processes through which gender stereotypes are transmitted and translated into attitudes, beliefs, interests, and behaviors) manifest on an implicit (unconscious) and an explicit (conscious) level. The implicit level refers to the automatic and less controllable distinction between men and women, whereas the explicit level refers to the purposeful and controllable distinguishment between what is typical of and appropriate for men and women (Fazio & Olson, 2003). Implicit and explicit gender stereotypes are only weakly correlated, in part because of explicit measures' sensitivity to social desirability or lack of insight in one's own stereotypes (Fazio & Olson, 2003). For instance, teachers can report that they do not differentiate between male and female students, whereas the observation of implicit gender messages in the classroom can prove otherwise (Jones & Myhill, 2004; Nürnberger et al., 2016).

A large body of research demonstrated that gender stereotypes can affect students' performance through psychological processes in students (e.g., Heyder & Kessels, 2017) or through psychological processes in teachers (e.g., Alan et al., 2018). As a result, gender stereotypes are often assumed to underly effects of student-teacher gender congruence on student performance but only a handful of studies empirically tested such mechanisms (Xu & Meier, 2021; Zhang, 2019). In a qualitative study, Xu and Meier (2021) tested whether female math teachers actively try to reduce gender stereotypes in female students, but they did not find evidence for this hypothesis. Zhang (2019) tested whether students' explicit gender stereotypes mediate the positive relation between student-teacher gender congruence and female students' math performance. Although the direct effect of student-teacher gender congruence was much larger, he found a small significant indirect effect through students' gender stereotypes (Zhang, 2019). Important gaps in the understanding of the role of gender stereotypes remain, including questions about the psychological processes in which explicit and implicit gender stereotypes are manifested. In the following paragraphs, I propose psychological mechanisms in students and teachers that may explain the role of student and teacher gender on student performance through the role of gender stereotypes.

1.4.1 THE STUDENT PERSPECTIVE

In their social cognitive theory, Bussey and Bandura (1999) argue that gender stereotypes develop through the observation of and in the interaction with one's environment. Three modes of influence are distinguished: direct information, social consequences to gender

related behaviors, and modelling (Bussey & Bandura, 1999). Direct information refers to learning from people, books, or tv series that communicate explicit, but mostly implicit statements about what is appropriate for men and women. Social consequences to gender related behaviors refer to learning from the reaction of the environment to behaviors that are seen as typically masculine or typically feminine. Behavior of men and women that is congruent with gender stereotypes is likely to be rewarded with attention and positive feedback, while behavior that is incongruent with gender stereotypes is likely to be penalized through lack of attention and negative feedback. The last mode of influence, modelling, is assumed to be the most persuasive form of transmission of gender-linked information (Bussey & Bandura, 1999). Modeling refers to the observation of men and women in the environment of an individual, including teachers in schools. Through observing gender roles and through observing behaviors of men and women, individuals learn about underlying gender norms and structures that describe what is typical and appropriate for men and women. Moreover, the more frequent an individual is confronted with models that confirm gender stereotypes, the stronger the gender stereotypical beliefs in an individual can become. The mere presence of male math teachers can strengthen students' gender stereotypical beliefs that describe math as a masculine subject. At the same time, the mere presence of female math teachers can contradict the gender stereotype weakening students' gender stereotypical beliefs.

Consequently, students' gender stereotypes can translate into students' performance. A large body of research demonstrated the association between gender stereotypes and performance (e.g., Heyder & Kessels, 2013). An often-cited study of Beilock et al. (2010) demonstrated that the presence of female teachers that are anxious of giving math negatively affected female students' math performance, but only of those female students who believed the gender stereotype that women are bad at math. Further, studies on stereotype threat provide convincing evidence that the activation of gender stereotypes can lead to performance deficits (e.g., Walton & Spencer, 2009).

Bussey and Bandura (1999) emphasized that gender stereotypes affect performance through mechanisms including self-efficacy beliefs. The internalization of gender stereotypes can lead to self-efficacy beliefs that adhere to perceived social gender norms. For instance, a female student with strong gender stereotypes may believe that she is bad at math because she is a woman. Being a woman in combination with the believe in the gender stereotype affects her self-efficacy in math. Self-efficacy beliefs include students' self-perceived abilities and academic self-concept which are found to influence students' performance (e.g., Chamorro-Premuzic et al. 2010; Greven et al. 2009; Marsh et al., 1988; Wach et al. 2015). Based on the above, I propose a mechanism that explains the relation between student-teacher gender congruence and student performance through students' gender stereotypes and self-efficacy beliefs. The mechanism is visualized in Figure 1.2.

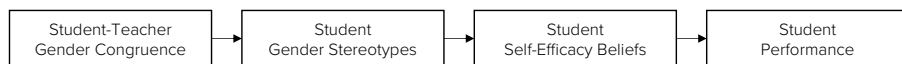


Figure 1.2. The Effect of Student-Teacher Gender Congruence on Student Performance

We empirically test the proposed mechanism in the second and third study of this dissertation. The second study focuses on the first part of the mechanism and examines the influence of student-teacher gender congruence on students' academic self-concept through students' explicit gender stereotypes led by the research question:

How does teacher-student gender congruence influence the academic self-concepts of female and male students?

The third study builds on the second study and examines the role of student-teacher gender congruence in student performance through students' implicit gender stereotypes and students' self-perceived efficacy. The research question is formulated as:

What psychological processes in students can explain the association between student–teacher gender congruence and student performance?

1.4.2 THE TEACHER PERSPECTIVE

Gender stereotypes can translate into gender biased expectations and behaviors towards other men and women. Gender schema theory explains that individuals' behavior towards men and women is guided by their gender schema's that consist of gender-typed information and experiences (Bem, 1981). Individuals with strong gender schema's make a strong distinction between what is typical and appropriate behavior for men and women (i.e., men and women are seen as fundamentally different). These gender schema's influence individuals' perception of men and women and guide their expectations of and behavior towards them.

In schools, teachers with strong gender schema's are likely to have different expectations of and behaviors towards male and female students. These different expectations and behaviors can affect student performance through differential attention, feedback, or grading of male and female students (Muntoni & Retelsdorf, 2018). Indeed, research on gender grading bias suggested that gender stereotypes can affect gender bias in teachers' grading (e.g., Protivínský & Münich, 2018).

As mentioned before, teachers' gender-related experiences can affect the strength of teachers' gender stereotypes. Female teachers in STEM and male teachers in languages may have weaker gender stereotypical beliefs than their counterparts because of their personal experience that contradicts the gender stereotypes. As a consequence, these teachers may have more equal expectations of male and female students which in turn may

result in more equal grading practices (i.e., less gender grading bias). Based on the above I propose a mechanism that explains the difference between male and female teachers' grading bias against male or female students through teachers' implicit gender stereotypes and gender-typed expectations. The mechanism is visualized in Figure 1.3. In the fourth and last study, we empirically test this mechanism led by the research question:

To what extent is teachers' grading gender biased and what is the role of teachers' gender and teachers' implicit gender stereotypes in gender grading bias?

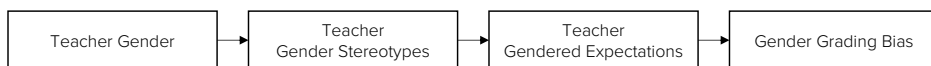


Figure 1.3. The Effect of Teachers' Gender on Teachers' Grading of Male and Female Students

1.5 METHOD

By means of four empirical studies with different methodology, this dissertation examines the relations between student and teacher gender, gender stereotypes, and student performance. Table 1.1 provides an overview of the method, population, subject, period of data collection, and sample size that were used in each study. In the first study, we collected administrative data of secondary schools. The use of administrative data resulted in a large and representative sample of Dutch secondary school students and their teachers. The data allowed us to test the association between student-teacher gender congruence and students' math, physics, Dutch language, and French language performance. Additional school data enabled us to explore the role of educational contexts in these effects.

Table 1.1. Research methods

	Method	Population	Subject	Period data collection	Sample size
Study 1	Secondary data analyses	Administrative data secondary schools	Math, Physics, Dutch language, French language	2020-2021	N > 50,000
Study 2	Vignette Experiment	Secondary school students	Math	2017	N = 410
Study 3	Longitudinal study	Secondary school students	Math, Dutch language	2018-2022	N = 329
Study 4	Field Experiment	Teachers in training	Math, Dutch language	2020-2021	N = 358

In the second study, we zoomed in on an individual-level mechanism from a student perspective. By means of a vignette study among students in a Dutch high school, we examined the influence of the mere presence of a male or female math teacher on both male and female students' gender stereotypes and academic self-concept in math. After being exposed to a gender manipulation in a short text, we measured students' explicit gender stereotypical beliefs and academic self-concept. In doing so, we were able to test causality in the relation between student-teacher gender congruence, students' gender stereotypes, and academic self-concept.

In the third study, we extended the examination of the individual-level mechanism from the student perspective. We used multisource data of a wave in a longitudinal research project to test the associations between student-teacher gender congruence, students' implicit gender stereotypes, self-efficacy beliefs, and performance in math and Dutch language. During family home visits, student data were collected by means of a combination of interviews, questionnaires, and an implicit association test (IAT) resulting in real-life and individual level data.

In the fourth study, we examined the influence of students' gender on male and female teachers' grading in math and Dutch language and we tested the role of teachers' implicit gender stereotypes and gendered expectations in the relation. We did so by means of a field experiment among teachers in training. Teachers in training were asked to grade real-student answers and were asked to complete a short questionnaire with implicit measures including an IAT. These components enabled us to minimize social desirability and conscious behaviors and to grasp real gender biases in teachers' associations, expectations, and behaviors.

1.6 AIM AND OUTLINE OF THE DISSERTATION

The general aim of this dissertation is to provide more insight into the role of student and teacher gender and gender stereotypes in male and female students' performance. More specifically, I aim to contribute to literature on representative bureaucracy in public administration by complementing representative bureaucracy literature with elements of social psychological theories on gender stereotypes to unravel distinct mechanisms of representation. In doing so, I provide new insights about the role of context in representation effects and about the individual-level mechanisms underlying representation effects. Additionally, I aim to contribute to the study of gender and gender stereotypes in education more broadly by integrating different streams in the literature to build a theoretical framework on where, when, and how effects of student and teacher gender on student performance emerge. Further, I show the value and suitability of a variety in research angles, methods, and populations to study the role of gender in education. Finally, I aim to contribute to education practice by providing more insight into the often unconscious and unintentional

role of gender and gender stereotypes in education. Thereby, I try to help practitioners to consider how to reduce negative consequences of gender and gender stereotypes in education.

The studies within this dissertation are divided into chapters. Chapter 2 includes the study on the role of educational context in the associations between student-teacher gender congruence and student performance. Chapters 3 and 4 include the studies on an individual-level mechanism from the student perspective. In these chapters, the role student-teacher gender congruence in students' gender stereotypes, self-perceived abilities, and performance is examined. Chapter 5 includes the study of an individual-level mechanism from the teacher perspective. In this study the role of teachers' gender, implicit gender stereotypes, and gender-typed expectations of student performance in gender bias in teachers' grading is investigated. The research model is depicted in Figure 1.4. This dissertation concludes with Chapter 6 in which the main findings of the studies are integrated and discussed.

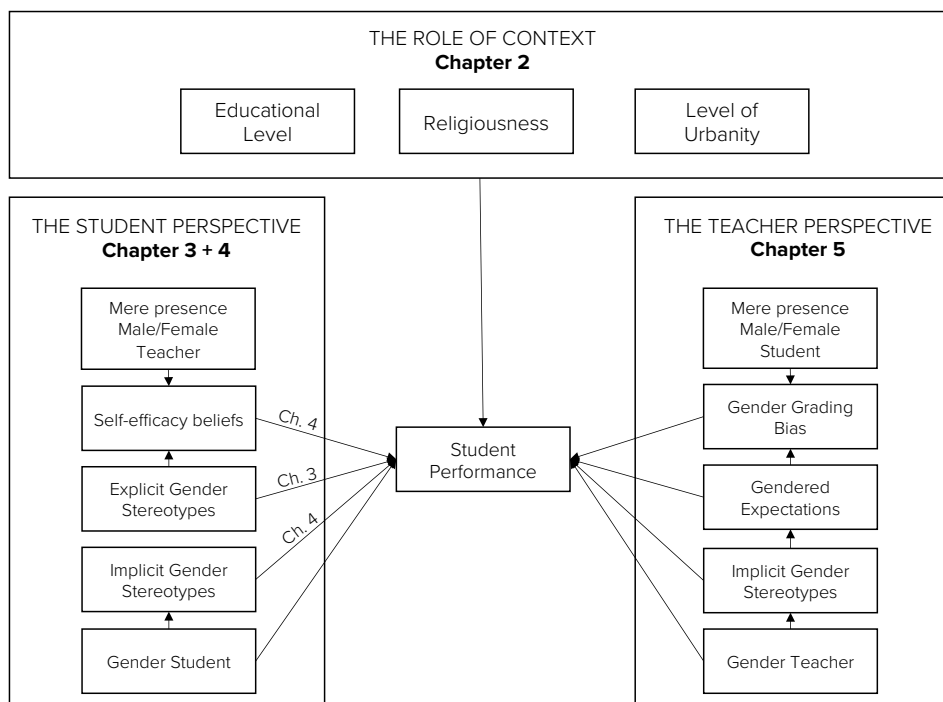


Figure 1.4. Research Model



CHAPTER TWO

Student-teacher gender congruence and student performance: The role of context

Currently under review at an international peer-reviewed journal as:

Doornkamp, L., Doornkamp F., Van der Pol, L.D., Groeneveld, S., Mesman, J., & Groeneveld., M. 'Student-Teacher Gender Congruence and Student Performance: the Role of Context.'

ABSTRACT

Student-teacher gender congruence is suggested to be related to increased student performance, but little is known about the contexts in which these effects occur. Using administrative data of secondary schools in The Netherlands ($N > 50,000$), this study examined the relation between student-teacher gender congruence and student performance in four subjects and explored the role of educational level, religiousness, and location. We found that gender congruence was positively related to female students' performance in math and physics and to male students' performance in Dutch language and French language. The role of educational context differed for male and female students across subjects encouraging future research to use qualitative methods to examine how context influences the role of gender in education.

2.1 INTRODUCTION

Gender gaps in students' performance are often linked to teacher characteristics such as teachers' gender. For instance, female students' math performance is found to be related to the presence of female math teachers (e.g., Marx & Roman, 2002), and the deteriorated relative performance of male students is suggested to be linked to the increasing feminization of the education system (Holmlund & Sund, 2008; Neugebauer et al., 2011). Congruence between teachers' gender and students' gender is assumed to be positively associated with student performance, but empirical evidence for the relation is ambiguous. Previous research showed positive effects of student-teacher gender congruence (e.g., Muralidharan & Sheth, 2016), negative effects (e.g., Lindahl, 2016), and null effects (e.g., Holmlund & Sund, 2008).

These ambiguous findings raise questions about the contexts under which positive effects of student-teacher gender congruence emerge (An et al., 2021). An important condition for positive effects of gender congruence to occur is 'gender salience', i.e., the extent to which gender is perceived as relevant to the context at hand (e.g., Keiser et al., 2002). When gender is perceived as salient, a gender-related self-concept is activated which allows gender to have an influence (Palan, 2001). In schools, gender may be salient in subjects such as math, physics, and languages because these are the subjects in which male and female students' performance is gender stereotyped as fundamentally different (men are stereotyped as naturally talented in math and physics, while women are stereotyped as naturally talented in languages). In these subjects, gender can become an important characteristic for students' identity, interactions, and performance (Hilliard & Liben, 2010; Kiefer & Sekaquaptewa, 2007; Shih et al., 1999). Nevertheless, the strength of gender stereotypes and the role gender stereotypes play in schools may differ in different contexts. Previous research indicates that gender stereotypes may be relatively strong among lower educational levels (Trusty et al., 2000; Turner et al., 2019), in religious schools (Schulze & Tomal, 2005), and in schools located in less-populated areas (García-Retamero et al., 2011).

In this study we examine the relation between teachers' gender and male and female students' performance in four secondary school subjects in the Netherlands. We aim to answer the following research questions:

Research question 1: *To what extent is student-teacher gender congruence associated with male and female students' performance in math, physics, Dutch language and French language?*

Research question 2: *What is the role of students' educational level, schools' religiousness, and schools' location in the relation between student-teacher gender congruence and students' performance?*

We retrieved student data from secondary schools' student administration systems in The Netherlands resulting in a unique dataset with over 50,000 students and their teachers nested in more than 70 schools. Answers to our research questions contribute to previous research on student-teacher gender congruence that mainly focused on female students' performance in math by extending the examination of effects of student-teacher gender congruence to both male and female students' performance in four secondary school subjects. Further, the exploration of the role of context in the relation student-teacher gender congruence and students' performance contributes to the understanding of where and when effects of student-teacher gender congruence may occur.

2.2 STUDENT-TEACHER GENDER CONGRUENCE AND STUDENTS' PERFORMANCE

The relation between student-teacher gender congruence and students' performance has been a topic of study in the fields of economics, education, and public administration for decades. Theories from these diverse fields of research suggest that the match between teachers' gender and students' gender can result in positive effects on students' performance. Positive effects of student-teacher gender congruence are suggested to be the result of three, possibly co-occurring, explanations: teachers' preferences, teachers' gender stereotypes, and students' role model perception (Aslam & Kingdon, 2011). The preference explanation implies that teachers reward students who have the same gender as them with higher grades (Aslam & Kingdon, 2011). In addition to grading, teachers are assumed to pay more attention and to give more positive feedback to students of the same gender (e.g., Grissom et al., 2015). In public administration, these intentional and active behaviors of teachers towards students they represent (student of the same gender) are referred to as active representation (Ricucci & Van Ryzin, 2017). The gender stereotype explanation is similar to the preference explanation but suggests that the favoring of students of the same gender happens implicitly and unconsciously based on teachers' gender stereotypes (Doornkamp et al., 2022). Gender stereotypes are socially constructed ideas that describe what men and women are like and should be like (Ellemers, 2018). Based on gender stereotypes, male students are described as being naturally talented in math while female students are described as being naturally talented in languages. As a result, teachers may have different expectations of and behaviors towards male and female students in different subjects. Lastly, the role-model explanation takes the perspective of the student and argues that students can perceive teachers with the same gender as a role model which stimulates behaviors that positively affect students' performance (Aslam & Kingdon, 2011). In public administration this mechanism is referred to as symbolic representation (Ricucci & Van Ryzin, 2017).

Although student-teacher gender congruence is expected to lead to positive effects,

empirical evidence for this association is mixed (for cross-national comparisons see An et al., 2021; Cho, 2012). Positive relations between student-teacher gender congruence and students' performance were found for both females and males (Dee, 2005; Muralidharan & Sheth, 2016), and only for females and not for males (Lee et al., 2019; Zhang, 2019). Negative relations between student-teacher gender congruence and students' performance have also been found for both females and males (Lindahl, 2016), and only for males and not for females (Doornkamp et al., 2021). Finally, some studies did not find a relation between student-teacher gender congruence and students' performance at all (Holmlund & Sund, 2008; Neugebauer et al., 2011).

2.2.1 THE ROLE OF CONTEXT

These ambiguous findings may imply that there are conditions under which positive as well as negative relations between student-teacher gender congruence and students' performance emerge (An et al., 2021). To our knowledge, thus far the role of different educational contexts in effects of student-teacher gender congruence has not been explored. Although public administration scholars identified organizational (Dhillon & Meier, 2022; Song, 2018) and political conditions (An et al., 2021) for positive effects of student-teacher gender congruence to occur, still little is known about the educational contexts that are favorable for student-teacher gender congruence effects.

Effects of student-teacher gender congruence are assumed to emerge in educational contexts in which gender is perceived as salient, i.e., important, by students and teachers (e.g., Keiser et al., 2002). Gender becomes a salient characteristic when a gender-related self-concept is activated (Palan, 2001). Gender-related self-concepts can be activated in contexts in which there is an observable differentiation between males and females, for instance by dress codes in which women have to wear skirts and males have to wear ties, or when differential language or labels are used to refer to males and females (Hilliard & Liben, 2010). Further, gender can become salient when a policy is directed to benefit women as a group (e.g., policies directed to improve female participation in STEM), when gender of the teacher changes the student-teacher relationship (e.g., in discussing contra-gender stereotypical decisions, students might want to talk to a teacher of the same gender), or when an issue is identified as a gender issue through the political process (e.g., the feminization of the education system) (Keiser et al., 2002).

The differentiation between males and females often happens on the basis of gender stereotypes. Gender stereotypes describe what men and women are like, and what men and women should be like regarding their appearances, behaviors, and capabilities (Ellemers, 2018). In schools, the extent to which gender salience is perceived by teachers and/or students may vary for males and females across subjects with gender stereotypes connected to it. Males and females' capabilities in STEM and languages are gender stereotyped as different: males are gender stereotyped as good at STEM and bad at languages, while females are gender stereotyped as good at languages and bad at STEM.

Negative gender stereotypes (i.e., bad at) are generally easier to internalize and more resistant to change than positive gender stereotypes (Baumeister et al., 2001). Therefore, gender can become a salient characteristic for the group that is stereotypically known for its performance deficits (Beilock et al., 2010; Spencer et al., 2016). As a result, gender may be relatively salient for females in STEM and for males in languages. Accordingly, effects of student-teacher gender congruence are likely to benefit female students' performance in STEM and male students' performance in languages.

Student-teacher gender congruence positively affects students' performance through the behaviors of teachers and/or students. Female teachers in STEM subjects and male teachers in language subjects may have experienced social barriers as a result of gender stereotypes throughout the process of becoming a female teacher in STEM or a male teacher in languages. Because of these personal experiences, female STEM teachers and male language teachers may reward students of the same gender with more points for their answers or give students of the same gender more positive reinforcements (i.e., compliments) in order to reduce social barriers and encourage students of the same gender to participate in the subjects. Further, female role models in STEM fields (Herrmann et al., 2016), and perhaps also male role models in language fields (DUO, 2020), are limited. As a result, female STEM teachers and male language teachers may function as important role models for female students in STEM and male students in languages, which can positively affect students' attitudes, behaviors, and ultimately, their performance (Hermann et al., 2016; Solanki & Xu, 2018). Based on the above we hypothesize:

H1a: Student-teacher gender congruence is positively related to female students' performance in math and physics.

H1b: Student-teacher gender congruence is positively related to male students' Dutch language and French language performance.

Apart from the type of subject, the extent to which gender stereotypes connected to the subjects play a role in effects of student-teacher gender congruence may differ across educational contexts including students' educational levels, schools' religiousness, and schools' location. Previous studies on socioeconomic status (SES) showed that stereotypical gender roles and behaviors are more pronounced in segments of the population with lower SES compared to segments with higher SES (Trusty et al., 2000). In education, students from lower SES segments were found to perceive higher career barriers, including social barriers that stem from gender stereotypes, than students from higher SES segments (Turner et al., 2019). As SES and educational level are closely related concepts (educational attainment is an important indicator for SES and is often used as one of the measures for SES (e.g., Saifi & Mehmood, 2011), the same reasoning might apply to students' educational level. Students in lower educational levels might perceive gender as more salient. Students' educational

pathways in the Netherlands seem to support this idea. The educational pathways of Dutch students reveal gender stereotypic choices: male students tend to choose education in STEM and female students tend to choose education in (health)care (Onderwijsraad, 2019). These gendered choices are the largest among the lowest educational levels (Bügel et al., 2011), possibly implying that gender stereotypes play a more important role in lower educational levels. In this context in which gender stereotypes are assumed to be relatively strong, female teachers in STEM and male language teachers may feel a stronger need to reduce social barriers that stem from gender stereotypes and encourage students of the same gender to participate in the subject, reinforcing effects of student-teacher gender congruence. Further, access to contra-gender stereotypical role models may be even more limited among lower educational levels. The few contra-stereotypical role models in this context may be of greater importance, increasing student-teacher gender congruence effects. We therefore hypothesize:

H2: The associations between student-teacher gender congruence and students' performance are stronger at lower educational levels.

Similarly, gender stereotypes in religious schools might be stronger than gender stereotypes in non-religious schools. Religiousness is often associated with more traditional views on roles of men and women (De Vries et al., 2022). Traditional views on males and females may result in relatively strong gender stereotypes. Indeed, previous research showed that students on religious schools hold stronger gender stereotypical beliefs than students from non-religious schools (Schulze & Tomal, 2005). As a result, we expect effects of student-teacher gender congruence to be stronger in religious schools.

H3: The associations between student-teacher gender congruence and students' performance are stronger in religious schools.

Finally, effects of student-teacher gender congruence might be stronger in schools in less-populated areas compared to heavily populated areas. In general, residents in rural areas express more conservative attitudes and follow more traditional values compared to residents in urban areas (Istenič, 2007). As people in heavy populated areas generally tend to be more progressive and heterogenous, people living in heavy populated areas have a bigger chance of meeting contra-gender stereotypical role models than people living in less-populated areas. Similarly, because of larger population sizes in heavy populated areas, people living in these areas have a bigger chance of meeting contra-gender stereotypical role models. As a consequence, individuals in less-populated areas may hold stronger gender stereotypical beliefs than individuals in heavily populated areas. Indeed, population size and gender stereotypical beliefs were found to be negatively related (Garcia-Retamero et al., 2011). Therefore, effects of student-teacher gender congruence might be stronger in

less-populated areas (Dhillon & Meier, 2022). The research model is visualized in Figure 2.1.

H4: The associations between student-teacher gender congruence and students’ performance are stronger in schools located in less populated areas.

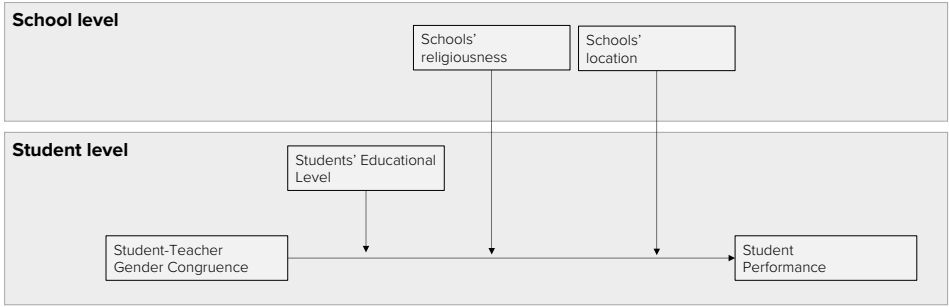


Figure 2.1. Research Model

2.3 METHOD

2.3.1 PROCEDURE

This study is part of the research project Girls in Science running from 2017 until 2022 in the Netherlands which examines gender socialization in the family and school context. The data for this study were collected between December 2019 and February 2021 through secondary schools’ student administration systems. In collaboration with The Implementation Group (TIG) we were able to subtract relevant data from these systems. TIG is an organization that develops administrative tools for secondary schools to inform schools about patterns in their student data. Approximately half of all secondary schools in the Netherlands are affiliated with TIG. Through this affiliation TIG has access to a great amount of secondary school data.

The affiliated schools were asked to participate in the research. Participation involved giving consent to TIG to transfer a selection of anonymized administrative data to the researchers for research purposes. The data included students’ gender, educational level per year, subject choice per year, student performances (grades) per year in four subjects (math, physics, Dutch language, and French language), and corresponding teachers’ gender for the subjects per year over the period 2011 until 2021. Additionally, school characteristics including school denomination, school structure, and level of urbanity were collected based on open access data. Participating schools received the results of the study and an invitation to a symposium on the role of gender in education. Seventeen percent of schools affiliated with TIG participated in the research resulting in an unique dataset with 403.984

students nested in 195 secondary schools.

2.3.2 SAMPLE

This study focused on the last year of lower secondary education when all students still participate in all subjects. In this year, students choose the subjects they want to participate in higher secondary education and ultimately want to graduate in. The year in which Dutch students choose their subjects and transfer from lower to higher secondary education is dependent on the educational level they are in. Students in the educational level that prepares for vocational education (vocational track) have four years of secondary education and choose their subjects in their second year. Students in the educational level that prepares for applied universities (higher vocational track) as well as students in the educational level that prepares for university (scientific track) have respectively five and six years of secondary education and choose their subjects in their third year. As a result, this study examines students' data from students in year 2 (vocational track) and year 3 (higher vocational track and scientific track).

Some Dutch secondary schools provide combinations of educational levels in lower secondary education. In this study we excluded students who combined or switched educational level. We included students' that started secondary education in 2011 until 2021. We include only those students with scores on all relevant variables per subject, resulting in $N = 83,271$ for math (51% male), $N = 65,902$ for physics (49% male), $N = 87,293$ for Dutch language (49% male), and $N = 58,242$ for French language (48% male).

2.3.3 MEASURES

Student-Teacher Gender Congruence

The independent variable refers to student and teacher having the same gender (0 = no gender congruence, 1 = gender congruence). Students' and teachers' gender were administrated in the secondary school administration system as male, female, or neutral. Students' ($N = 2$) and teachers' ($N = 133$) that were administrated as neutral were regarded as missing data.

The distribution of male and female teachers across the subjects is skewed but similar for male and female students (see Table 2.1). The percentages of students with a female teacher for the subjects are: math 37%, physics 26%, Dutch language 69%, and French language 79%. The distribution of male and female teachers over the subject in our sample is comparable with the distribution in the Netherlands (math 45% female, physics 21% female, Dutch language 74% female, and French language 82% female (DUO, 2020)). The skewed distribution is not expected to introduce bias, because teachers do not select their students (or vice versa).

Student Performance

The dependent variable refers to students' grades for math, physics, Dutch language, and

French language. Students' grades were obtained from the calculated averages ranging from 1.0 to 10.0 administrated in the secondary school administration systems. Grades between 0.0 and 1.0 were regarded as missing data (N = 9).

Students' Educational Level

The first moderating variable, educational level, was obtained from students' 'ILT-code' (Integral Student Count) that was administrated in the secondary school administration systems. The ILT-code is a governmental code that includes students' educational level. The ILT-code variable was recoded into the educational variable with 1= vocational track, 2= higher vocational track, and 3= scientific track.

Schools' Religiousness

The second moderating variable, schools' religiousness, was deduced from schools' websites. Schools with an open character not referring to any religion or explicitly stating not to be a religious school were coded as 0 = not religious. Schools with religious roots referring to religious values but explicitly stating to be open to students from any (non) religious background were coded as 1 = religious roots. Schools that explicitly stated that they practice religion in their education were coded as 2 = practicing religious. Students on schools whose degree of religiousness could not be coded based on the information provided on their website were regarded as missing data (N = 14719).

Schools' location

The third moderating variable refers to the address density per squared kilometer of the schools' location. We obtained data on address density in the postal code of the participating schools using open governmental data. We recoded level of urbanity variable in 0= suburban, less-populated area (less than 1500 addresses), and 1= urban, heavy populated area (more than 1500 addresses), based on the breaking point between 'moderate' and 'strong urban' governmental classification (CBS, 2022).

2.3.4 DATA ANALYSES

The data inspection and further analyses were conducted using R statistical software. To check for normal distributions Q-Q plots were inspected. The normality checks indicated that some variables had heavy tails, however this type of minor divergence from normality is unlikely to cause any major problems given the large sample sizes. After checking the assumptions, which were not violated, no clear signs of outliers were observed. This in combination with the large sample sizes and the scales of the variables, outliers are assumed not to be influential.

To test our hypotheses, we performed multilevel ANOVA's using the lme4 package (Bates et al., 2015). Additional post hoc tests were conducted to examine whether the mean scores of students with gender congruent teachers were significantly different from students

with gender incongruent teachers (hypothesis 1). To test the robustness of our results for hypothesis one, we conducted the analyses on the largest possible sample size per subject (results reported in Appendix A). Further, post hoc analyses were performed to test whether differences in mean scores for student-teacher gender congruence were observed for different educational levels (hypothesis 2), for schools' religiousness (hypothesis 3), and for schools' location (hypothesis 4).

Table 2.1. Range, Means, Standard Deviations, and Distributions of Central Variables

			Mean (SD) and Distribution		
		Range	Full Sample N = 83,271	Male Students N = 41,223	Female Students N = 42,048
Math	1 Grade	1 – 10	6.51 (1.11)	6.42 (1.08)	6.59 (1.13)
	2 Gender Congruence	0, 1	50/50	37/63	63/37
	3 Educational Level	1, 2, 3	46/27/27	48/27/25	44/27/29
	4 Schools' Religiousness	0, 1, 2	26/61/13	25/61/13	26/60/13
	5 Schools' Location	0, 1	39/61	39/61	39/61
		Range	Full Sample N = 65,902	Male Students N = 32,289	Female Students N = 33,613
Physics	1 Grade	1 – 10	6.50 (0.96)	6.52 (0.94)	6.48 (0.97)
	2 Gender Congruence	0, 1	50/50	25/75	74/26
	3 Educational Level	1, 2, 3	34/33/33	37/33/31	32/33/35
	4 Schools' Religiousness	0, 1, 2	25/60/15	25/60/15	25/59/16
	5 Schools' Location	0, 1	34/66	35/65	34/66
		Range	Full Sample N = 87,293	Male Students N = 43,157	Female Students N = 44,136
Dutch	1 Grade	1 – 9.8	6.55 (0.74)	6.33 (0.70)	6.77 (0.70)
	2 Gender Congruence	0, 1	50/50	69/31	31/69
	3 Educational Level	1, 2, 3	45/27/28	47/27/26	42/27/30
	4 Schools' Religiousness	0, 1, 2	26/61/13	26/61/13	26/61/13
	5 Schools' Location	0, 1	38/62	38/62	38/62
		Range	Full Sample N = 58,242	Male Students N = 27,808	Female Students N = 30,434
French	1 Grade	1 – 10	6.53 (1.11)	6.21 (1.08)	6.83 (1.06)
	2 Gender Congruence	0, 1	49/51	79/21	21/79
	3 Educational Level	1, 2, 3	24/37/39	25/38/38	23/36/41
	4 Schools' Religiousness	0, 1, 2	26/59/15	27/58/15	26/60/14
	5 Schools' Location	0, 1	32/68	32/68	32/68

Note: Distributions are presented in percentages. Gender congruence 0= no congruence, 1= congruence; Educational level 1= vocational track, 2= higher vocational track, 3= scientific track; Schools' religiousness 0= not religious, 1= religious roots, 2= practicing religious; Schools' location 0= suburban, 1= urban.

2.4 RESULTS

2.4.1 STUDENT-TEACHER GENDER CONGRUENCE AND STUDENTS' PERFORMANCE

Table 2.1 presents the range, means, standard deviations, and distributions for student-teacher gender congruence, students' performance, and the contextual variables in math, physics, Dutch language, and French language. Table 2.2 presents the means for the different student-teacher gender combinations. Both Table 2.1 and 2.2 show that female students had higher average grades than male students in math, Dutch language, and French language, but not in physics.

Table 2.2. Student-Teacher Gender Congruence and Student Performance

	Male Student		Female Student	
	Male Teacher	Female Teacher	Male Teacher	Female Teacher
Math N = 83,271	6.36 ^a	6.36 ^a	6.51 ^b	6.56 ^c
Physics N = 65,902	6.46 ^a	6.45 ^a	6.39 ^b	6.47 ^a
Dutch N = 87,293	6.35 ^a	6.28 ^b	6.75 ^c	6.73 ^c
French N = 58,242	6.22 ^a	6.12 ^b	6.78 ^c	6.77 ^c

Note: Means with different superscripts in rows are significantly different from each other on $p < 0.001$, Standard Errors range from 0.03 to 0.04.

Further, Table 2.2 shows that in math female students with female teachers had significantly higher grades than female students with male teachers. Male students' math performance did not differ for the gender of their teachers. In physics, male students with male teachers, male students with female teachers, and female students with female teachers had (approximately) the same average grades, only female students with male teachers had significantly lower grades. In Dutch language and French language, female students' performance did not differ for the gender of their teacher. However, male students with male teachers had significantly higher grades than male students with female teachers in both Dutch language and French language. Based on these results, hypothesis 1 should be accepted: student-teacher gender congruence was positively related to performance in math and physics for female students (H1a) and performance in Dutch language and French language for male students (H1b). Robustness checks on the largest possible samples confirmed the results (see Appendix A).

2.4.2 THE ROLE OF CONTEXT

In the next step we tested whether students' educational level, schools' religiousness, and schools' location moderated the relation between student-teacher gender congruence and students' performance. Table 2.3 and Table 2.4 present the results of the moderation analyses for male (Table 2.3) and female (Table 2.4) students' performance.

As expected, student-teacher gender congruence explained variance in student

performance for females in math and physics and for males in Dutch language and French language. Additionally, Table 2.4 shows that student-teacher gender congruence explained variance in female students' Dutch language performance. Based on our hypotheses we expected student-teacher gender congruence to matter in particular in the lowest educational level (vocational track), in religious schools, and in suburban located schools. We therefore expected (the largest) negative significant difference between the incongruent and congruent condition in the bottom half of Table 2.3 and 2.4 for the vocational track, religious schools, and suburban located schools. We will discuss the results for each contextual factor separately.

Table 2.3. Multilevel ANOVA of Male Student-Teacher Gender Congruence on Students' Performance, Moderated by Students' Educational Level, Schools' Religiousness, and Schools' Location with Post Hoc Test.

	Math		Physics		Dutch		French	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Gender Congruence	0.34	1.65	0.35	0.06	85.77***	71.77**	20.13***	2.17
Educational Level	285.67***	227.08***	638.34***	567.64***	462.64***	439.81***	426.79***	303.81***
Religiousness	1.29	1.34	2.20	2.77	2.05	2.17	0.23	0.50
Location	0.20	0.19	0.31	0.41	2.28	0.88	0.40	1.09
GCxEduLevel		1.94		49.50***		2.12		1.59
GCxReligiousness		6.26**		8.14***		1.72		3.45*
GCxLocation		0.00		1.49		12.88***		8.70**
Conditional R2	0.08	0.08	0.14	0.15	0.16	0.16	0.14	0.14
	Mean	Difference	Mean	Difference	Mean	Difference	Mean	Difference
GC0x EduLevel1	6.32		6.48		6.26		5.90	
GC1x EduLevel1	6.35	-0.03	6.37	0.11***	6.34	-0.08***	5.98	-0.08*
GC0xEduLevel2	6.26		6.16		6.17		6.04	
GC1xEduLevel2	6.24	0.01	6.34	-0.18***	6.23	-0.06***	6.05	-0.01
GC0x EduLevel3	6.56		6.84		6.44		6.43	
GC1x EduLevel3	6.61	-0.04	6.76	0.08**	6.54	-0.10***	6.44	-0.01
GC0xReligiousness0	6.29		6.42		6.21		6.11	
GC1xReligiousness0	6.36	-0.07**	6.40	0.02	6.27	-0.06***	6.19	-0.08*
GC0xReligiousness1	6.40		6.40		6.31		6.15	
GC1xReligiousness1	6.37	0.03	6.48	-0.08***	6.39	-0.08***	6.23	-0.08***
GC0xReligiousness2	6.46		6.66		6.34		6.10	
GC1xReligiousness2	6.47	-0.01	6.58	0.08*	6.46	-0.11***	6.05	0.05
GC0xLocation0	6.37		6.46		6.25		6.10	
GC1xLocation0	6.39	-0.02	6.48	-0.01	6.36	-0.11***	6.08	0.03
GC0xLocation1	6.39		6.53		6.33		6.14	
GC1xLocation1	6.41	-0.02	6.50	0.02	6.38	-0.05***	6.23	-0.09***

Note: In the top half, F-Values of the Multilevel ANOVA are reported, in the bottom half the gender congruent condition was post hoc tested against the gender incongruent condition (i.e., second row vs first row), ***p<0.001, **p < 0.01, *p< 0.05

Table 2.4. Multilevel ANOVA of Female Student-Teacher Gender Congruence on Students' Performance, Moderated by Students' Educational Level, Schools' Religiousness, and Schools' Location with Post Hoc Test.

	Math		Physics		Dutch		French	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Gender Congruence	6.95 ^{***}	2.54	14.12 ^{***}	4.78 [*]	17.89 ^{***}	8.09 ^{***}	1.05	0.26
Educational Level	427.95 ^{***}	377.66 ^{***}	948.43 ^{***}	760.20 ^{***}	711.17 ^{***}	626.73 ^{***}	548.62 ^{***}	380.29 ^{***}
Religiousness	1.18	1.07	3.95 [*]	3.93 [*]	2.39	2.43	0.97	0.92
Location	0.96	1.40	0.16	0.08	0.01	0.01	1.58	0.86
GCxEduLevel		0.88		36.96 ^{***}		2.09		1.98
GCxReligiousness		1.35		0.76		1.12		0.28
GCxLocation		12.38 ^{***}		1.45		0.00		4.76 [*]
Conditional R2	0.09	0.09	0.16	0.16	0.19	0.19	0.13	0.13
	Mean	Difference	Mean	Difference	Mean	Difference	Mean	Difference
GC0x EduLevel1	6.51		6.38		6.72		6.63	
GC1x EduLevel1	6.55	-0.04 [*]	6.54	-0.16 ^{***}	6.72	0.01	6.57	0.06
GC0xEduLevel2	6.34		6.23		6.62		6.63	
GC1xEduLevel2	6.37	-0.03	6.12	0.11 ^{***}	6.58	0.04 ^{**}	6.63	-0.00
GC0x EduLevel3	6.79		6.75		6.95		7.05	
GC1x EduLevel3	6.79	-0.00	6.81	-0.07 ^{***}	6.91	0.03 [*]	7.07	-0.02
GC0xReligiousness0	6.48		6.33		6.66		6.75	
GC1xReligiousness0	6.51	-0.03	6.39	-0.06	6.64	0.02	6.76	-0.01
GC0xReligiousness1	6.52		6.42		6.80		6.84	
GC1xReligiousness1	6.58	-0.05 ^{***}	6.43	-0.01	6.76	0.04 ^{***}	6.83	0.01
GC0xReligiousness2	6.63		6.60		6.83		6.71	
GC1xReligiousness2	6.63	0.01	6.65	-0.04	6.81	0.02	6.68	0.03
GC0xLocation0	6.56		6.47		6.77		6.79	
GC1xLocation0	6.63	-0.07 ^{**}	6.49	-0.02	6.74	0.03	6.82	-0.03
GC0xLocation1	6.53		6.43		6.76		6.75	
GC1xLocation1	6.51	0.02	6.49	-0.06 ^{**}	6.73	0.03 ^{**}	6.70	0.05 [*]

Note: In the top half, F-Values of the Multilevel ANOVA are reported, in the bottom half the gender congruent condition was post hoc tested against the gender incongruent condition (i.e., second row vs first row), ***p<0.001, **p < 0.01, *p< 0.05

Students' educational level

The interaction between student-teacher gender congruence and educational level was not significant in math, Dutch language, and French language indicating that the relation between student-teacher gender congruence and students' performance did not differ for students' educational level in these subjects. However, in physics, the interaction between gender congruence and educational level explained variance in both males' and females' physics performance. For males, the post hoc analyses revealed that gender congruence was related to higher grades in the higher vocational track while gender congruence was related to lower grades in the vocational track and the scientific track. For females, gender congruence was related to higher grades in the vocational and scientific track, whereas gender congruence was related to lower grades in the higher vocational track. These findings are visualized in Figure 2.2. In all, hypothesis 2 can only be accepted in the subject

physics for female students in the vocational and scientific track and for male students in the higher vocational track, in all other contexts the hypothesis should be rejected.

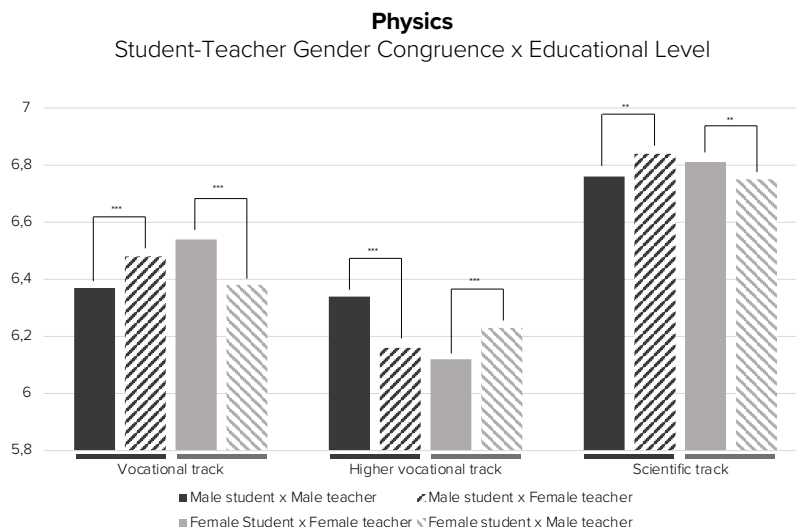


Figure 2.2. Student-Teacher Gender Congruence and Educational Level on Students' Physics Performance

Schools' religiousness

The interaction between student-teacher gender congruence and schools' religiousness was significant for males in math, physics, and French language. The interaction between gender congruence and religiousness was not significant for male students in Dutch language, nor for female students in any of the subjects. For males in math, the post hoc analyses revealed that in non-religious schools, male students with male teachers had slightly higher average grades than male students with a female teacher. In schools with religious roots and in schools that practice religion, gender congruence was not associated with male students' math performance. In physics, gender congruence was associated with higher grades for males in schools with religious roots whereas gender congruence was associated with lower grades in schools that practice religion. In non-religious schools gender congruence did not affect male students' physics grades. In French language, gender congruence was related to higher grades for males in non-religious schools and in schools with religious roots. Gender congruence was not related to male students' performance in religious schools. The findings are visualized in Figure 2.3. In all, hypothesis 3 should be rejected as student-teacher gender congruence was not associated with higher average grades in any of the subjects for neither male students nor female students.

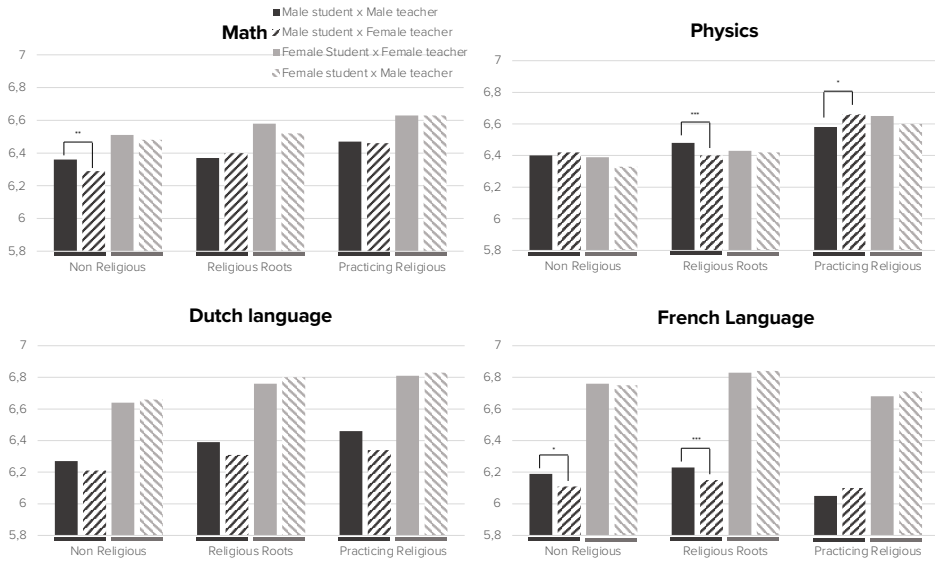


Figure 2.3. Student-Teacher Gender Congruence and Schools' Religiosity on Students' Performance by Subject

Schools' location

The interaction between student-teacher gender congruence and schools' location was significant for male students in Dutch language and French language and for females in math and French language. For males in Dutch language, the post hoc analyses showed that gender congruence was related to higher grades in suburban-located schools. Gender congruence was not related to male students' Dutch language performance in schools that are located in urban areas. In French language, gender congruence was related to higher grades in schools located in urban areas, and not in schools located in suburban areas. For females in math, student-teacher gender congruence was related to higher grades in schools located in suburban areas and not in schools located in urban areas. In French language, gender congruence in schools in urban areas was related to lower grades. Gender congruence was not related to female students' performance in schools in suburban areas. The findings are visualized in Figure 2.4. In all, hypothesis 4 should only be accepted for male students in Dutch language and female students in math.

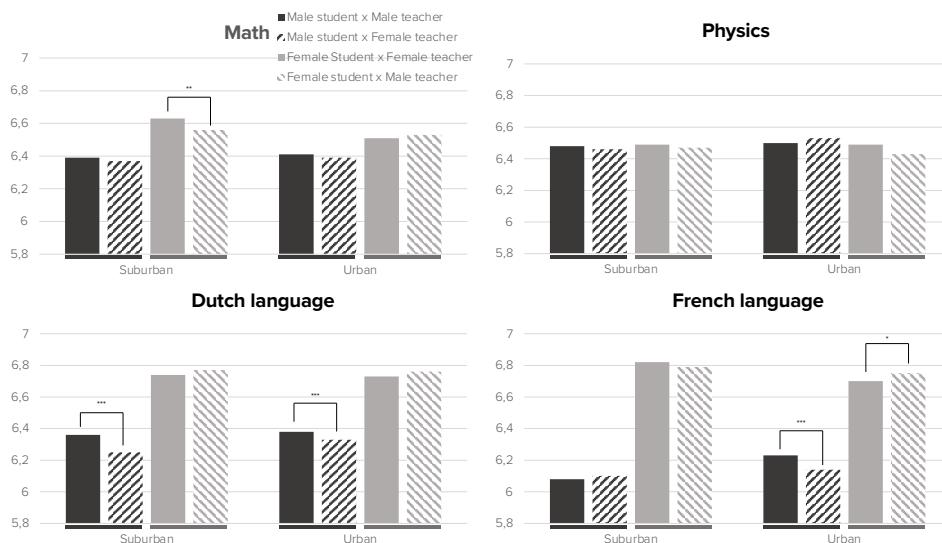


Figure 2.4. Student-Teacher Gender Congruence and Schools' Location on Students' Performance by subject

2.5 DISCUSSION AND CONCLUSION

This study tested the relation between student-teacher gender congruence and students' performance and examined the role of students' educational level, schools' religiousness, and schools' location. We found that student-teacher gender congruence was positively related to female students' performance in math and physics and to male students' performance in Dutch language and French language. Further, we found differential roles for different educational contexts for males and females and across subjects.

2.5.1 GENDER CONGRUENCE AND PERFORMANCE

Our results support previous research in the fields of economics, education, and public administration that found positive effects of student-teacher gender congruence for both males and females (e.g., Dee, 2005), but adds to this literature that positive effects of student-teacher gender congruence emerge in different subjects for male and female students. We found that female students benefited from female teachers in math and physics but not in Dutch language and French language, and that male students benefited from male teachers in Dutch language and French language but not in math and physics. We argued that this differential role of student-teacher gender congruence for males and females could be the result of differences in the extent to which males and females perceived gender salience in given subjects (Doornkamp et al., 2021; Keiser et al., 2002). We suggested that gender was salient for female teachers and students in STEM and for male teachers and students in

languages because these are the contexts in which these groups are stereotypically thought to have performance deficits (e.g., Spencer et al., 2016). In these contexts, effects of student-teacher gender congruence may have emerged through teachers' and/or students' attitudes and behaviors. From a teacher perspective, female teachers in STEM and male teachers in languages might have tried to (intentionally or unintentionally) reduce social barriers that stem from gender stereotypes by rewarding students of the same gender with higher grades or by giving more positive feedback to them (Aslam & Kingdon, 2011, Grissom et al., 2015). From a student perspective, female teachers in STEM and male teachers in languages might have functioned as important contra-gender stereotypical role models affecting students' attitudes, behaviors, and ultimately performance (e.g., Solanki & Xu, 2018).

An additional remark with regard to these findings should be made. In physics, female student-teacher gender congruence and performance was not *positively* related as female students with a female teacher did not perform better than average. Instead, female students with a female teacher performed as well as male students with both male and female teachers. Female students with a male teacher performed worse than the average (worse than all other groups) indicating that student-teacher gender *incongruence* had a negative effect on female students' physics performance. Negative effects of student-teacher gender incongruence have been found in previous research (Dee, 2005; Solanki & Xu, 2018). However, it should be noted that in contrast to our finding, these studies found that the negative effects of gender incongruence occurred simultaneously with positive effects of gender congruence (Dee, 2005; Solanki & Xu, 2018). Negative effects of gender incongruence could have stemmed from male teachers' differential treatment of male and female students or from negative attitudes or self-beliefs in physics in female students as a result of having a male physics teacher. Perhaps, male physics teachers had strong associations between physics and masculinity (Doornkamp et al., 2022). Therefore, male physics teachers might have underestimated female students' capabilities in physics which could have led to stricter grading or generating less positive reinforcements for female students (Denessen et al., 2022). From a student perspective, female students' performance might have been weakened by having a male physics teacher through the mechanism of stereotype threat (Kapitanoff & Pandey, 2017). Female students' physics performance could have been suppressed by the stereotypical belief that women are bad at physics. Whereas female physics teachers could have reduced effects of stereotype threat on female students' performance by being a role model, male physics teachers could not (Kapitanoff & Pandey, 2017).

2.5.2 THE ROLE OF CONTEXT

We expected gender stereotypes to be relatively strong and therefore effects of student-teacher gender congruence to occur in lower educational levels (Trusty et al., 2000; Turner et al., 2019), in religious schools (Schulze & Tomal, 2005), and in schools located in less-populated areas (Garcia-Retamero et al., 2011). However, our results did not support the

hypotheses. We found differential patterns for educational context for male and female students and across the subjects.

Evidence for the role of students' educational level was only found in the subject physics and the role was opposite for male and female students. In the vocational and scientific track, gender congruence led to positive effects for female students' performance while gender congruence led to negative effects for male students' performance. In the higher vocational track our findings were opposite: gender congruence was related to better performance for male students and worse performance for female students. The contradictory findings across educational levels raise questions about the dissimilarities in students and teachers in different educational levels. To gain more insight in these dissimilarities, future research could use qualitative research methods to systematically compare students and teachers in different educational levels.

We found no role of schools' religiousness for female students and the role of schools' religiousness for male students was contrary to what we expected. Gender congruence was associated with better performance of male students in non-religious schools for the subjects math and French language, and in schools with religious roots for the subject physics. Only for male students in physics we found an effect in religious schools, but this effect was negative: gender congruence was associated with lower grades for male students. Perhaps, specific student characteristics that are related to school religiosity could have acted as a buffer against the role of gender stereotypes and student-teacher gender congruence effects. For instance, students in religious schools generally perform better than students in non-religious schools (Jeynes, 2002). Relatively good performance is associated with having more self-perceived ability (Marsh et al., 2005), and students with more self-perceived ability are found to be less susceptible for effects of student-teacher gender congruence (Eble & Hu, 2019).

For males in Dutch language and females in math, gender congruence led to better performance in less-populated areas and not in heavily populated areas which supported our hypothesis. These findings are similar to those of Dhillon and Meier (2020) who found positive gender congruence effects for females in math in rural areas. It should be noted that this finding was more complex as math performance of female students decreased when the female teacher percentage increased (Dhillon & Meier, 2020). In addition, in French language, gender congruence was associated with higher grades for males and lower grades for females in heavily populated areas. Further, schools' location did not affect the relation between student-teacher gender congruence and performance in physics. These findings, and the other findings regarding the role of context, emphasize that effects of student-teacher gender congruence can differ in magnitude and direction in different contexts, but a clear pattern in these differential findings was not observed in this study.

2.5.3 LIMITATIONS

Our findings should be interpreted within this study's limitations. First, as a result of our large

dataset very small differences in means can become statistically significant (Khalilzadeh & Tasci, 2017). Statistically significant effects smaller than 0.05 (on the 95% confidence interval) could be considered as practically insignificant. The results of our moderation analyses showed statistically significant mean differences smaller than 0.05 indicating that students with gender congruent teachers achieved less than 0.05 higher average grades than students with gender incongruent teachers. This difference is very small and dissolves when average grades are round to 1 decimal, which is the case in most of Dutch secondary schools. Further, these effects might become statistically insignificant when student-teacher gender congruence effects are controlled for by predictors such as academic self-concept and student IQ which have been found to be important predictors of students' performance (e.g., Wach et al., 2015). Nevertheless, the small differences in average grades for male and female students with a teacher of the same gender compared to a teacher of the other gender can have important consequences. In an education system in which calculated end averages are the basis for decisions with regard to students' educational pathways (which is the case in the Netherlands), 0.1 can make the difference for passing a grade, moving up an educational level, or admission to further education.

Second, multicollinearity was introduced in some of our models and, as a consequence, variances were inflated. The models without interaction terms showed no signs of multicollinearity indicating that variance inflation was caused by the introduction of the interaction terms which does not create a multicollinearity problem (Disatnik & Sivan, 2016; Shieh, 2010).

Finally, our dataset relied on administrative data of secondary schools. The secondary schools' administration systems were sometimes inconsistent in their registration. For instance, students' grades can be registered under different columns with different terminology. Because of these inconsistencies in the administration of grades, several schools provided incomplete datasets. We asked participating schools to improve their registration and offered specific instructions and help for these improvements. Although some schools were able to improve their data for this study, not all schools were able to do so. Nevertheless, because we used registration data we still were able to collect a large amount of data without bias and/or systematic missings.

2.5.4 CONCLUSION

In all, our study showed that student-teacher gender congruence not necessarily leads to better performance of male and female students. Our findings indicated that having a teacher of the same gender led to positive effects in performance in secondary school subjects for the group of students that was confronted with a negative gender stereotype connected to that subject. We theorized that these effects of student-teacher gender congruence could differ across educational contexts because of differences in the extent to which gender stereotypes play a role. Indeed, our results showed differential effects of student-teacher gender congruence in different contexts encouraging future research to

further explore the role of context including factors such as socioeconomic status, schools' size, and schools' structure. Nevertheless, the differential effects of student-teacher gender congruence were not always consistent raising questions about *how* context influences student-teacher gender congruence effects. Future research might want to use qualitative designs to explore the mechanisms that explain how context affects the role of gender in education.

2.6 APPENDIX

2.6.1 APPENDIX A – ROBUSTNESS CHECK

Table Robustness check Student-Teacher Gender Congruence and Student Performance

	Male Student		Female Student	
	Male Teacher	Female Teacher	Male Teacher	Female Teacher
Math N = 97082	6.35 ^a	6.35 ^a	6.50 ^b	6.54 ^c
Physics N = 76286	6.44 ^a	6.46 ^a	6.38 ^b	6.47 ^a
Dutch N = 102267	6.36 ^a	6.28 ^b	6.75 ^c	6.74 ^c
French N = 68097	6.25 ^a	6.15 ^b	6.81 ^c	6.79 ^c

Note: Means with different superscripts in rows are significantly different from each other on $p < 0.001$. Standard Errors range from 0.02 to 0.03.



CHAPTER **THREE**

The individual level effect of symbolic representation: An experimental study on teacher-student gender congruence and students' perceived abilities in math

Previously published as: Doornkamp, L., Van den Bekerom, P., & Groeneveld, S. (2019). The individual level effect of symbolic representation: An experimental study on teacher-student gender congruence and students' perceived abilities in math. *Journal of Behavioral Public Administration*, 2(2), 1-11.

ABSTRACT

Studies on representative bureaucracy have often confirmed the positive performance effects of bureaucracies mirroring the demographic characteristics of their clientele. However, little is known about the underlying individual-level mechanisms leading to these outcomes. In this study, theoretical ideas from representative bureaucracy literature and social and educational psychology are combined in a new model that explains effects of passive representation from the perspective of the individual client in the educational field. It is hypothesized that positive effects of gender congruence on students' academic self-concepts are mediated by gender stereotypical beliefs of students. This mediation is expected to be moderated by the self-confidence of the teacher. Results of a survey experiment among students in a Dutch high school do not support the hypothesized relationships. The study does reveal gender differences in stereotypical beliefs and academic self-concepts though. Furthermore, the academic self-concept for math of both male and female students is higher if the math teacher is a woman. The study concludes with a discussion of the findings and avenues for future research on the role of stereotypical beliefs in the association between gender representation and student performance.

3.1 INTRODUCTION

Representative bureaucracy literature argues that governmental organizations should mirror demographic characteristics of their clientele because representative organizations would meet democratic values and perform better (Groeneveld & Van de Walle, 2010; Keiser et al., 2002; Meier et al., 1999; Sowa & Selden, 2003). Although positive effects of representation have been confirmed by a variety of studies, little is known about underlying individual-level mechanisms explaining these effects. This may be attributed to the aggregate level of analysis of most studies, making it impossible to identify theoretical mechanisms that are grounded in the behavior of individual bureaucrats and clients (Nicholson-Crotty et al., 2016).

The literature distinguishes between active, passive, and symbolic representation. Positive effects of representation are generally attributed to bureaucrats being actively representative, meaning that bureaucrats (intentionally) act to improve outcomes for the clients they represent (e.g., Meier et al., 2006). The presence of minority bureaucrats may also indirectly affect outcomes for minority clients by creating an organizational culture in which nonminority bureaucrats become aware of the interests and needs of minority clients (Nicholson-Crotty et al., 2016). Symbolic representation shifts the focus from the bureaucrat to the perspective of the client and assumes that clients' attitudes and behavior change as a result of the mere presence of bureaucrats with a similar background (e.g., Gade & Wilkins, 2013).

Only a handful of studies have focused on individual level effects of symbolic representation (e.g., Gade & Wilkins, 2013; Guul, 2018; Nicholson-Crotty et al. 2016; Song, 2018; Theobald & Haider-Markel, 2008). For instance, Guul (2018) found that clients increased their job seeking efforts if they were matched with a job counselor with the same gender. Still, much remains unknown as to what psychological processes on the part of the client are responsible for changes in clients' attitudes and behavior. The present study, which focuses on the field of education, unravels one possible mechanism explaining individual effects of symbolic representation by combining representative bureaucracy literature with insights from social and educational psychology.

Educational performance is strongly dependent on students' perceived abilities, referred to as academic self-concepts (e.g., Marsch et al., 1988). For subjects with gender related stereotypes connected to them, students' academic self-concepts are argued to be influenced by stereotypical beliefs (Beilock et al., 2010; Marsch et al., 1988; Pajares & Miller, 1994). Combining these insights with representative bureaucracy literature, it is hypothesized that gender congruence in the student-teacher relationship in math education positively affects academic self-concepts, because the teacher either confirms (for male students) or rejects stereotypes (for female students) connected to the subject. The empirical study is therefore guided by the following research question:

How does teacher-student gender congruence influence the academic self-concepts of female and male students?

3.2 REPRESENTATIVE BUREAUCRACY AND PERFORMANCE IN EDUCATION

A lot of research on representative bureaucracy focusses on the educational sector. Education is one of the largest branches of the public sector and a cornerstone of our society in that it builds the knowledge and skills of citizens (Meier & Stewart, 1992). The three mechanisms underlying representation and performance are argued to occur in education. Teachers may for instance tailor their teaching to the needs of students they represent in order to increase these students' performance, or use terms during class and in test questions that are culturally sensitive. Teachers may also inform their nonminority colleagues about qualities and needs of minority students, which leads them to adjust their teaching methods to minority students' needs, increasing their performance (Grissom et al., 2015). The mere presence of a representative teacher may also affect represented students' performance as a result of attitudinal and behavioral changes on the part of the student.

Symbolic representation explains performance effects of representation from the perspective of the client. Clients are argued to be more responsive to, and receptive of bureaucrats with a similar background (Gade & Wilkins, 2013; Guul, 2018). When confronted with a bureaucrat of a similar background, the represented client may (unwittingly) change his or her attitude, behavior, and cooperation with a program or policy, increasing the output for him or herself. Hence, students may change their attitudes and behavior as a consequence of the presence of a teacher with a similar background: school identification may increase, a representative teacher may be a role model, or a teacher may reject certain stereotypes, all of which can increase student performance (Beilock et al., 2010; Grissom et al., 2015; Zhang, 2019).

3.2.1 SYMBOLIC REPRESENTATION, STEREOTYPES, AND ACADEMIC SELF-CONCEPTS

Studies on symbolic representation have studied its effects as a change in observable behavior or attitudes of clients such as changes in effort for or judgements about a program (Guul, 2018; Nicholson-Crotty et al., 2016). However, studying less observable processes underlying changes in behavior, attitudes, and performance may help to better understand why and how effects of symbolic representation occur in an educational setting. Studies in educational and social psychology point at the role of stereotypes and self-concepts in students' performance. Studies show that student performance is strongly dependent on students' academic self-concepts, i.e., beliefs of self-worth, and judgements about one's performance in a subject (e.g., Pajares & Miller, 1994). Stereotypical beliefs are argued to influence these academic self-concepts (e.g., Beilock et al., 2010).

A strong gender stereotype about math (i.e., men are good at math and women are bad at math) is expected to affect female and male students' beliefs and judgements about their own performance in math. Studies on gender differences in academic self-concepts in math suggest that those differences can be attributed to stereotypical beliefs about the

subject (Marsh et al., 1988; Pajares & Miller, 1994). This idea is confirmed by many scholars; multiple studies found a difference in female and male students' math performance and partly attributed it to the stereotypical belief connected to the subject (Beilock et al., 2010; McFarland et al., 2016; Steele, 2003; Zhang, 2019).

Teacher-student gender congruence, which is the gender match between teacher and student, is expected to impact the aforementioned psychological processes in students. A teacher can serve as evidence for or against the stereotype connected to math, influencing students' stereotypical beliefs and academic self-concepts. Female students are expected to perform better with a female teacher because a female teacher serves as evidence against the stereotype that women are bad at math. The rejection of the stereotype decreases female students' stereotypical beliefs, thereby increasing their beliefs and judgements about their capabilities and performance. The stronger academic self-concepts result in better performance. Vice versa, male students' performance would increase with a male teacher, because a male teacher confirms the stereotype that men are good at math, strengthening the male students' stereotypical belief which in-turn strengthens their academic self-concepts and performance. This results in the following hypotheses:

H1: Teacher-student gender congruence increases students' academic self-concepts.

H2: The effect of teacher-student gender congruence on academic self-concepts is mediated by stereotypical beliefs.

The self-confidence of the teacher might moderate the relationship between gender congruence and stereotypical beliefs. For instance, Beilock et al. (2010) show that girls' self-confidence in math decreases if their female teacher is anxious about teaching math. By being anxious, the female teacher confirms the negative stereotypical belief about math rather than rejecting it. In this case, female students' stereotypical beliefs may be stronger with an insecure female math teacher, because she confirms stereotypical beliefs, reducing female students' academic self-concept. The conceptual model is depicted in Figure 3.1.

H3: The indirect effect of gender congruence on academic self-concepts is moderated by teacher self-confidence.

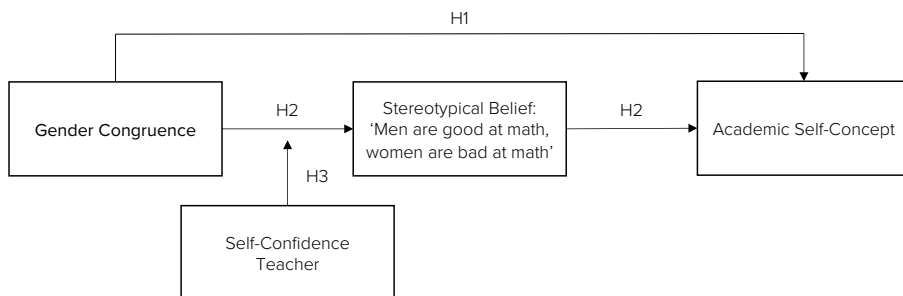


Figure 3.1. Research Model

3.3 METHOD

An online survey experiment was conducted to test the hypotheses. The survey experiment was run at a Free Reformed high school that teaches regular education in the three main educational levels of the Dutch secondary education system. Because of the religious character of the school, it is difficult to generalize the findings of this study to other school contexts. However, the weak external validity of the study is compensated for by the strong internal validity of the experimental design. Furthermore, the religious character of the high school may increase gender socialization. As a result, the variation in stereotypical beliefs among students is expected to be rather low. This low variation makes it harder to find effects of the manipulation on students' stereotypical beliefs. If an effect is found in this setting, it will be even more likely the effect will be found in other school contexts.

The sample ($N = 410$) was drawn from the third and fourth grades from all educational levels.¹ Students' timetables determined which classes were able to participate in the survey. In consultation with the schedulers of the school, five classes from the lowest educational high school level, six classes from the intermediate level, and six classes from the highest educational level were selected. The survey experiment took place during mentoring hours in computer classrooms.

Respondents were presented a vignette describing the characteristics of a math teacher (Figure 3.2). The information on the vignette varied in terms of the gender of the teacher (female/male) and the self-confidence of the teacher (unconfident/confident), resulting in four experimental groups (Table 3.1). Based on the last digit of the students' six-digit student number (which is based on the moment of registration at the high school), students were randomly assigned to one of the four experimental groups.

Table 3.1. 2x2 Survey Experiment Design

		Gender Math Teacher	
		Female	Male
Self-Confidence	Unconfident	Group 1	Group 3
Math Teacher	Confident	Group 2	Group 4

[1.Anouk/2.Daan] is a math teacher on a high school. [1.She/2.He] has always found math very interesting. [1.She/2.He] achieved good grades for math in high school and then started studying mathematics. [1.She/2.He] teaches math to all levels for 5 years now, [a. but still finds it very scary. [1.She/2.He] stumbles over words, drops a chalk or gets red when [1.she/2.he] explains something./b. but still really enjoys it. [1.She/2.He] finds fun ways to make math problems less boring and stands with visible pleasure in front of the class.] [a. Nevertheless, [1.her/2.his] boss is happy with [1.her/2.him] and [1.she/2.he] hopes to continue teaching for a while./b. [1.Her/2.His] boss is happy with [1.her/2.him] and [1.she/2.he] hopes to continue teaching for a while.]

Figure 3.2. Text Vignette

3.3.1 MEASURES

Gender of the teacher

The gender of the teacher in the vignette was indicated by the names Anouk (popular Dutch female name) and Daan (popular Dutch male name) and the pronouns ‘she’ and ‘he’ (see Figure 3.2). A dummy variable was created with a value of ‘1’ for a female teacher and a value of ‘0’ for a male teacher.

Self-confidence teacher

Self-confidence of the teacher is measured by describing the teacher as insecure (using characteristics such as ‘turning red’, ‘dropping the chalk’, ‘stumbling words’) or self-confident (using characteristics such as ‘making math fun’, and ‘enjoying teaching’, see Figure 3.2). A dummy variable was created with a value of ‘1’ for a self-confident teacher and a value of ‘0’ for an unconfident teacher. A manipulation check confirms that this manipulation was successful ($r = 0.831$; $p < 0.01$).

Gender of the student

The gender of the respondent was measured by using the categories ‘male’, ‘female’, ‘I cannot say’, and ‘I do not want to say’. Two students chose the option ‘I cannot say’ and one student chose ‘I do not want to say’. These answers were coded as missing. A dummy variable was created with a value of ‘1’ for a female student and a value of ‘0’ for a male student.

Stereotypical belief

Stereotypical belief taps the extent to which respondents believe the idea that men are good at math and women are bad at math. Stereotypical belief is measured using eight items (for the exact items see Appendix A). Four items were based on the gender schema

measurement scale of Smetackova (2015). A pre-test of the survey demonstrated that the items based on the Smetackova (2015) measure were interpreted as too extreme, which might result in low variation. Therefore, four additional more nuanced statements were included. Respondents were asked to respond to the statements on a 5-point Likert scale (1 'totally agree', 5 'totally disagree'). The variable stereotypical belief is constructed by calculating the average of the items for each respondent ($\text{Alpha}=0.937$).

Academic self-concept

Academic self-concept taps the respondents' beliefs in and judgements of their mathematical abilities. To measure the academic self-concept, six statements from Arens et al. (2017) as well as four statements from the TIMMS study (for the exact items see Appendix A) were used.² Adding the more emotionally focused statements of the TIMMS study creates a measurement scale that measures both rational and emotional aspects of the academic self-concept. Respondents were asked to respond to the statements on a 5-point Likert scale (1 'totally true', 5 'totally false'). The variable academic self-concept is constructed by calculating the average of the items for each respondent ($\text{Alpha}=0.942$).

Average math grade

Because extant research shows that prior achievement in a subject greatly influences the academic self-concept for that subject (e.g., Marsh et al., 1988), the respondent's average grade for math will be included in the model as a covariate. In the Netherlands, most institutions grade exams, papers, and theses on a scale from 1 (very poor) to 10 (outstanding). Students were asked to look up their average grade for math (rounded to one decimal place) before filling in the survey. Students could answer '0' if they did not want to share their average grade, which is coded as missing. Adding average grade for math to the models will soak up noise in the dependent variable (academic self-concept) that might exist due to possible over (or under) estimation of one's mathematical abilities (c.f., Wang et al., 2017).

After a listwise deletion of respondents who have missing values on the variables used in the analyses, the total number of observations is 394. Table 3.2 shows descriptive statistics of and correlations between the variables in the analysis. The results show that there is no correlation between students' average math grade and the manipulations of gender of the teacher and teacher confidence, which is a prerequisite for using independent covariates in experimental designs (Wang et al., 2017).

3.4 RESULTS

To test the hypotheses, a moderated mediation analysis was carried out using STATA 15's 'sem' command. The direct effects of the moderated mediation analysis on academic self-concepts are presented in Table 3.3. In the first model, the effects of gender of the student (with male student as the reference category) and gender of the teacher (with male teacher as the reference category) are assessed as independent variables. In the second model, the effect of gender congruence is tested by adding a two-way interaction term for gender of the student and gender of the teacher. Finally, in the third model, the moderating effect of teacher confidence (with unconfident as the reference category) is tested by adding a three-way interaction term for gender of the student, gender of the teacher, and teacher confidence. The structure of the third model mirrors Hayes's (2018) conditional moderated mediation model (Model 13). For each model, bootstrapping (N = 5,000) is used to obtain normal-based bootstrapped confidence intervals around the indirect effect. Table 3.4 presents the indirect effects of the moderated mediation analysis.

Table 3.2. Descriptive Statistics and Correlations (N = 394)

	Mean	SD	Min.	Max.	1	2	3	4	5
1 Gender Student (Female = 1)	0.56	0.50	0	1	1.000				
2 Gender Teacher (Female = 1)	0.51	0.50	0	1	0.055*	1.000			
3 Confidence Teacher (Self-Confident = 1)	0.51	0.50	0	1	0.014	-0.015	1.000		
4 Average Math Grade	6.76	1.20	2.6	9.9	0.146*	0.014	-0.010	1.000	
5 Stereotypical Beliefs	2.36	0.90	1	5	-0.264*	-0.057*	0.008	-0.057*	1.000
6 Academic Self-Concept	3.39	0.88	1	5	-0.111*	0.086*	-0.011	0.560*	-0.154*

The first part of Model 1 tests the effect of gender of the student and gender of the teacher on the mediator variable, stereotypical beliefs, without the two-way interaction term. The results show that gender of the student has a statistically significant, negative effect on stereotypical beliefs, indicating that female students have weaker stereotypical beliefs about gender and math than male students. The effect of gender of the teacher is not statistically significant. Whether the teacher is male or female makes no difference in explaining students' stereotypical beliefs.

The second part of Model 1 tests the effect of gender of the student and gender of the teacher on the dependent variable, academic self-concepts. The variable stereotypical beliefs has a statistically significant, negative effect on academic self-concept. Hence, the stronger one's stereotypical beliefs about gender and math, the lower one's academic self-concept. The effect of gender of the student is also statistically significant, indicating that female students have a lower academic self-concept than male students. Moreover, the

results show that, in general, students with a female teacher have a higher academic self-concept than students with a male teacher. In addition, students with a higher average for math have a higher academic self-concept.

Table 3.3. Direct Effects of Moderated Mediation Analysis (N = 394)

	Model 1 Coef.	Model 2 Coef.	Model 3 Coef.
M = Stereotypical Beliefs			
Gender Student (1 = Female)	-0.472*** (0.092)	-0.531*** (0.124)	-0.674*** (0.184)
Gender Teacher (1 = Female)	-0.077 (0.088)	-0.143 (0.131)	-0.341+ (0.181)
Confidence Teacher (1 = Self-Confident)			-0.159 (0.171)
Average Math Grade	-0.014 (0.036)	-0.014 (0.035)	-0.015 (0.036)
Gender Student x Gender Teacher		0.119 (0.176)	0.427+ (0.248)
Gender Student x Confidence Teacher			0.277 (0.251)
Gender Teacher x Confidence Teacher			0.407 (0.264)
Gender Student x Gender Teacher x Confidence Teacher			-0.624+ (0.357)
Constant	2.754*** (0.242)	2.788*** (2.34)	2.883***
Y = Academic Self-Concept			
Stereotypical Beliefs	-0.176*** (0.044)	-0.176*** (0.044)	-0.176*** (0.044)
Gender Student	-0.439*** (0.075)	-0.433*** (0.099)	-0.433*** (0.103)
Gender Teacher	0.142* (0.069)	0.148 (0.102)	0.148 (0.101)
Average Math Grade	0.425*** (0.029)	0.425*** (0.029)	0.425*** (0.029)
Gender Student x Gender Teacher		-0.012 (0.139)	-0.012 (0.139)
Constant	1.106*** (0.237)	1.103*** (0.246)	1.103*** (0.241)
var (e.m)	0.756 (0.045)	0.756 (0.044)	0.749 (0.046)
var (e.y)	0.466 (0.032)	0.466 (0.031)	0.466 (0.032)

Note: Unstandardized coefficients, *** p<0.001, ** p<0.01, * p<0.05, + p<0.1

Hypothesis 1 predicts that gender congruence increases students' academic self-concepts, and Hypothesis 2 predicts that this effect is mediated by stereotypical beliefs. To test these hypotheses, an interaction term for gender of the student and gender of the teacher was added to Model 2. The results of Model 2 show that the effects of the interaction term on both stereotypical beliefs and academic self-concept are not statistically significant. Figure 3.3 depicts the predictive margins of stereotypical beliefs for the two-way interaction term. It shows that there is a significant difference in the stereotypical beliefs of male and female students, but this effect is not moderated by the gender of the teacher. Moreover, Table 3.4 shows that the indirect effect of the two-way interaction term on academic self-concepts is not statistically significant. Hence, gender congruence has no direct effect on academic self-concepts and this effect is not mediated by students' stereotypical beliefs, which means that Hypotheses 1 and 2 cannot be corroborated.

Hypothesis 3 predicts that the indirect effect of gender congruence on stereotypical beliefs is moderated by teacher confidence. For example, when confronted with an unconfident female teacher, girls' stereotypical beliefs will be stronger compared to situations in which female students are confronted with a confident female teacher. Model 3 tests the effect of the three-way interaction term for gender of the student, gender of the teacher, and teacher confidence on stereotypical beliefs. The results show that the three-way interaction term is statistically significant at the 90 percent confidence level. The coefficients for the independent variables that predict academic self-concept are, logically, the same as those in Model 2.³ Table 3.4 shows that the indirect effect of the three-way interaction term is also statistically significant at the 90 percent confidence level.

Table 3.4. Indirect Effects of Moderated Mediation Analysis (N = 394)

	Model 1 Coef.	Model 2 Coef.	Model 3 Coef.
Gender Student (1 = Female)	0.083** (0.025)	0.094** (0.030)	0.119** (0.039)
Gender Teacher (1 = Female)	0.014 (0.016)	0.025 (0.023)	0.060+ (0.033)
Confidence Teacher (1 = Self-Confident)			0.028 (0.030)
Average Math Grade	0.002 (0.006)	0.002 (0.006)	0.003 (0.006)
Gender Student x Gender Teacher		-0.021 (0.031)	-0.075+ (0.044)
Gender Student x Confidence Teacher			-0.049 (0.044)
Gender Teacher x Confidence Teacher			-0.072 (0.048)
Gender Student x Gender Teacher x Confidence Teacher			0.110+ (0.065)

Note: Unstandardized coefficients, *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, + $p < 0.1$

To better understand the effects of the interaction term, the predictive margins for the three-way interaction term are depicted in Figure 3.4. Figure 3.4 only depicts the effects for teacher-student gender congruence. It shows that, for both male and female gender congruence, the effect of teacher confidence on stereotypical beliefs is negative, but lacks statistical significance. In other words, the effect of the interaction between gender of the student and gender of the teacher is not moderated by teacher confidence. The fact that the three-way interaction in Model 4 is significant can in all likelihood be explained by the effect of the gender of the student. Hence, the indirect effect of gender congruence on academic self-concepts is not moderated by teacher confidence, which means that Hypothesis 3 cannot be corroborated.⁴

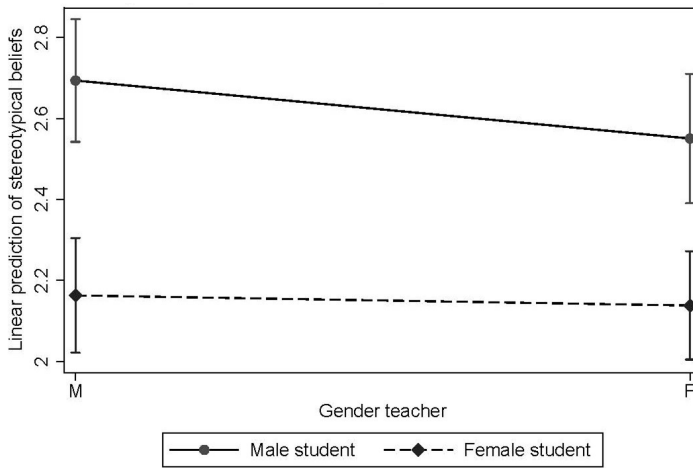


Figure 3.3. Adjusted Predictions of Two-Way Interaction with 90% Cis

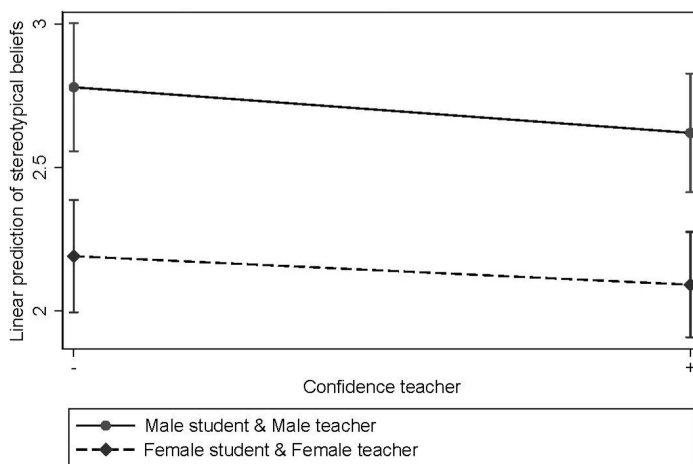


Figure 3.4. Adjusted Predictions of Three-Way Interaction with 90% Cis

3.5 DISCUSSION AND CONCLUSION

Integrating insights from educational and social psychology in representative bureaucracy theory, this study examined a mechanism through which individual level teacher-student gender congruence would lead to an increase in academic self-concepts of students in math in secondary education. A survey-experiment was conducted to test hypotheses on the association between teacher-student gender congruence, stereotypical beliefs about gender and math, and academic self-concepts. To soak up any noise in the dependent variable academic self-concept that might exist due to over estimation of students' academic self-concept, all models included students' average grade for math.

Based on the representative bureaucracy literature, it was hypothesized that gender congruence increases academic self-concepts for both female and male students, but no positive effects were found. Using insights from social and educational psychology it was further hypothesized that stereotypical beliefs mediate the relationship between teacher-student gender congruence and academic self-concepts and that this mediation is moderated by the self-confidence of the teacher. The analysis found no statistically significant interaction effects of gender of the teacher and gender of the student on both stereotypical beliefs and academic self-concepts. The effect of a three-way interaction of gender of the student, gender of the teacher, and teacher confidence on stereotypical beliefs was statistically significant; however, predictive margins showed no significant effects for the three-way interaction for teacher-student gender congruence specifically. In all, the analyses did not support the hypotheses.

The analyses showed a dominant effect of the gender of the student on both stereotypical beliefs and academic self-concepts. The difference in academic self-concepts for female and male students is consistent with the literature about gender differences in math (e.g., Marsch et al. 1988). As educational and social psychology suggest, stereotypical beliefs might explain differences in math performance of female and male students. These effects, however, are relatively modest which could be explained by the fact that average math scores were added to the models. The statistically significant relationships between gender of the student, stereotypical beliefs and academic self-concepts indeed strengthen the belief in a mediating role of stereotypical beliefs. The analyses also showed a positive effect of gender of the teacher on students' academic self-concept, meaning that both male and female students' academic self-concepts increase with a female teacher. Whereas studies on individual level effects of representation tend to focus on representation effects for the represented, the results of this study suggest that representation of stereotypically disadvantaged groups might be beneficial for both the represented and non-represented group.

Although this study found significant effects for the gender of the student on stereotypical beliefs, and significant effects for gender of the student and gender of the teacher on academic self-concepts of students, effects for teacher-student gender congruence on stereotypical beliefs and academic self-concepts were not found. The lack of support

for effects of teacher-student gender congruence might be caused by limitations of the study's methodology. Stereotypical beliefs were measured through items that explicated stereotypical beliefs. While, if explicitly asked, people tend to disagree with gender stereotypical beliefs to a higher degree, they may still have an implicit bias regarding the competencies of men and women in specific areas of expertise such as math (Steffens & Jelenec, 2011). Future research should therefore explore methods that allow for the measurement of implicit associations between gender and competencies in math, such as implicit association tests.

Moreover, lack of support for the hypotheses may be caused by the experimental design of the study. Stereotypical beliefs and academic self-concepts of students might be influenced by characteristics of their teacher as a result of the regular interactions students and teachers repeatedly have, over a longer time period. The survey experiment also might not have been able to capture a real-life teacher-student relationship but instead may have tested the effect of passive representation in the short term. The finding that representation does not have an effect in the short term raises the question if, on what terms, and after how many interactions effects of symbolic representation do occur. These considerations provoke interest in further testing of the proposed model with other methods.

This study contributed to representative bureaucracy literature by combining two theoretical ideas into a new model explaining individual level effects of symbolic representation. Although the hypotheses were not supported by the data, the results call for further exploration of the hypothesized mechanism. Significant gender differences in stereotypical beliefs and academic self-concepts and a significant effect of gender of the teacher on academic self-concepts of both male and female students provoke interest in further exploration of the role of teacher gender in aforementioned psychological processes in students. Such research could shed further light on the role stereotypical beliefs of citizen-clients may play in their reception of public services being delivered by public professionals with different backgrounds.

3.6 NOTES

1. In the Netherlands the educational system exists of primary education (one level, grade 1 to 8; age 4 till 12) and secondary education. Secondary education exists of three educational levels in which grades are starting from 1 again: VMBO (lowest level; prepares for vocational education, grade 1 to 4; age 12 till 16), HAVO (intermediate level; prepares for applied sciences, grade 1 to 5; age 12 till 17) and VWO (highest level; prepares for university, grade 1 to 6; age 12 till 18).
2. TIMMS & PIRLS (2015). About TIMMS 2015. Retrieved from: <https://timssandpirls.bc.edu/timss2015/>
3. The coefficients predicting academic self-concept slightly differ across the models due to bootstrapping.
4. Since students were exposed to the math teacher's gender and confidence at the same time, one could argue that the confidence of the teacher dampens the effect of gender congruence. As a robustness check, we ran the analyses on a subsample without the unconfident teachers. The results confirmed the results of the original analyses on the total sample.

3.7 APPENDIX

This appendix provides supplementary material as published on <https://www.journal-bpa.org/index.php/jbpa/article/view/64>. Appendix A provides information on the operationalization of the variables used in the analyses. Appendix B contains information about the survey and the survey questions used in the analyses.

3.7.1 APPENDIX A - OPERATIONALIZATION OF THE VARIABLES

Gender of the teacher

The gender of the teacher in the vignette was indicated by the names Anouk (popular Dutch female name) and Daan (popular Dutch male name) and the pronouns 'she' and 'he' (see Figure 3.2). A dummy variable was created with a value of '1' for a female teacher and a value of '0' for a male teacher.

Self-confidence teacher

Self-confidence of the teacher is measured by describing the teacher as insecure (using characteristics such as 'turning red', 'dropping the chalk', 'stumbling words') or self-confident (using characteristics such as 'making math fun', and 'enjoying teaching', see Figure 3.2). As a manipulation check, students had to rate the self-confidence of the teacher in the vignette. The manipulation check confirms that this manipulation was successful ($r = 0.831$; $p < 0.01$). A dummy variable was created with a value of '1' for a self-confident teacher and a value of '0' for an unconfident teacher.

Gender of the student

Students were asked their gender by the question: 'What is your gender?' with the answer options: 'male', 'female', 'I cannot say' and 'I do not want to say'. Two students chose the option 'I cannot say' and one student chose 'I do not want to say'. These answers were coded as missing. A dummy variable was created with a value of '1' for a female student and a value of '0' for a male student.

Stereotypical belief

Stereotypical belief is defined as belief in socially constructed ideas about skills and opportunities for women and men (e.g., Marsh et al., 1988). The stereotypical belief connected to math is the idea that men are good at math and women are bad at math. In the survey, students had to respond to eleven statements on a 5-point Likert-scale (1 being totally agree, 5 totally disagree). Five of the items were retrieved from the gender schema measurement scale created by Smetackova (2015). Only the items appropriate for measuring stereotypical beliefs connected to math were used from the initial nineitem scale created by Smetackova (2015). A pre-test of the survey demonstrated that the items based on the Smetackova (2015) measure were interpreted as too extreme, which might result

in low variation and non-response. Therefore, six additional more nuanced statements were included. After a factor and reliability analysis three items were removed due to cross loadings. Eight items for the measurement scale were retained (KMO measure of sampling adequacy is 0.913 and all KMO values for individual items are greater than 0.860. Cronbach's Alpha is 0.937).

The eight items used to construct the variable stereotypical belief are:

- men are often better at math than women (Smetackova, 2015)
- women do not have the 'right brain cells' for math (Smetackova, 2015)
- men are born with talent for math (Smetackova, 2015)
- men are better at logical thinking than women (Smetackova, 2015)
- I think that in general girls need to work harder for math than boys
- I think boys understand math faster than girls
- I think girls need more time to understand complicated math than do boys
- I think boys are better in solving complicated math problems than are girls

The variable stereotypical belief is constructed by calculating the average of the items for each respondent.

Academic self-concept

Academic self-concepts are defined as beliefs and judgements about one's performance in a subject (Pajares & Miller, 1994). In the survey students had to respond to twelve statements on a 5-point Likert-scale (1 being totally true, 5 totally false). Six of these statements were based on the measurement scale for academic self-concepts for math used by Arens et al. (2017). Arens et al. (2017) retrieved items for their measurement scale from the Project for the Analysis of Learning and Achievement in Mathematics (PALMA) that analyses adolescents' developments in math (Pekrun et al., 2007). The measurement scale of Arens et al. (2017) was complemented by six more emotionally focused statements from the Trends in International Mathematics and Science Study (TIMSS). TIMSS assesses mathematical and scientific achievement in multiple countries. After a factor and reliability analysis two items were removed due to cross loadings. Ten items for the measurement scale were retained (KMO measure of sampling adequacy is 0.939 and all KMO values for individual items are greater than 0.900. Cronbach's Alpha is 0.942).

The ten items used to construct the variable academic self-concept are:

- In math, I am a talented student (Arens et al., 2017)
- It is easy to understand things in math (Arens et al., 2017)
- I can solve math problems well (Arens et al., 2017)
- It is easy for me to write math tests (Arens et al., 2017)
- It is easy for me to learn something in math (Arens et al., 2017)

- If the math teacher asks me a question, I can answer it correctly most of the time (Arens et al., 2017)
- I worry often that the math class will be too hard for me (TIMSS)
- I am just not good in math (TIMSS)
- Math makes me nervous (TIMSS)
- It does not matter how hard I try, I receive bad grades for math (TIMSS)

The variable academic self-concept is constructed by calculating the average of the items for each respondent.

Average math grade

Before filling in the survey, students were asked to look up their average grade for math. Students filled in their average at the beginning of the survey (rounded to one decimal place). Students could answer '0' if they did not want to share their average grade, 13 students chose to not share their average.

3.7.2 APPENDIX B - SURVEY

Introduction of the survey: Informed consent

My name is Laura and I am a Master student at Leiden University. I would like to ask you some questions related to mathematics. It is important for my research that you answer the questions as honestly as possible. Read the questions carefully and give the answer most true to you. Your answers are anonymous, meaning that nobody will know what you have answered. Even I can't find out.

Completing the survey takes approximately 10 minutes. The survey starts when your press 'start the survey'. Should you encounter any difficulties while filling in the questionnaire, please raise your hand to ask me for help.

It is not mandatory to participate in the research. If you do not want to participate, do not start the survey. I will come to you as soon as possible to learn why you would rather not participate. One student chose to not participate in the research.

Questions and coding

1. What is your average grade for math? Respondents entered average grade rounded to one decimal place. Respondents entered '0' if they did not want to share their average, these were coded as missing.
2. What is your gender? Answer options: a) Male; b) Female; c) I cannot say; d) I do not want to say. Answers c and d were coded as missing. Male was coded 0; female 1.

Vignette and coding

1. [Anouk] is a math teacher on a high school. [She] has always found math very interesting. [She] achieved good grades for math in high school and then started

studying mathematics. [She] teaches math to all levels for 5 years now, [but still finds it very scary. [She] stumbles over words, drops a chalk or gets red when [she] explains something.] [Nevertheless], [her] boss is happy with [her] and [she] hopes to continue teaching for a while.

2. [Anouk] is a math teacher on a high school. [She] has always found math very interesting. [She] achieved good grades for math in high school and then started studying mathematics. [She] teaches math to all levels for 5 years now, [but still really enjoys it. [She] finds fun ways to make math problems less boring and stands with visible pleasure in front of the class.] [Her] boss is happy with [her] and [she] hopes to continue teaching for a while.
3. [Daan] is a math teacher on a high school. [He] has always found math very interesting. [He] achieved good grades for math in high school and then started studying mathematics. [He] teaches math to all levels for 5 years now, [but still finds it very scary. [He] stumbles over words, drops a chalk or gets red when [he] explains something.] [Nevertheless], [his] boss is happy with [him] and [he] hopes to continue teaching for a while.
4. [Daan] is a math teacher on a high school. [He] has always found math very interesting. [He] achieved good grades for math in high school and then started studying mathematics. [He] teaches math to all levels for 5 years now, [but still really enjoys it. [He] finds fun ways to make math problems less boring and stands with visible pleasure in front of the class.] [His] boss is happy with [him] and [he] hopes to continue teaching for a while.

Gender of the teacher was coded as 0) male; 1) female. Self-confidence of the teacher was coded as 0) unconfident; 1) confident.

Manipulation check confidence of the teacher

On a scale from 1 to 10, how would you rate the self-confidence of the teacher in the vignette (1 being very unconfident, 10 being very confident)? Respondents entered a number between 1 and 10.

Statements and coding

Academic self-concept: Read the following statements and indicate to what extent the statements are true to you. Answer options: totally true, true, not true and not false, false, totally false

1. In math, I am a talented student
2. It is easy to understand things in math
3. I can solve math problems well
4. It is easy for me to write math tests
5. It is easy for me to learn something in math

6. If the math teacher asks me a question, I can answer it correctly most of the time.

Read the following statement and indicate to what extent the statements are true to you.

Answer options: totally true, true, not true and not false, false, totally false

7. I worry often that the math class will be too hard for me
8. I am just not good in math
9. Math makes me nervous
10. If I do my best I can receive good grades in math
11. If I wanted to I could receive good grades in math
12. It does not matter how hard I try, I receive bad grades for math.

Statements 1-6, 10 and 11 were recoded so that a higher score means a more positive academic self-concept.

Stereotypical beliefs

Read the following statements and indicate to what extent you agree with the statements.

Answer options: totally agree, agree, do not agree and do not disagree, disagree, totally disagree

1. I think that in general girls need to work harder for math than boys
2. I think math is easier for girls than for boys
3. I think boys need to practice longer to understand math problems than girls
4. I think boys understand math faster than girls
5. I think girls need more time to understand complicated math than do boys
6. I think boys are better in solving complicated math problems than are girls

Read the following statements and indicate to what extent you agree with the statements.

Answer options: totally agree, agree, do not agree and do not disagree, disagree, totally disagree

7. men are often better at math than women
8. women do not have the 'right brain cells' for math
9. men are born with talent for math
10. men are better at logical thinking than women
11. men and women do not differ in their capabilities

Statements 1, and 4-10 were recoded so that a higher score means stronger stereotypical beliefs.



CHAPTER **FOUR**

Understanding the symbolic effects of gender representation: A multi-source study in education

Previously published as: Doornkamp, L., Groeneveld, S., Groeneveld, M.G., Van der Pol, L.D., & Mesman, J. (2022): Understanding the Symbolic Effects of Gender Representation: A Multi-Source Study in Education, *International Public Management Journal*, DOI: 10.1080/10967494.2021.1991534

ABSTRACT

Symbolic effects of bureaucratic representation assume that outcomes for citizen-clients change in response to the mere presence of bureaucrats with similar backgrounds. The social-psychological mechanisms in clients that may explain these changes are barely examined, though. Based on multi-source data on male and female high school students in the Netherlands, this study empirically tests a theoretical model that links student-teacher gender congruence to students' performance in math and Dutch language through the mediating role of students' gender-stereotypical beliefs and self-perceived abilities. The empirical analyses do not support the hypotheses. For student performance in math, no effects were found, while for Dutch language, a negative gender-representation effect was found for male students. In all, the results point at divergent relations between gender congruence, self-perceived abilities and performance for male and female students across subjects. The null results in this study suggest that symbolic effects of gender representation may be affected by national context and socio-economic status which should be considered as new frontiers for future research on representative bureaucracy.

4.1 INTRODUCTION

Despite the major expansion in studies on representative bureaucracy in the past decades, little is known about the individual-level mechanisms explaining the effects of representation in public organizations (Meier, 2019). Representativeness of public organizations is associated with inclusiveness in decision-making which contributes to the legitimacy and effectiveness of public policy (e.g., Sowa & Selden, 2003). The majority of empirical studies on representation effects in public organizations typically focuses on an aggregate level, modeling increased output for a group of clients based on the percentage of employees in the public organization from that same group. This aggregate focus fails to identify the individual-level mechanisms that drive representation effects (Meier, 2019; Nicholson-Crotty et al., 2016; Zhang, 2019). For decades, research assessments have found the existence of representation effects and subsequently ascribed it to *assumed* mechanisms, mostly situated at the individual level (Nicholson-Crotty et al., 2016).

The study of individual-level mechanisms has intensified in recent years. However, these studies have mostly taken the perspective of the bureaucrat, focusing on the behavior of public officials to explain representation effects (e.g., Bardbury & Kellough, 2011; Keiser et al., 2002; Nicholson-Crotty et al., 2016; Sowa & Selden, 2003; Wilkins & Keiser, 2006; Wilkins, 2007). Only a handful of studies have taken the perspective of the client, explaining representation effects in terms of changes in attitudes and behaviors of the client as a result of the mere presence of a bureaucrat from a similar background (Guul, 2018). Clients matched with a bureaucrat sharing the same characteristics were found to rate higher program satisfaction (Gade & Wilkins, 2013), increase job-seeking efforts (Guul, 2018), perceive bureaucratic actions as more legitimate (Theobald & Haider-Markel, 2008) and as more trustworthy and fair (Ricucci et al., 2014). Although these studies measured changes in attitudes and behaviors in clients, the underlying (un- or subconscious) psychological processes producing these changes remain unknown. In this study, we theorize and empirically test a psychological mechanism on the individual level from the perspective of the citizen-client that so far has been under researched (Figure 4.1). In doing so, we try to disclose one of many possible pathways for symbolic representation, contributing to the understanding of representation effects.

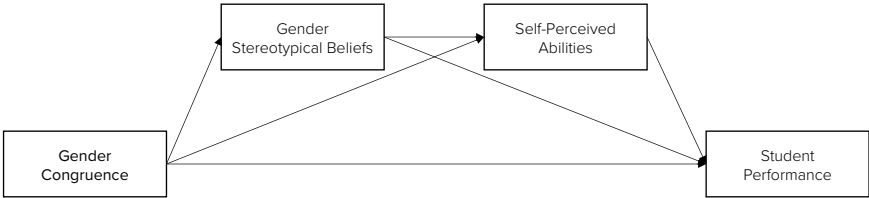


Figure 4.1. Research Model

Like many previous studies on representative bureaucracy (e.g., Dhillon & Meier, 2020; Grissom et al., 2015; Meier & Stewart, 1992; Meier et al., 1999; Nicholson-Crotty et al., 2016; Vinopal, 2018), we empirically test our proposed model in the context of schools, more specifically in Dutch secondary education. We focus on the effects of gender representation on male and female high school students' performance in the subjects of math and Dutch language. Previous studies on this topic, that focused on the effect of female teachers on female students' math performance, found positive effects of gender congruence (Agyapong, 2018; Keiser et al., 2002; Song, 2018; Zhang, 2019). We added to this stream in the literature by means of a survey-experimental study on the effect of the gender of a math teacher on stereotypical attitudes and perceived abilities of male and female students, finding an effect of brief hypothetical exposure to a female math teacher on both male and female students' academic self-concept (Doornkamp et al., 2019). The current study builds on this study, elaborates on its theoretical model and tests new hypotheses based on multi-source data including data on student performance of 329 high school students and their teachers in the Netherlands. In doing so, it acknowledges that empirical research should include real-life and enduring interactions to be able to detect attitudinal and behavioral changes in citizens-clients, if any. Our research is guided by the question: *what psychological processes in students can explain the association between student-teacher gender congruence and student performance?*

This study's contribution is threefold. First, this study empirically tests a theoretical mechanism underlying symbolic representation in which the psychological processes in the citizen-client, not the bureaucrat, are central. Second, in so doing, this study introduces social cognitive theory to representative bureaucracy literature which provides a basis for the study of social-psychological mechanisms in clients to explain representation effects in a variety of bureaucratic contexts. Third, while effects of symbolic representation are often tested through survey-experimental research, if tested at all, this study uses multi-source observational data that improves the ecological validity of the research.

The paper proceeds as follows: the next section elaborates on the theoretical framework. Then the study context as well as its research design and methods are described, followed by the study's results. The paper concludes with the discussion of the findings and outlines avenues for future research.

4.2 REPRESENTATIVE BUREAUCRACY

The concept of representative bureaucracy was coined by Kingsley (1944), who argued that bureaucracies are part of the democratic constitutional state and should therefore comply with democratic principles. Consequently, bureaucracies' demographic composition *should* reflect the demographic composition of the people they serve. Besides the normative argument for representative bureaucracies, a large body of research found that

public organizations with high levels of representativeness perform better than their less-representative counterparts (see Kennedy, 2014; Meier, 2019 for a review of normative and theoretical underpinnings, see Groeneveld & Van de Walle, 2010; Riccucci & Van Ryzin, 2017 for state of the art).

The representative bureaucracy literature distinguishes between passive, active, and symbolic representation. Passive representation refers to the extent to which the socio-demographic composition of the workforce of a public organization mirrors its clientele. Research on passive representation shows disparity in hierarchical and functional representation for minorities. Moreover, public organizations with higher levels of minority representation were found to serve important democratic values, meet fairness and equality goals, and perform better (Riccucci & Van Ryzin, 2017). In schools, research on passive representation has found that representativeness of the workforce is associated with fewer drop outs (Pitts, 2007), more inclusive school policies (Roch et al., 2010), and better performance for female and ethnic minority students (Agyapong, 2018; Dhillon & Meier, 2020; Keiser et al., 2002; Meier & Stewart, 1992; Meier & Bothe, 2001).

These outcomes of representativeness are generally assumed to originate on the individual level through the interaction between the bureaucrat and the citizen-client. Most studies that focus on this individual level take the perspective of the bureaucrat, explaining representation effects in terms of a bureaucrat acting in concordance with the values or interests of a client because of their similar social origins and shared values. This is referred to as active representation (e.g., Meier et al., 2006; Meier et al., 2005; Park, 2012). For bureaucrats to be able to engage in this minority representative role-taking behavior, bureaucrats need discretionary room and must experience salience of the minority status to the policy area at hand (Keiser et al., 2002).

In education, teachers have discretion in shaping and teaching their classes. Additionally, characteristics such as ethnicity and gender are salient as ethnic minorities and women are historically disadvantaged in education (Keiser et al., 2002). Research on representativeness in education shows that contextual factors such as more class days, longer teaching hours, and smaller student-teacher ratios enhance representation effects assuming active representation plays a role (Dhillon & Meier, 2020; Song, 2018). Minority teachers may spend more time on the students they represent and generate more positive reinforcements (Grissom et al., 2015; Keiser et al., 2002; Zhang, 2019). Moreover, Nicholson-Crotty et al. (2016) demonstrated that minority teachers were more likely to assign gifted services to minority students than their nonminority counterparts.

Symbolic representation shifts the focus from the bureaucrat to the citizen-client. Representation effects are explained in terms of changes in the attitudes and behaviors of the citizen-client as a result of the mere presence of the bureaucrat (Riccucci et al., 2018). Some studies empirically examined and found evidence for changes in attitudes and behaviors of the citizen-client as a result of being matched with a bureaucrat sharing the same characteristics as them (Gade & Wilkins, 2013; Guul, 2018; Riccucci et al., 2014;

Theobald & Haider-Markel, 2008). While studies on representation that were situated in education have not been able to empirically unravel the effects of symbolic and active representation, some studies suggest symbolic representation is likely to occur. A qualitative study in education in China conducted by Xu and Meier (2021) suggests that teachers do not actively pursue the interests of the students they represent and that representation effects in education in China primarily result from symbolic representation. Vinopal (2018) found that parents of minority students were more involved in the presence of a minority teacher than in the presence of a nonminority teacher which could be a result of both symbolic and active representation. Zhang (2019) found a small and partial effect of female teachers on female students' math performance mediated by their gender-stereotypical beliefs. While Zhang (2019) interprets this result as evidence for active representation, this may also reflect effects of both active and symbolic representation. In a previous study in Dutch education we added to this stream in the literature by measuring attitudinal changes in students by means of a vignette experiment that allowed to single out the effect of symbolic representation. Although we found that students' academic self-concept is enhanced in the presence of a female teacher, this was found for both male and female students (Doornkamp et al., 2019).

Although studies on gender representation in education have mostly focused on effects of gender congruence on female students, the logic of representation might also apply to male students. Although Zhang (2019), in his study on gender representation in Chinese middle schools, did not find representation effects for male students in math, English, or Chinese scores, studies on the gender gap in education have found positive effects of same-sex teachers on the academic performance of both female and male students (e.g., Dee, 2005).

Based on the above, the first hypothesis can be formulated:

H1: Student-teacher gender congruence has a positive effect on student performance.

Despite increased attention for symbolic representation in the study of representative bureaucracy, still little is known about *how* symbolic representation actually works. How do attitudinal and behavioral changes emerge in the citizen-client as a result from the presence of a bureaucrat with whom one shares similar socio-demographic characteristics? This is theoretically explored in the next section.

4.2.1 SOCIAL COGNITIVE PROCESSES AS MECHANISM UNDERLYING SYMBOLIC REPRESENTATION EFFECTS

Social-cognitive theory (SCT) emphasizes the role of the environment in a person's knowledge acquisition and personal development; by observing and interacting with others people learn about social norms and structures that influence attitudes, beliefs, and

behaviors. As this study focusses on gender representation (i.e., the influence of the student-teacher gender congruence on students' performance) we will discuss SCT in the context of gender development. Note that SCT can be applied in several contexts and therefore might contribute to the understanding of bureaucratic representation more broadly.

SCT of gender development emphasizes the role of the environment in a person's knowledge acquisition about the sexes to identify males and females, and consequently themselves (Blakemore et al., 2009). By observing and interacting with men and women an individual learns about existing gender norms and rules. Bussey and Bandura (1999) distinguish three ways through which an individual learns: direct information, social consequences to gender related behaviors, and modeling. The last one, modeling, is the most pervasive and powerful as it is the fastest to cognitively process (see Bussey & Bandura, 1999 for an extensive explanation). In short, by observing models (i.e., men and women in the environment) people learn about what is possible and appropriate for men and women. This gender-linked information generates new forms of attitudes, beliefs, and behavior that conform to these gender rules and structures (Bussey and Bandura, 1999).

Following SCT reasoning, symbolic gender representation can be expressed in terms of bureaucrats functioning as a model influencing a client's conception of what is appropriate and possible for him or her as a man or a woman changing his or her attitudes, beliefs, and behavior producing representation effects. Just by being a man or woman, each and every model (including bureaucrats) in the environment of a client can have a greater or lesser impact on a client's gender development. Applied to the school context, teachers can pass along information about what is academically appropriate and possible for male and female students changing students' attitudes, beliefs and behaviors in education.

4.2.1.1 Self-Perceived Abilities

Important to SCT of gender development is the role of models on people's self-perceived abilities, i.e., believe in the capability to master given levels of attainment (Bussey and Bandura, 1999). The concept of self-perceived abilities is often used to grasp self-efficacy beliefs in an educational setting. Self-perceived abilities of students are found to be strongly associated with academic achievement (Marsh et al., 2005). Although the causality in this relation does not work in one direction (self-perceived abilities can predict academic achievement and vice versa), studies show the predictive power of self-perceived abilities in the prediction of academic achievement (Chamorro-Premuzic et al., 2010, Spinath et al., 2006). A large body of research shows that an individual's beliefs and judgements about his or her own capabilities to perform a certain academic task (i.e., self-perceived abilities), can predict actual performance in that task (e.g., Marsh et al., 1988; Chamorro-Premuzic et al., 2010; Greven et al., 2009; Wach et al., 2015). Even when controlled for general mental ability (intelligence), self-perceived abilities contribute to the prediction of academic achievement (Spinath, et al., 2006; Greven et al., 2009; Wach et al., 2015). Moreover, motivational influences such as self-perceived abilities can be stronger predictors of

academic achievement than cognitive influences in some subjects (Weber et al., 2013).

The gender of a teacher can affect students' self-perceived abilities through their success of having become a teacher as a man or a woman, passing along gender-linked information about what the sexes are capable of in that subject. Social-cognitive theory argues that as soon as an individual is aware of his or her gender, success or failure of someone similar to themselves (man or woman) can influence the individual's beliefs and judgements to master comparable activities (Bussey & Bandura, 1999). As students are aware of their own gender, students can perceive the success of teachers similar to themselves as a representation of their own capabilities in a subject affecting their self-perceived abilities (Bussey & Bandura, 1999). These strengthened self-perceived abilities positively affect their actual performance in a subject (e.g., Wach et al., 2015). This leads to hypothesis 2.

H2: Self-perceived abilities mediate the association between student-teacher gender congruence and student performance.

4.2.1.2 Gender-Stereotypical Beliefs

'Stereotypes reflect general expectations about members of particular social groups' (Ellemers, 2018: .276). Gender-stereotypical beliefs are an individual's ideas about attributes of men and women based on these generalized preconceptions (Bussey & Bandura, 1999). Gender-stereotypical beliefs in individuals emerge through the internalization of repeated modeling of gender-typed behavior in the environment of an individual (Bussey & Bandura, 1999; Spencer et al., 2016).

The gender of a teacher can affect students' gender-stereotypical beliefs by modeling a (contra) gender-stereotypical pattern strengthening or weakening students' gender-stereotypical beliefs connected to a subject. For math and Dutch languages, the subjects of the current study, stereotypes prescribe the capabilities for men and women; men are good at math, women are good at Dutch language. Having a male or female teacher for a subject that is generally seen as either masculine (e.g., math) or feminine (e.g., language), provides students with a stereotypical or contra-stereotypical example. By confirming gender-prototypic behaviors (for instance a male math teacher) students' gender-stereotypical beliefs can become stronger, whereas models contradicting gender-prototypic behaviors (female math teacher) can reduce gender-stereotypical beliefs.

These stereotypical beliefs can affect performance through the internalization of stereotypes reflected in self-perceived abilities (Plante et al., 2013). Stereotypes can either facilitate or obstruct optimal performance (i.e., the stereotype 'men are good at math, women are bad at math' can facilitate male students' performance, whereas female students' performance can be obstructed), the negative obstructing effect is assumed to be easier to internalize and more resistant to disconfirmation than the positive facilitating effect of stereotypes (Baumeister et al., 2001). The psychological state in which students believe that they belong to a group stereotypically known for its performance (deficits) affects students'

beliefs and judgements about their own capabilities to succeed (i.e., self-perceived abilities) affecting their performance (Beilock et al., 2010; Spencer et al., 2016). Similarly, research on self-efficacy beliefs and academic achievement propose stereotypical beliefs as an explanation for gender differences in self-efficacy beliefs that cannot be explained based on prior achievement (Marsh et al, 1988; McFarland et al., 2016; Pajares & Miller, 1994; Wach et al., 2015).

Thus, math and Dutch language teachers can affect student performance not only through being perceived as a representation for what is achievable for the sexes, but also through being a negation or a confirmation of the stereotypes connected to the subject affecting students' gender-stereotypical beliefs and self-perceived abilities. Based on the above reasoning, we formulate hypothesis 3.

H3: Stereotypical beliefs mediate the indirect effect of self-perceived abilities in the association between student-teacher gender congruence and student performance.

4.3 STUDY DESIGN AND METHOD

4.3.1 STUDY CONTEXT

This study is part of the longitudinal study Girls in Science running from 2017 until 2022 in the Netherlands.¹ Although the Netherlands ranks in the highest quartile of the Global Gender Gap Index Rating indicating that gender inequality is relatively mild (Global Gender Gap Report, 2020), gender-stereotypical beliefs still play an important role in Dutch society. While Dutch men and women are equal under the law and gender equality has improved over the last decades, Dutch society is still characterized by gender role segregation (Daalmans et al., 2017). For instance, educational and occupational careers of men and women in the Netherlands display gender-stereotypical patterns (Onderwijsraad, 2019). Female students are still underrepresented in technical and mathematical studies and occupations, whereas male students are underrepresented in education and occupations in the healthcare-sector (Onderwijsraad, 2019). Moreover, 60% of Dutch women who are on the labour market work part-time, this is three times more than Dutch men resulting in a pay gap, unequal distribution of caretaking tasks, and more (see OESO, 2019). Hence, we believe gender is a highly salient status characteristic in Dutch education.

4.3.2 DATA COLLECTION

The research project Girls in Science collects data on children and adolescents from 10 to 20 years old to examine gender socialization in the family and school context. This study reports on school-related data from the first wave (collected in 2019-2020) and includes adolescents of on average 15.44 ($SD = 0.33$) and 17.49 ($SD = 0.53$) years old.

Both adolescents and their parents participated in the study. Eligible families in the Western and Middle region of the Netherlands were selected from municipality records. Inclusion criteria were biological parents living on the same address with a first child of 17 years old and a second child of 15 years old with a maximum age difference of 36 months. Exclusion criteria were severe psychical or intellectual handicaps of parent or child, being born outside the Netherlands and/or not speaking the Dutch language. Eligible families were invited in writing (letter and brochure) to participate in the study. Families were able to respond to the invitation by returning an answering card or responding via the website of the research project. In contrast to comparable studies in the field of pedagogy and social psychology, this study's recruitment method allows for a calculation of the response rate. Of the 3,316 eligible families, 6% were willing to participate ($N = 199$) whereas 14% indicated not to be willing to participate in the study. The rest of the families (80%) did not respond to the invitation. Families may have been reluctant to participate due to the relatively significant effort that participation required from them.

Firstly, father, mother, first and second child of the families were requested to each complete an online set of questionnaires which took 30-40 minutes. Secondly, the family was visited in a two-hour home visit. The home visit included interviews with both children, questionnaires, computer tasks, and different interaction-tasks which were filmed. At the end of every home visit pictures were taken of report cards of recent grades and of the final test of primary school of both children. The extensive data collection resulted in a comprehensive dataset on adolescents and their family and school context.

The adolescents reported on in this study were visited between September 2019 and February 2020. These adolescents were in, or graduated from, secondary education, making this sample most suitable for this study's analysis as the proposed mechanism requires subject specific teachers, rather than general primary teachers. The vast majority of the adolescents (86%) were in secondary education, the remaining 14% graduated from secondary education, which was no more than one and a half year ago, reported on their experiences in secondary education. 62% of the adolescents reported on in this study were in or graduated from the highest educational level of the Dutch secondary school system, 24% were in or graduated from the intermediate level, and 14% were in or graduated from the lowest level.² The average national distribution across the educational levels is 22% in the highest level, 22% in the intermediate level, 50% in the lowest level and 6% other (Dutch Report Education Inspectorate, 2019). Our sample is thus relatively highly educated, as are the mothers (average age of 48.64; $SD = 3.38$, at the moment of the home visit) and fathers (50.93; $SD = 4.68$) of the adolescents in our sample. 76% of the mothers and 72% of the fathers have obtained a bachelor's degree at the University of Applied Sciences or higher compared to 33% in the Dutch population (Education in Numbers 2020). In sum, the adolescents in our sample are members of families of a relatively high socio-economic status.

Only the adolescents with scores on all relevant variables are included in the study resulting in $N = 329$ (42% male). Although these adolescents are nested in families, intra-

class correlation coefficients show that the educational data of siblings are independent ($ICC < 0.05$ for both subjects). The independence indicates that grades in Dutch language and math of adolescents within one family differ as much as grades from adolescents belonging to different families. As the design effect is smaller than 2 ($deff = 1.03$), the analyses can be performed using general linear modeling (Peugh, 2010).

4.3.3 MEASURES

This study makes use of measures included in the online questionnaire as well as measures included in the structured interview and computer task conducted during the home visits.

Gender Congruence

The independent variable indicating student and teacher having the same gender, is a binary variable with 1 = gender congruence and 0 = no gender congruence. Students indicated their gender in the online questionnaire.³ The gender of their teachers in math and Dutch language were asked during the structured interview in the home visit (e.g., 'What is the gender of your math teacher?'). The students expressed their teachers' gender as either male or female. Only a few students had both a male and female teacher for a subject in one year or, when graduated, forgot (the gender of) their teacher; these cases are regarded as missing data ($N = 10$).

The distribution of male and female teachers across math and Dutch language is skewed, but similar for female and male students. 75% of the students in our sample has a female Dutch language teacher, in respect to 25% with a male Dutch language teacher. In math, 38% of the students has a female math teacher, whereas 62% has a male math teacher. The distribution of female and male teachers over the subjects in our sample is comparable with the distribution in the Netherlands. Dutch language teachers in the Netherlands are disproportionately female (72%). For math, the distribution is slightly more balanced with 58% male and 42% female teachers (numbers based on open educational data 2018: DUO, 2020). The skewed distribution is not expected to introduce bias, because teachers do not select their students (or vice versa).

Student Performance: Grade in Math and Grade in Dutch Language

Grade for math and grade for Dutch language are the calculated averages ranging from 1.0 to 10.0 obtained from photographed final exams (14%), secondary school reports (58%), or screenshots from online grade administration systems (28%). For some students, grades are missing because reports or screenshots were not available. These cases are regarded as missing data ($N = 17$). Additionally, reports with international grades (scale 1 to 7) ($N = 6$), letters instead of numbers ($N = 6$), or assessments in text ($N = 2$) were regarded as missing data. Also, reports that stem from one and a half year before the home visit ($N = 4$), wrongly photographed reports ($N = 1$), reports without grades for math ($N = 9$) or Dutch language ($N = 4$), or both ($N = 2$) were regarded as missing data.⁴

Self-Perceived Abilities

The first mediator in the analyses is students' beliefs and judgements about their capabilities in math and Dutch language. The self-perceived abilities are measured using three subject-specific items completed on a 5-point Likert-scale (*how good do you think you are at [...]*) with 1 = very bad until 5 = very good. The full scale consists of twelve items representing beliefs and judgements about students' capabilities in four subjects (Math, Dutch language, humanities and science). The scale is an translation of the math, first language and science items used in earlier research on self-perceived abilities of adolescents (Chamorro-Premuzic et al., 2010; Spinath et al., 2006), and is complemented with items for humanities.

The subscale for math includes the items 1) multiplying and dividing, 2) calculating with capacity measures, and 3) estimating the outcome of a sum with large numbers. The subscale for Dutch language includes 1) text comprehension, 2) grammar, and 3) writing a text. The variables 'self-perceived abilities in math' and 'self-perceived abilities in Dutch language' are constructed by calculating the average of the items for each respondent. Cronbach's Alpha is 0.83 for math and 0.64 for Dutch language.

Gender-Stereotypical Beliefs

The second mediator is assessed by a computerized version of the arts-science Implicit Association Test (IAT) (Nosek et al., 2002a).⁵ The IAT was built with E-prime 3.0 based on the task on the Harvard project Implicit demonstration website and the Nosek et al. (2002a) paper. The IAT was conducted on laptop computers and reaction time and accuracy were automatically recorded. Each adolescent's level of implicit stereotypes was determined by calculating differences in reaction time and accuracy in congruent and incongruent blocks resulting in a d-score (using scoring algorithm by Greenwald et al., 2003). A positive d-score represents stronger implicit gender-stereotypical beliefs (e.g., math is for males, Dutch language for females). Negative d-scores represent contra-stereotypical beliefs (e.g., math is for females, Dutch language for males). To reduce possible order effects of the presentation of congruent and incongruent blocks, two versions of the IAT were used, one in which the congruent block was administrated first and one in which the incongruent block was administrated first (Nosek et al., 2005). Adolescents starting with the congruent blocks had significantly higher differences scores ($p=0.000$). We do not expect that this will introduce any bias as adolescents were randomly assigned to the version starting with congruent blocks or the version starting with incongruent blocks. The randomization ensured that the different gender/teacher combinations in math and Dutch language included about the same number of students who started with a congruent version as students who started with an incongruent version.

4.3.4 ANALYTICAL STRATEGY

Previous research on gender stereotypes in education has mostly focused on math, assessing whether female students suffer from the stereotype ‘men are better at math than women’ (e.g., Steele, 2003). Some studies include language education to examine the role of the stereotype assuming that ‘women are better with language than men’ (Plante et al., 2013; Wach et al., 2015). In this study, we will include both subjects (math and Dutch language), estimating the effect of gender congruence through gender-stereotypical beliefs and self-perceived abilities for male and female students in math and Dutch language. As the effect of gender congruence on student performance through gender-stereotypical beliefs and self-perceived abilities is expected to vary for male and female students in math and Dutch language, the hypotheses testing will be split in four separate analyses for male and female students in math and Dutch language. In math, it is expected that female math teachers *reduce* female students’ gender-stereotypical beliefs, strengthening their self-perceived abilities and increasing their performance, whereas for male students, male math teachers are expected to *reinforce* gender-stereotypical beliefs strengthening the students’ self-perceived abilities, and increasing their performance. In Dutch language, it is expected that female Dutch language teachers *reinforce* female students’ gender-stereotypical beliefs, strengthening their self-perceived abilities and increasing their performance, whereas for male students, male Dutch language teachers are expected to *reduce* gender-stereotypical beliefs strengthening their self-perceived abilities, increasing their performance.

Table 4.1. Descriptive Statistics

	Mean (SD)			Correlations Math (p-value)			Correlations Dutch language (p-value)		
	Full Sample N = 329	Male Students N = 138	Female Students N = 191	GC	Grade	SPA	GC	Grade	SPA
Math	GC	0.47 (0.500)	0.61 (0.490)						
	Grade	6.828 (1.176)	6.690 (1.191)	-0.085 (0.124)					
	SPA	3.565 (0.937)	3.911 (0.807)	0.055 (0.324)	0.167** (0.002)				
	GSB	0.317 (0.402)	0.297 (0.408)	-0.066 (0.234)	0.058 (0.297)	0.017 (0.754)			
	GC	0.51 (0.501)	0.21 (0.409)	-0.063 (0.257)	0.025 (0.652)	-0.143** (0.009)	0.052 (0.346)		
Dutch	Grade	6.607 (0.854)	6.339 (0.826)	-0.032 (0.563)	0.421** (0.000)	-0.059 (0.284)	0.079 (0.151)		
	SPA	3.431 (0.775)	3.302 (0.783)	-0.008 (0.878)	0.029 (0.598)	-0.041 (0.464)	0.034 (0.539)	0.368** (0.000)	
	GSB	0.317 (0.402)	0.297 (0.408)	-0.066 (0.234)	0.058 (0.297)	0.017 (0.754)	0.052 (0.346)	0.055 (0.324)	-0.034 (0.537)
	GC	0.51 (0.501)	0.21 (0.409)	-0.063 (0.257)	0.025 (0.652)	-0.143** (0.009)	0.052 (0.346)		

Note: GC = Gender Congruence, SPA = Self-Perceived Abilities, GSB = Gender Stereotypical Beliefs ** p<0.01, * p<0.05, + p<0.1

4.4 RESULTS

Table 4.1 presents the means and standard deviations for the full sample and split by adolescent gender. In addition, correlations for math and Dutch language for the full sample are presented. Significant correlations were only found for self-perceived abilities and grades in both math and Dutch language.

The main results for the first hypothesis, that, student-teacher gender congruence has a positive effect on student performance, are presented in Table 4.2. ANOVA analyses were conducted to assess variation in math and Dutch language grades for the different combinations of the student-teacher gender dyad.

Table 4.2. Differences in Grades for Student-Teacher Gender Combinations (N = 329)

	Mean Math Grade (SD)	Mean Dutch language Grade (SD)
Male student x Male teacher	6.639 (1.261)	6.025 (0.806) ^a
Male student x Female teacher	6.768 (1.081)	6.423 (0.815) ^b
Female student x Male teacher	6.991 (1.154)	6.779 (0.801) ^c
Female student x Female teacher	6.820 (1.166)	6.809 (0.835) ^c
Significance Between Groups	p = 0.203	p < 0.001

Note: means with different superscripts are significantly different from each other, p < 0.05

For math, Table 4.2 reveals no statistically significant differences in grades for the various student-teacher combinations. Contrary to what was expected, student math performance is not related to having a teacher of the same gender.

A different pattern applies for Dutch language. Table 4.2 shows statistically significant variation in Dutch language grades for student-teacher gender combinations. Post-hoc analyses were conducted to test the hypothesized association. The analyses reveal that male students with a male Dutch language teacher achieve 0.399 (p = 0.021) lower average grades in Dutch language compared to male students with a female teacher. Gender congruence thus has a negative effect for male students' performance in Dutch language. For female students, no such effect was found. Female students' Dutch language performance is not associated with having a female Dutch language teacher. Additionally, the analyses show that female students achieve higher grades in Dutch language than male students regardless of the gender of their teachers. In all, hypothesis 1 is not supported by the data.

Despite the absence of a direct effect, we test the hypothesized indirect effects as Hayes (2018) argues that a correlation between X and Y as precondition for a mediation effect is not necessarily needed. To test the first step in the proposed mechanism explaining the association between student-teacher gender congruence and student performance,

self-perceived abilities is added to the model. The hypothesized mediation effect of self-perceived abilities is assessed using the single mediation model (model 4) in PROCESS. PROCESS is an OLS regression-based macro for SPSS which incorporates bootstrapping techniques for estimating indirect effects (Hayes, 2018).⁶ We report bootstrap estimates of indirect effects derived from 10,000 bootstrap samples along with 95% CIs, as ‘there is relatively little added value to increasing bootstrap samples above 10,000’ (Hayes 2018: 103). We run the model for math and Dutch language for males and females separately, resulting in four analyses of which the findings are presented in Table 4.3. The first and third models present the results of regressing student-teacher gender congruence on students’ self-perceived abilities. The second and fourth models present the results of regressing self-perceived abilities on students’ grades controlled for student-teacher gender congruence. The indirect effect of student-teacher gender congruence on students’ grades through self-perceived abilities are also included.

Table 4.3. The Effect of Gender Congruence on Grades through Self-Perceived Abilities

Male Students				
	Math		Dutch language	
	(1) SPA	(2) Grade	(3) SPA	(4) Grade
Constant	3.790 (0.109) p = 0.000	6.170 (0.508) p = 0.000	3.309 (0.075) p = 0.000	5.275 (0.287) p = 0.000
Gender Congruence	0.198 (0.140) p = 0.160	-0.160 (0.209) p = 0.446	-0.033 (0.164) p = 0.841	-0.387* (0.161) p = 0.017
Self-Perceived Abilities		0.158 (0.127) p = 0.216		0.347** (0.084) p = 0.000
Indirect effect		0.031 (0.038) -0.019 0.127		-0.012 (0.059) -0.134 0.105
Observations	138	138	138	138
R ²	0.014 p = 0.160	0.014 p = 0.385	0.000 p = 0.841	0.147** p = 0.000
Female Students				
	Math		Dutch language	
	(1) SPA	(2) Grade	(3) SPA	(4) Grade
Constant	3.394 (0.086) p = 0.000	5.843 (0.309) p = 0.000	3.603 (0.105) p = 0.000	5.370 (0.288) p = 0.000
Gender Congruence	-0.211 (0.141) p = 0.136	-0.100 (0.168) p = 0.554	-0.109 (0.123) p = 0.380	0.072 (0.126) p = 0.567
Self-Perceived Abilities		0.338** (0.086) p = 0.000		0.391** (0.074) p = 0.000
Indirect effect		-0.072 (0.050) -0.182 0.017		-0.043 (0.050) -0.148 0.046
Observations	191	191	191	191
R ²	0.012 p = 0.136	0.081** p = 0.000	0.004 p = 0.380	0.129** p = 0.000

Note: Standard errors in parentheses. SPA = self-perceived abilities. Size of bootstrap sample for calculations of the indirect effect = 10,000, 95% confidence interval. ** p<0.01, * p<0.05, + p<0.1

Table 4.4. Double Mediation of Gender Stereotypical Beliefs and Self-Perceived Abilities in the Relation Between Gender Congruence and Students' Grades

Male Students		Math			Dutch language	
	(1) GSB	(2) SPA	(3) Grade	(4) GSB	(5) SPA	(6) Grade
Constant	0.304 (0.056) p = 0.000	3.715 (0.120) p = 0.000	6.148 (0.506) p = 0.000	0.285 (0.039) p = 0.000	3.392 (0.088) p = 0.000	5.138 (0.299) p = 0.000
Gender Congruence	-0.011 (0.071) p = 0.882	0.201 (0.140) p = 0.153	-0.151 (0.208) p = 0.470	0.060 (0.085) p = 0.487	-0.016 (0.163) p = 0.924	-0.402* (0.160) p = 0.013
Gender Stereotypical Beliefs		0.249 (0.168) p = 0.140	0.379 (0.250) p = 0.132		-0.292+ (0.164) p = 0.077	0.251 (0.162) p = 0.125
Self-Perceived Abilities			0.133 (0.128) p = 0.298			0.367** (0.084) p = 0.000
Observations	138	138	138	138	138	138
R ²	0.000 p = 0.882	0.030 p = 0.125	0.031 p = 0.242	0.004 p = 0.487	0.023 p = 0.204	0.162** p = 0.000
Female Students		Math			Dutch language	
	(1) GSB	(2) SPA	(3) Grade	(4) GSB	(5) SPA	(6) Grade
Constant	0.360 (0.036) p = 0.000	3.425 (0.106) p = 0.000	5.855 (0.322) p = 0.000	0.320 (0.055) p = 0.000	3.575 (0.114) p = 0.000	5.361 (0.290) p = 0.000
Gender Congruence	-0.075 (0.060) p = 0.211	-0.218 (0.142) p = 0.127	-0.102 (0.169) p = 0.548	0.017 (0.065) p = 0.795	-0.110 (0.124) p = 0.374	0.072 (0.126) p = 0.571
Gender Stereotypical Beliefs		-0.085 (0.173) p = 0.626	-0.030 (0.205) p = 0.886		0.088 (0.139) p = 0.528	0.036 (0.141) p = 0.797
Self-Perceived Abilities			0.338** (0.086) p = 0.000			0.390** (0.074) p = 0.000
Observations	191	191	191	191	191	191
R ²	0.008 p = 0.211	0.013 p = 0.293	0.081** p = 0.001	0.000 p = 0.795	0.006 p = 0.558	0.130** p = 0.000

Note: GSB = Gender Stereotypical Beliefs, SPA = Self-Perceived Abilities ** p<0.01, * p<0.05, + p<0.1

For math, the findings are presented in the first two models. Model 1 reveals that student-teacher gender congruence does not have an effect on self-perceived abilities in math for either male or female students. Similar to the findings on student performance but contrary to what was expected, student-teacher gender congruence is not associated with students' self-perceived abilities in math. Model 2 presents the findings of regressing both student-teacher gender congruence and self-perceived abilities in math on math grades. For male students, no significant effects were found for either gender congruence or self-perceived abilities. Hence, both the gender of their teachers and self-perceived abilities are not associated with male students' math grades. For female students, model 2 shows a significant effect of self-perceived abilities on math grades suggesting that stronger self-perceived abilities in math are associated with higher math grades.

In accordance with the insignificant effects of student-teacher gender congruence and self-perceived abilities on math grades, the confidence intervals of the indirect effects

include zero for both males (BootLLCI -0.019, BootULCI 0.127) and females (BootLLCI -0.182, BootULCI 0.017). This confirms that self-perceived abilities do not mediate the association between gender congruence and grades in math for either male or female students.

For Dutch language, the findings are included in models 3 and 4. Model 3 shows no effect for gender congruence on either male or female students' self-perceived abilities in Dutch language. In contrast to what was expected, having a teacher of the same gender is not related to students' self-perceived abilities in Dutch language. Model 4 presents the findings of regressing both gender congruence and self-perceived abilities on Dutch language grades. For male students, significant effects of both variables are found. Contrary to the expectation, controlling for self-perceived abilities, student-teacher gender congruence predicts lower Dutch language grades for male students. Self-perceived abilities are positively associated with male students' Dutch language grades, controlling for gender congruence. For female students, model 4 only reveals a positive significant effect of self-perceived abilities indicating that, controlled for student-teacher gender congruence, stronger self-perceived abilities predict higher grades in Dutch language. An effect for gender congruence on female students' Dutch language grades was not found.

For both male and female students, the bootstrap confidence interval of the indirect effect of gender congruence on student performance through self-perceived abilities in Dutch language includes zero (male students BootLLCI -0.134, BootULCI 0.105, female students BootLLCI -0.148, BootULCI 0.046), indicating that self-perceived abilities do not mediate the association between gender congruence and performance in Dutch language for either male or female students. Therefore, hypothesis 2 should be rejected.

To examine the role of gender-stereotypical beliefs, in a last step gender-stereotypical beliefs is added to the model. The serial multiple mediator model is assessed with 'Model 6' in PROCESS (Hayes, 2018). Again, bootstrap estimates of indirect effects derived from 10,000 bootstrap samples along with 95% CIs are reported. Table 4.4 presents the findings of regressing student-teacher gender congruence on gender-stereotypical beliefs (model 1 and 4), regressing both gender congruence and stereotypical beliefs on self-perceived abilities (model 2 and 5), and regressing gender-stereotypical beliefs and self-perceived abilities on students' grades controlled for student-teacher gender congruence (model 3 and 6). The mediation effects of gender-stereotypical beliefs and self-perceived abilities, as well as the double mediation effect and the total effect of the model are presented in Table 4.5.

Table 4.4 shows a similar pattern as Table 4.3, the addition of gender-stereotypical beliefs to the model does not change any of the previously found relations. Model 1 and 4 show that, contrary to what was expected, gender congruence is not associated with students' gender-stereotypical beliefs. Having a teacher of the same gender is not related to students' beliefs about what is capable for the sexes in either math or Dutch language. Model 2 shows that gender-stereotypical beliefs are not related to students' self-perceived abilities in math. Model 5 shows a negative effect, on the 0.10 level, of gender

stereotypical beliefs on male students' self-perceived abilities in Dutch language. Stronger gender stereotypical beliefs of male students (beliefs about what is capable for the sexes in math and Dutch language) seem to be negatively associated with their beliefs about their own capabilities in the subject. For female students, no such effect was found. Lastly, model 3 and 6 reveal that gender-stereotypical beliefs are not associated with students' performance in math or Dutch language, for either male or female students.

Table 4.5 summarizes the indirect effects of student-teacher gender congruence on students' grades through gender-stereotypical beliefs, self-perceived abilities and both gender-stereotypical beliefs and self-perceived abilities. Table 4.5 reveals no indirect effect of student-teacher gender congruence on students' grades through students' gender-stereotypical beliefs for either male or female students in either math or Dutch language. Moreover, Table 4.5 shows no evidence for a serial mediation effect of student-teacher gender congruence on students' grades through gender-stereotypical beliefs and self-perceived abilities. Hence, hypothesis 3 should be rejected.

The results reported above are the findings based on the basic models. To control for effects of student age and educational level of the students, all models have also been tested with these two variables as covariates. The addition of the covariates reveals a significant negative effect, on the 0.01 level, of male students' gender stereotypical beliefs on their self-perceived abilities in Dutch language (basic model: $p = 0.077$, model with covariates: $p = 0.009$). All other effects and patterns found are similar to the findings of the basic model. No other conclusions are to be drawn regarding the hypotheses.

Table 4.5. Indirect Effects of Gender Congruence on Students' Grades

Male students						
	Math			Dutch language		
	Effect Size (SE)	LLCI	ULCI	Effect Size (SE)	LLCI	ULCI
Total Mediated Effect	0.022 (0.051)	-0.071	0.136	0.003 (0.065)	-0.127	0.132
Mediation through GSB	-0.004 (0.033)	-0.079	0.062	0.015 (0.025)	-0.031	0.072
Mediation through SPA	0.027 (0.036)	-0.020	0.120	-0.006 (0.062)	-0.132	0.116
Double mediation	-0.000 (0.004)	-0.010	0.008	-0.006 (0.011)	-0.031	0.016
Female students						
	Math			Dutch language		
	Effect Size (SE)	LLCI	ULCI	Effect Size (SE)	LLCI	ULCI
Total Mediated Effect	-0.069 (0.054)	-0.184	0.030	-0.042 (0.051)	-0.151	0.049
Mediation through GSB	0.002 (0.021)	-0.039	0.053	0.001 (0.009)	-0.020	0.020
Mediation through SPA	-0.074 (0.050)	-0.185	0.014	-0.043 (0.050)	-0.150	0.048
Double mediation	0.002 (0.006)	-0.007	0.018	0.001 (0.004)	-0.008	0.009

Note: Direct effect refers to $X \rightarrow Y$ controlled for M1 and M2, Total Mediated Effect is the sum of the indirect effects of $X \rightarrow Y$. Size of bootstrap sample for calculations of the indirect effect = 10,000. Standard deviations of the 10,000 bootstrap estimates of the indirect effect between parentheses (SE). LLCI = bootstrapped lower level confidence interval; ULCI = bootstrapped upper level confidence interval, 95% confidence interval.

4.5 DISCUSSION AND CONCLUSION

Despite the fact that empirical research on representative bureaucracy and performance has expanded extensively in recent years, to date individual-level mechanisms explaining the effects of representation on performance in public organizations have hardly been examined. In particular, individual-level mechanisms explaining representation effects from the perspective of the citizen-client are subject to speculation. This study contributes to representative bureaucracy theory by examining a psychological process that may underlie symbolic representation effects. We empirically tested our model in the school context and hypothesized an effect of student-teacher gender congruence on students' performance via students' gender-stereotypical beliefs and self-perceived abilities in the subjects math and Dutch language. The model was tested on multi-source data on high school students in the Netherlands collected in the first wave of a longitudinal study among families in the Netherlands.

We separated our analyses for math and Dutch language and for female and male students. By doing so, our proposed model was tested four times; for male and female students in math and male and female students in Dutch language. Our study provides no evidence for the hypothesized positive representation effects and the theorized mechanism underlying symbolic representation in any of the four cases.

For math, we did not find that gender representation was associated with either male or female students' math performance, nor did we find evidence for a role of students' self-perceived abilities or gender-stereotypical beliefs in the hypothesized associations. We did find that self-perceived abilities can predict math scores for female students, but not for male students, suggesting different dynamics of self-perceived abilities in math for male and female students.

Also for female students in Dutch language, our results showed no evidence for a relation between female Dutch language teachers and female students' Dutch language performance, self-perceived abilities, or gender-stereotypical beliefs. The results did show that stronger self-perceived abilities of female students in Dutch language are related to higher grades in the subject.

For male students in Dutch language, our study showed surprising findings. In contrast to the positive gender-representation effects we expected, we found a *negative* gender-representation effect for male students in Dutch language. The results showed that male students with a male Dutch language teacher had lower grades than male students with a female Dutch language teacher. The proposed psychological processes described in our model cannot explain this negative effect: neither male students' ideas about males' and females' capabilities in Dutch language, nor male students' judgements about their own capabilities in the subject were associated with having a male Dutch language teacher. The results did show that stronger gender stereotypical beliefs can predict lower self-perceived abilities and that stronger self-perceived abilities of male students in Dutch language can

predict higher grades in the subject.

The null-findings and negative gender-representation effect in Dutch language for male students contradict the evidence for effects of gender representation on student performance from previous studies (Agyapong, 2018; Dhillon & Meier, 2020; Keiser et al., 2002; Song, 2018; Zhang, 2019), and positive representation effects in general (Riccucci & Van Ryzin, 2017). While our study is not the first to find negative representation effects (Doornkamp et al., 2021; Keiser et al., 2002), previously found negative effects have been assumed to be ‘spillover effects’, suggesting that, to a certain degree, the majority may benefit from minority representation effects as well (Riccucci et al., 2018). However, as the positive representation effects for female students (the minority gender in our study) are absent in our study, the negative gender-representation effects for male students seem to stand on their own. This, and the null-findings with respect to most hypotheses, lead us to explore other explanations.

Let us first more closely consider the national context in which this study has been conducted. In representative bureaucracy theory, it is argued that for gender-representation effects to occur gender has to be salient to the policy at hand (it has to be a men’s issue or a women’s issue) (Keiser et al., 2002). The psychological processes theorized in our model also depend on this notion (the role of gender-stereotypical beliefs and self-perceived abilities). Gender is assumed to be salient to education as a consequence of enduring disadvantages of women compared to men, as becomes apparent from gender differences in achieved educational level (Evans et al., 2020). Historically, women have been discriminated against based on gender stereotypes in which they are assumed to have lower academic capabilities (e.g., Hausmann, 2014; Prentice & Carranza, 2002; Szymanowicz & Furnham, 2011). Even now, educational access for girls and women is not self-evident everywhere in the world. Probably, it is no coincidence that in recent studies effects of gender representation in education have been found in countries in which gender inequality in educational attainment still exists: Ghana, India, South Korea, and China rank in the third quartile of the Global Gender Gap Index Rating; subindex Gender Gap in Educational Attainment (Agyapong, 2018; Dhillon & Meier, 2020; Song, 2018; Zhang, 2019; see Global Gender Gap Report, 2020). However, educational attainment in the Netherlands, the country in which this study has been conducted, is equal for men and women (see Global Gender Gap Report, 2020). Although the educational careers of Dutch male and female students display gender-stereotypical patterns (see Onderwijsraad, 2019), the role of gender (and gender stereotypes) might be less salient and explicit to Dutch boys and girls in high schools than assumed, causing the drivers of gender-representation effects to diminish (or even disappear) (cf. Fay et al., 2020; Meier, 2019).

More importantly, the sample in our study consists of adolescents mainly from Dutch families with a higher SES compared to the nation’s average. Stereotypical gender roles and behaviors are more pronounced in segments of the population with lower SES compared to segments with a higher SES (e.g., Trusty et al., 2000). As a result, the role of gender might

be less salient and explicit for the students in our sample not only because of relatively high gender equality in the Dutch national context, but also because of gender roles being less pronounced in high-SES families, decreasing the likelihood of gender-representation effects.

So far, in representative bureaucracy theory salience of minority status to the policy at hand has been discussed in the context of active representation; minority bureaucrats need to experience salience of their minority status to the policy at hand for their willingness to actively represent. Our study points to the role of salience in the context of symbolic representation; for symbolic representation effects to occur the minority citizen-client has to experience their minority status to be salient in the context of the bureaucratic interaction. This might inspire scholars to apply social cognitive theory to the study of symbolic representation in a variety of bureaucratic interactions in order to explore the role of modeling effects in contexts that may vary as to the saliency of social norms related to a minority status.

Moreover, the national and socioeconomic context in which interaction between the bureaucrat and citizen-client takes place, seem to play an important part in the degree to which salience is experienced and representation effects are likely to occur. Future research on bureaucratic representation might want to study the role of national and socioeconomic context in experienced salience of minority status to the policy at hand as a condition for symbolic representation to occur.

The negative gender-representation effect for male students in Dutch language found in our study could not be explained by the proposed psychological mechanism including gender-stereotypical beliefs and self-perceived abilities. The negative representation effect might have been caused by other mechanisms from the perspective of the student (symbolic representation), or by mechanisms from the perspective of the teacher (active representation). From the perspective of the student, teachers similar to themselves might strengthen pressures to succeed. Stress and pressure to succeed can create a psychological state in which negative thought is promoted and optimal academic performance is impeded (Autin & Croizet, 2012).

From the perspective of the teacher, teachers' grading might be influenced by the gender of a student. Research on gender grading bias shows that male students are systematically graded more strictly by teachers than female students, resulting in a negative grading bias against boys (Bonnesronning, 2008; Cornwell et al., 2013; Falch & Naper, 2013; Lindhal, 2007; Lavy, 2008; Matějů & Smith, 2014; Protivínský & Münich, 2018). For Dutch language, this grading bias in combination with the stereotype that male students would perform worse in Dutch language than female students could (unconsciously) influence teacher grading practices. In particular, male teachers might be stricter in grading male students in Dutch language as family studies indicate that men hold stronger explicit stereotypical attitudes about gender than women (Blakemore & Hill, 2008; Endendijk et al., 2013; Tenenbaum & Leaper, 2002). Also, social role theory proposes that fathers are more likely to socialize

children, especially boys, in accordance with prevailing gender stereotypes in society (Eagly et al., 2000), which may also apply to male teachers. And although the teacher may show contra-stereotypical behavior (male Dutch language teacher), this teacher may still believe the gender stereotype connected to their subject is true (Xu & Meier, 2021). To receive a better understanding of (negative) representation effects, future research might want to persist to explore the (unconscious) psychological processes at the individual level underlying both symbolic and active representation.

The findings of this study should be interpreted in light of its limitations. In addition to the study's national context and the relatively high SES of the families included in the sample, the composition of the sample as to the gender of teachers across subjects led to statistical limitations. The skewed distribution of teachers in Dutch language and math resulted in small subsamples of male students with male Dutch language teachers and female students with female math teachers, whereas, theoretically, they were key to our study. Although bootstrapping techniques account for irregularities of sampling distributions and increase power for the calculation of indirect effects, this may have weakened the statistical power and, hence, the likelihood of Type 2 errors. This may have driven the null findings, too.

Furthermore, the educational context in which our study is conducted makes a strong case for active representation because of teachers' discretion and salience of minority status in education (Keiser et al., 2002). Therefore, in such a context, both symbolic and active representation may occur simultaneously and should be seen as complementary processes. Hence, it is difficult, if not impossible, to disentangle effects of symbolic and active representation in the empirical analyses. That said, social cognitive theory provided us with strong theoretical argumentation for the mechanism through which symbolic representation occurs. More importantly, as student-teacher gender congruence is an exogenous variable and teachers and students do not have any chance to influence the student-teacher-match, no systematic biases were expected.

Despite these limitations, our study provides some important implications for the study of symbolic effects of gender representation in education specifically. First, when studying psychological mechanisms explaining symbolic effects of gender representation, scholars should consider differences in mechanisms for male and female students across subjects. Our results show divergent relations between the same variables for male and female students in math and Dutch language.

Second, contrary to our expectations, our results show no evidence for effects of gender-stereotypical beliefs. We used a measure of implicit gender stereotypical beliefs, while previous studies have relied on survey measures of gender stereotypical beliefs (Doornkamp et al., 2019; Zhang, 2019). It may be relevant to explore the role implicit and espoused gender stereotypes may play in symbolic effects of gender representation. This could contribute to further theorizing about the importance of unconscious bias compared to opinionated beliefs in such psychological processes.

Finally, this study's null effects have drawn attention to the characteristics of our

sample, emphasizing the need for comparative research on representation effects across countries. Additionally, scholars in representative bureaucracy should consider and further theorize on the role of SES and its intersection with gender (and race), both in education and in other public services. Moreover, this study calls for further exploration of (unconscious) psychological mechanisms grounded in the human interactions producing representation effects. The reverse effects and suggested conditions for symbolic effects of gender representation to occur point at avenues for further exploration of psychological mechanisms and their conditions in the effects of symbolic representation.

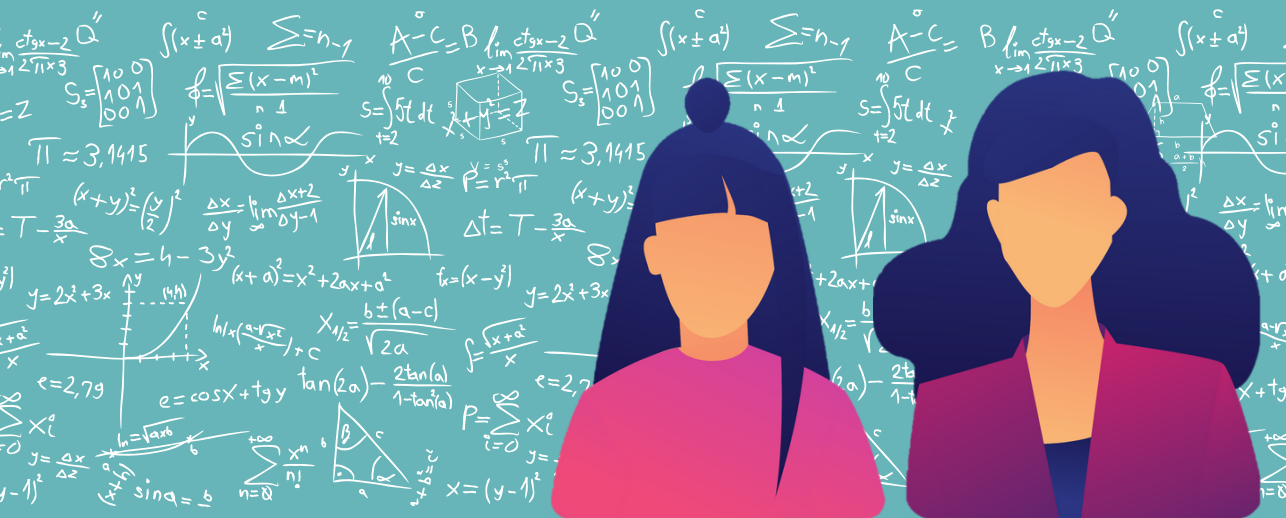
4.6 NOTES

1. The longitudinal study has been discussed with a data protection officer and has been approved by Leiden University ethics committee.
2. In the Netherlands, secondary education exists of three educational levels: VMBO (lowest level; prepares for vocational education, four years; age 12 till 16), HAVO (intermediate level; prepares for applied sciences, five years; age 12 till 17) and VWO (highest level; prepares for university, six years; age 12 till 18).
3. As representative bureaucracy theory uses the term '*gender* congruence', for consistency purposes we also use the term 'gender' to indicate 'sex'. The questionnaire included a non-binary option for respondents to indicate their sex. This option was not used. Consequently, we treated the gender congruence concept as a binary concept although we are aware gender is not binary.
4. To assess whether the missing data is random, ANOVA analyses were conducted comparing students with and without missing grades for the main variables in our model. The analyses show that the groups with and without missing grades do not differ significantly.
5. The IAT measures the association of female and male attributes with the school related concepts 'arts' and 'science'. During the IAT, participants are requested to sort words into groups by pressing keys. In congruent blocks, female names (e.g., 'Julia') and feminine words (e.g., 'Dutch language') need to be sorted in the 'language and culture' group. In incongruent blocks female names and masculine words (e.g., 'Math') need to be sorted in the 'nature and technology' group. Reaction time and accuracy in sorting these names and words are administrated and used to calculate participants' implicit gender stereotypes associated in the school context.
6. Bootstrapping is a resampling process in which an empirically derived representation of the sampling distribution is generated. The indirect effect for each taken random sample from the original sample is calculated constructing the confidence intervals. As a result, irregularity of sampling distribution is better respected and inferences are more likely to be accurate. Especially in small samples, the bootstrap technique is

useful as the bootstrapping increases statistical power (Hayes, 2018).

4.7 ACKNOWLEDGEMENTS

The authors would like to thank Kenneth J. Meier and Dimiter Toshkov for their constructive and insightful feedback on previous versions of this article.



CHAPTER **FIVE**

Understanding gender bias in teachers' grading: The role of gender stereotypical beliefs

Previously published as: Doornkamp, L., Van der Pol, L.D., Groeneveld, S., Mesman, J., Endendijk, J.J., & Groeneveld, M.G. (2022). Understanding Gender Bias in Teachers' Grading: The Role of Gender Stereotypical Beliefs. *Teaching and Teacher Education*, 118, 103826.

ABSTRACT

This study experimentally tested the influence of secondary school students' gender on Dutch language and math teachers' grading (N = 358) and examined the role of teachers' gender and gender stereotypes in gender grading bias. Results showed that grading, on average, was not gender biased. However, differences between teachers' gender grading bias were related to their gender stereotypes. Although we found no direct effect of teachers' gender on grading bias, for math we found an indirect effect through their gender stereotypes. This study provides evidence for the role of teachers' gender stereotypes in gender grading bias that thus far had only been assumed.

5.1 INTRODUCTION

Gender bias in teachers' grading refers to the differences in grades between male and female students with the same level of academic skills (Protivínský & Münich, 2018). As grades are interpreted as objective measures of students' level of academic skills, gender biases in teachers' grading can have far-reaching consequences. For instance, gender bias in teachers' grading was found to be associated with students' enrollment in advanced level math courses which has long-term implications for students' future careers (Lavy & Sand, 2018).

Despite the multitude of studies on the topic of gender grading bias, still little is known about the causal effect of students' gender on teachers' grading as experimental designs to study the effect are rare (Hanna & Linden, 2012; Hinnerich et al., 2011). Further, although scholars urged that the role of teachers' characteristics in gender grading bias should be explored (Falch & Naper, 2013; Lavy, 2008; Matějů & Smith, 2015; Protivínský & Münich, 2018), thus far, only a few studies empirically tested the association between teachers' gender and gender grading bias (Lavy, 2008; Lindahl, 2016). Moreover, the role of teachers' implicit gender stereotypes has often been assumed, but not empirically tested, to underly gender grading bias (Breda & Ly, 2015; Burgess & Greaves, 2013; Gibbons & Chevalier, 2008; Lavy, 2008; Lavy & Sand, 2018; Matějů & Smith, 2015).

In the current study we use a field experiment among Dutch teachers in training in the subjects Dutch language and math to contribute to the understanding of gender grading bias. We aim to answer the research question: *to what extent is teachers' grading gender biased and what is the role of teachers' gender and teachers' implicit gender stereotypes in gender grading bias?* By manipulating gendered names on tests, we examine the extent to which teachers' grading is influenced by students' perceived gender. We further investigate whether gender grading bias is associated with teachers' gender and implicit gender stereotypes and we propose and empirically test a model that explains the potential influence of the gender of the teacher on gender grading bias through teachers' implicit gender-typed associations and gender-typed expectations. In doing so, we are able to make inferences about causality, about the role of teacher characteristics and about a mechanism underlying gender grading bias.

5.2 GENDER GRADING BIAS

The literature on gender grading bias consists of several studies without a clear-cut theory that binds them together. Studies on the phenomenon of gender grading bias date back to at least 1967 (Caldwell & Hartnett, 1967). Since then, the topic has been studied in the fields of economics, education, and sociology, each using different terminologies to refer to the same phenomenon. Gender grading bias (Berg et al., 2020), gender grading gap (Falch & Naper, 2013), gender disparities in grading (Cornwell et al., 2013), gender bias

in examination (Stobart et al., 1992), evaluation bias (Breda & Ly, 2015), assessment bias (Lindahl, 2016), and teacher discrimination (Lavy, 2008) all refer to teachers' systematic over- or underassessment of a group of students based on students' gender. The literature on gender grading bias treats gender as a binary concept as it describes teachers' differential grading of male versus female students (e.g., Protivínský & Münich, 2018). We are aware that gender can entail more than binary categories male and female. However, societies generally are still organized in this binary way, categorizing humans as either male or female. This categorization lies at the basis of gender stereotypes and the accompanying gender discriminative behaviors in everyday life (Hyde et al., 2019). As we study the role of gender stereotypes in teachers' actual grading behavior, we follow the gender binary approach.

The vast majority of studies on the topic used secondary data and relied on the comparison of students' exam scores obtained from non-blind grading (generally the teacher is the grader and is aware of student gender) to students' exam scores obtained from blind grading (the assessor is not aware of student gender, these scores are interpreted as objective measures of students' academic level of skills). If the gender differences in test scores awarded by teachers (non-blind grading) were larger than the gender differences in blindly graded standardized entrance exams (Protivínský & Münich, 2018), central exit exams (Falch & Naper, 2013), externally graded state exams (Lavy, 2008), or national tests (Lindahl, 2016), these studies concluded that there is gender grading bias. Generally, blind grading produces lower test scores for all students (male and female) than non-blind (teacher) grading (Hinnerich et al., 2011). More importantly, whereas blind grading shows no or little differences in average male and female test scores, teachers do seem to differentiate between male and female students in their grading, i.e., gender grading bias (e.g., Protivínský & Münich, 2018).

The differences in male and female scores as a result of teachers' grading is generally to the disadvantage of male students, meaning that male students systematically receive lower scores than female students in teacher grading (Protivínský & Münich, 2018). Tens of previous studies found evidence for this grading bias against male students across several subjects, in different countries, and for different educational levels (for an overview see Protivínský & Münich, 2018). Whereas only a handful of studies found biased grading against female students in different subjects like math, but also in humanities and English (Breda & Ly, 2015; Gibbons & Chevalier, 2008; Lavy & Sand, 2018), or did not find gender grading bias at all (Hanna & Linden, 2012; Hinnerich et al., 2011). Notably, studies that did not find gender bias in grading were the only few studies that have experimentally tested the influence of students' gender on teachers' grading (Hanna & Linden, 2012; Hinnerich et al., 2011). Hence, little is known about whether there is a causal relation between students' gender and teachers' grading. Insight in cause-effect provides the basis for methods to reduce grading bias. In all, most of the evidence on the topic of gender grading bias points in the direction of a grading bias in teachers' grading against male students, we therefore hypothesize:

H1: Perceived male gender of students is negatively associated with teachers' grading.

5.2.1 THE ROLE OF TEACHERS' GENDER AND GENDER STEREOTYPES IN GENDER GRADING BIAS

The role of teacher characteristics in gender grading bias has often been suggested in the literature (e.g., Matějů & Smith, 2015), but only a handful of studies have empirically tested the associations between teacher characteristics and gender grading bias (Lavy, 2008; Lavy & Sand, 2018; Lindahl, 2016). Teachers' gender, age, experience, number of sons and daughters, and marital status were found to be related to gender grading bias, indicating that gender grading bias is indeed sensitive to teacher characteristics (Lavy, 2008; Lavy & Sand, 2018).

Teachers' own gender (whether they classify themselves as male or female) can play an important role in gender grading bias. In schools, some subjects are stereotyped as typically masculine (e.g., math) or typically feminine (e.g., languages). Although the magnitude of the 'math = masculine' gender stereotype can differ across cultures, the content of the stereotype generally is the same (Breda et al., 2020; Nosek et al., 2009). A female teacher in a stereotypically masculine subject (e.g., a female math teacher) may have accumulated counter-gender stereotypic experiences that affect her gender bias in grading (Lavy & Sand, 2018). Because of her personal experience that a woman can perform well in a subject that is stereotyped as masculine, a female math teacher can have weaker stereotypic ideas and beliefs regarding male and female students in math (Bussey & Bandura, 1999). These experiences might even result in counter-stereotypic beliefs that might turn into counter-stereotypic behaviors (Crisp et al., 2009; Endendijk et al., 2013). In general, though, experiences affect the strength of gender stereotypes but not the direction of the stereotype (i.e., gender stereotypes are flexible but remain gender stereotypic) (Charlesworth & Banaji, 2021; Xu & Meier, 2021). Therefore, counter-stereotypic experiences are more likely to result in weaker gender stereotypical beliefs and behaviors (for a meta-analysis see Lenton et al., 2009). Following this logic, female teachers would have weaker gender grading bias in math than male teachers, and male teachers would have weaker gender grading bias in languages than female teachers (Lavy, 2008).

H2a: Female math teachers display less gender grading bias than male math teachers.

H2b: Male Dutch language teachers display less gender grading bias than female Dutch language teachers.

Indeed, teachers' gender stereotypes are presumed to underly gender grading bias (Breda & Ly, 2015; Hanna & Linden, 2012; Hinnerich et al., 2011; Lavy, 2008; Lavy & Sand, 2018;

Lindahl, 2016; Matějů & Smith, 2015). Gender stereotypes are socially constructed ideas that provide descriptions of what men and women are like, and prescriptions about what men and women should be like (Ellemers, 2018). The content of gender stereotypes is assumed to be similar across different cultures, however it should be noted that recent research demonstrated that culture can shape how men and women are perceived based on gender stereotypes (Obioma et al., 2021). Obioma et al. (2021) showed that the content of the gender stereotype ‘men are agentic, women are communal’ is similar in Germany and Nigeria. However, in their study, in Germany men were perceived as more agentic than women while men and women were perceived as equally communal. In Nigeria, men and women were perceived as equally agentic while women were perceived as more communal than men (Obioma et al., 2021). Notwithstanding differences in perception based on gender stereotypes, individuals who have strong gender stereotypes make a strong distinction between men and women and between typically masculine and typically feminine attributes and behaviors (i.e., gender-typed individuals/gender-schematic individuals) (Bem, 1981). When confronted with someone else’s behavior, these individuals are likely to perceive the behavior in terms of congruence with their ideas about and expectations of males and females (Bem, 1981). Behavior in line with these ideas and expectations are rewarded, whereas incongruent behavior is penalized. Consequently, the distinction between males and females based on gender stereotypes can result in different evaluations of the same behavior of males and females (e.g., gender grading bias) (e.g., Burgess & Borgida, 1999). Following this logic, teachers who have strong gender stereotypes are likely to make a strong distinction about what is typical and appropriate for male and female students. As a result, the same behavior in males and females is evaluated differently which makes gender bias in grading more likely.

An individual can have implicit and explicit gender stereotypes. Implicit gender stereotypes refer to the *automatic and less controllable* distinctions individuals make, whereas explicit gender stereotypes refer to individuals *purposeful and controllable* distinguishment about what is typical of and appropriate for males and females (Fazio & Olson, 2003). Implicit and explicit gender stereotypes are only weakly correlated, in part because of explicit measures’ sensitivity to social desirability or lack of insight in one’s own stereotypes (Fazio & Olson, 2003). Teachers can report that they do not distinguish between male and female students, whereas observations or implicit measures can prove otherwise (Jones & Myhill, 2004; Nürnberger et al., 2016). In a country such as the Netherlands, in which gender equality is considered important, implicit measures of gender stereotypes could provide a better understanding of the (implicit) role of gender stereotypes in teachers’ grading behaviors.

In this study, we distinguish between implicit gender-typed associations and implicit gender-typed expectations. Both are unconscious cognitive processes through which one differentiated between male and female students. Implicit gender-typed associations refer to teachers’ automatic and unconscious associations between math and masculinity, and between languages and femininity. Generally, people tend to associate math with

masculinity and languages with femininity based on the traditional belief that math is the domain in which men can excel and that languages is the domain in which women can excel (Nosek et al., 2002b). As a result, teachers can have stereotypical perceptions of male and female students' capabilities in math which can turn in to different assessments of male and female students (Riegle-Crumb & Humphries, 2012). Math teachers who have strong associations between math and masculinity and between language and femininity make a strong distinction between males' and females' capabilities in the subjects, which might result in a stronger gender grading bias. On the other hand, teachers who have weak associations between gender and subjects might be more equal in their grading of male and female students.

In a related vein, implicit gender-typed expectations refer to teachers' automatic and unconscious expectations regarding male and female students' talent and effort based on gender stereotypes. Teachers can expect female students to have less talent in math than male students (e.g., Espinoza et al., 2014; Heller et al., 2001; Riegle-Crumb & Humphries, 2012), and male students to have less talent in languages than female students (e.g., Schmenk, 2004; Siegle & Reis, 1998). On the other hand, based on negative stereotypes regarding male students' interest, motivation, and achievements in schools (Brown & Stone, 2016), teachers can expect less effort from male students than from female students (Heyder & Kessels, 2017; Jones & Myhill, 2004; Siegle & Reis, 1998). Again, teachers who differ in their expectations of male and female students might display gender bias in their grading.

Indeed, literature on gender bias in job applications shows that negative expectations can lead to stricter assessments and positive expectations can lead to more lenient assessments (Powell, 1986). These positive or negative expectations can be informed by gender stereotypes. As a result, female applicants have less chance of being assessed positively by recruiters for jobs that are perceived as typically masculine (e.g., management positions), and male applicants have less chance of being assessed positively by recruiters for jobs that are perceived as typically feminine (e.g., communication positions) (Cole et al., 2004). Following this logic, teachers who have negative expectations regarding male or female students' talent or effort in a subject, might assess these students more strictly, and vice versa. At the same time, teachers with more equal expectations regarding male and female students' talent and effort, might assess male and female students more equally.

In sum, we expect that the stronger a teacher perceives male and female students as different in terms of their suitability, talent and effort in math or languages, the stronger the teacher shows differential grading of male and female students (Dasgupta, 2004; Gansen, 2019). Vice versa, we expect teachers with weaker gender-typed associations and gender-typed expectations regarding students' talent and effort to have weaker gender grading bias.

H3: Weaker implicit gender-typed associations and weaker gender-typed expectations regarding students' talent and effort are related to weaker gender grading bias.

5.2.2 TEACHERS' GENDER, GENDER STEREOTYPES, AND GENDER GRADING BIAS

Based on the literature discussed thus far it is likely that implicit gender-typed associations and gender-typed expectations at least partly mediate the relation between gender of the teacher and gender grading bias. That is, the implicit associations between gender and math and languages may be weaker among female math teachers and male language teachers compared to male math teachers and female language teachers (Martin & Dinella, 2012; Smeding, 2012). Consequently, these weaker implicit gender-typed associations help shape biased expectations regarding male and female students' talent and effort in the subjects (Denessen et al., 2022; Muntoni & Retelsdorf, 2018; Van den Bergh et al., 2010). Teachers' expectations are shaped by a number of co-occurring factors including teachers' implicit associations (Denessen et al., 2022). Teachers' expectations may in turn be related to gender grading bias. The proposed mediation model is depicted in Figure 5.1.

H4: The association between teachers' gender and gender grading bias is mediated by teachers' implicit gender-typed associations and implicit gender-typed expectations.

Female math teachers may have weaker gender-typed associations, that are in turn related to weaker gender-typed expectations regarding students' talent and effort, and weaker gender grading bias than male math teachers. Male Dutch language teachers may have weaker gender-typed associations, that are in turn related to weaker gender-typed expectations regarding students' talent and effort, and weaker gender grading bias than female Dutch language teachers.

Our hypotheses concentrated on teachers' gender and implicit gender stereotypes in relation to gender grading bias without hypothesizing the direction of effects (i.e., effects on grading bias against male or female students). Because the examination of psychological mechanisms that could underly teachers' gender bias in grading is still largely uncharted territory, we decided to explore the full spectrum of teachers' gender stereotypes and gender grading bias, without excluding participants with contra-stereotypical associations, expectations, or grading biases a priori. We thus focused on the extent to which the distinction between male and female students based on gender stereotypes can result in differential grading of male and female students. However, our data allows us to explore the direction of the (mediating) effects of implicit gender stereotypes both on grading biases against male and female students in Dutch language and math. These additional explorations will provide a complete picture of the role of implicit gender stereotypes in gender grading bias.

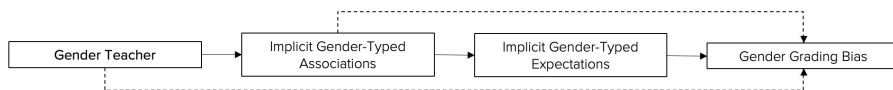


Figure 5.1. The Indirect Effect of Teachers' Gender on Gender Grading Bias

5.3 METHOD

This study's field experiment is part of the longitudinal research project 'Girls in Science' that examines gender socialization in the family and school context in the Netherlands from 2017 until 2022. The experimental data for this study were collected between February 2020 and August 2021 among teachers in training in the subjects Dutch language and math in secondary education. For the sake of legibility, in the following sections we use the term 'teacher/teachers' to refer to the participating teachers in training in our study.

5.3.1 SAMPLE

The teachers who participated in this study were recruited through their trainers. Trainers from all the 18 institutions that provide regular teacher training programs for secondary education in Dutch language and math in the Netherlands were invited to cooperate by e-mail (trainers from online institutions were not invited). Thirty-two trainers from 15 institutions agreed to cooperate, the others did not participate because of time issues or they did not respond to the invitation. The trainers either asked their students (the teachers) to participate in the research or facilitated the researcher to ask the students (teachers) to participate.

Based on the calls, 159 Dutch language teachers and 268 math teachers signed up to participate in the research. In all, 28 cases for Dutch language and 41 cases for math were regarded as missing data because their evaluations got lost in the Dutch mail system (2 Dutch language, 5 math), they did not send back their evaluations (24 Dutch language teachers, 26 math), they did not complete the evaluation (1 Dutch language, 5 math), or they did not complete the questionnaire (1 Dutch language, 5 math). Only the participants with scores on all relevant variables were included in the study, resulting in $N = 131$ for Dutch language (mean age = 33.73, $SD = 12.098$, 76% female), and $N = 227$ for math (mean age = 29.44, $SD = 12.027$, 56% female). Participants that were excluded from the analyses did not deviate on background characteristics from the participants that we included.

The distribution of male and female teachers across math and Dutch language is skewed, but similar to the gender distribution in Dutch secondary schools. Dutch language teachers in the Netherlands are disproportionately female (72%). For math, the distribution is slightly more balanced with 58% male and 42% female teachers (numbers based on open educational data 2018: DUO 2020).

Both the Dutch language and math teachers performed internships at secondary schools for 10 hours (Dutch language $SD = 8.53$, math $SD = 7.49$) a week. For Dutch language, 27% of

the participants received their training on a university (first degree) whereas 73% received their training on a university for applied sciences (second degree). Furthermore, 41% of the Dutch language teachers had experience with teaching. For math, 15% received their training at university, whereas the other 85% received it at a university for applied sciences. In math, 37% of the teachers in training had experience with teaching.

5.3.2 STUDY DESIGN

The participating teachers each (1) graded a set of 24 tests made by last-year students in higher secondary education, (2) assessed each test on invested talent and effort, and (3) filled in an online questionnaire. Data collection took place at home and participation required approximately 2.5 hours. In return for their participation, teachers received individual feedback on their grading practices, the results of the study, and a small gift.

The student names above the 24 tests were manipulated. Originally, 10 female students and 14 male students completed the Dutch language tests and 14 female students and 10 male students completed the math tests. The 24 tests were divided into two groups with the same average scores; one group was given names that are generally considered as male, whereas the other group was given names that are generally considered as female. We tried to maximize the ecological validity of our experiment within the limitations of its practical feasibility. To create conditions that are transferable to the real-life exam evaluation situation, we designed a grading experiment using real school exams with real students' answers. The study design is depicted in Figure 5.2.

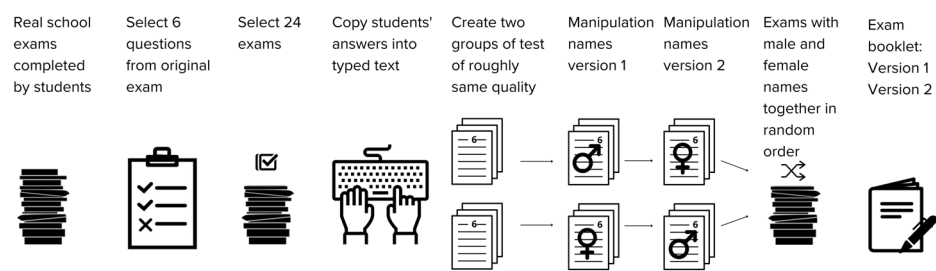


Figure 5.2. Study Design

The Dutch language and math school exams and students' answers to the exams that were used in the experiment were retrieved from a Dutch language and a math teacher from different Dutch secondary schools in the western region of the Netherlands. Both exams were conducted in 2019 by students in the final year of their secondary education.

We selected six open-answer questions from both school exams (Dutch language and math) to avoid asking too much time from the participants. We used the following two

criteria in the selection of these questions: (1) there had to be ample room for participants' own interpretation in their assessment of the given answers to the question, and (2) there was sufficient between-student range in the points given for answers to a question by the original assessor to ensure that the quality of the answers to the question differed.

From the groups of students who originally took the Dutch language exam ($N = 39$) and the math exam ($N = 45$), 24 students for each subject were selected to mimic real-life situations in which teachers must evaluate the tests of one class with approximately 25 students. The selected students' answers (including mistakes, misspellings, and notes) were copied into typed text and printed to eliminate any reference to the gender of the original student through for example handwriting.

To ensure that we could form two groups of tests (one group with names that are generally considered as male and one group with names that are generally considered as female) of equal quality in terms of answers given to the selected questions, we recruited 10 experienced teachers for each subject to blindly grade the 24 tests (i.e., without student names). Based on these blind evaluations, two groups were created with 11 tests each.

We used common Dutch female names (8), Dutch male names (8), Turkish female names (4), and Turkish male names (4) that were randomly distributed over the groups of tests (for an overview of the names see Appendix A). To reflect the Dutch multicultural society, we used Turkish names in addition to Dutch names as people with a Turkish migration background are, after the native Dutch, the largest ethnic-cultural group in Dutch society (CBS, 2020). To control for an unforeseen effect of differences in students' answers we created two versions. The tests with male names in the first version had female names in the second version, and the tests with female names in the first version had male names in the second version. The versions were randomly distributed across the participants. Furthermore, to control for grading order and grading fatigue, the order of tests was randomly distributed and different for the two versions. Additionally, the first test in the set alternated within each version.

In addition to the evaluation of the tests, participants were asked to estimate students' talent and effort based on each test. The 24 tests, including these questions, were presented in a booklet. A short introduction with general information about the experiment was added, but nothing was said about the gender grading bias component. To prevent participants from being primed, participants were told that the study was on the topic of grading accuracy in general. After completing their participation in the study, teachers in training were fully informed about the goal of the study.

The booklet also included an answer model provided by the original assessors, with a brief description of the main elements of a correct answer for each question. The maximum number of points per question for a good answer (which was similar to the original assessment) was given to the participants, but not further specified. The maximum total number of points for the Dutch language exam was 10 whereas for math it was 26.

During the evaluations of the tests, participants were asked to write the name of the

student on a provided form. This was done to ensure that participants read the name of the student and thus became aware of the student's gender (the manipulation). Ethical approval for this study was provided by the Ethics Committee of the Faculty of Governance and Global Affairs (University Leiden).

5.3.3 MEASURES

Gender grading bias

Gender grading bias refers to bias in evaluations against either male or female students. The mean scores for the tests with a male name (11) and the tests with a female name (11) were calculated based on the total number of points rewarded for the answers to the six questions. Then, gender grading bias was calculated by subtracting the mean male test score from the mean female test score. As a result, gender grading bias is a continuous variable including negative and positive scores. The negative scores indicate a gender grading bias against females, whereas positive scores indicate a gender grading bias against males.

Implicit gender-typed expectations

The gender stereotypical expectations of the teachers were measured for two aspects: talent for the subject and invested effort. After the teacher evaluated a single test, the teacher was asked to estimate the (fictitious) student's talent and effort on a scale from 1 to 10 (i.e., on a scale from 1 to 10 to what extent do you think this student has talent for the subject/invested time in preparing the test). We did not put any explicit emphasis on gender of the student.

Similar to the gender grading bias measure, mean scores were calculated for the tests with male names (11) and the tests with female names (11). Then, gender bias in expectations was calculated by subtracting the mean score for the male tests from the mean score on the female tests. Again, negative scores indicate a bias against females (males have more talent/put more effort than females), whereas positive scores indicate a bias against males (females have more talent/put more effort than males).

Implicit gender-typed associations

The implicit gender-typed associations of the teachers were assessed with the arts-science Implicit Association Test (IAT). The IAT was built in LimeSurvey, an online questionnaire tool, and was based on the task on the Harvard project Implicit demonstration website and the Nosek et al. (2002a) paper. The IAT measures participants' implicit association between female and male attributes and school related concepts 'languages' and 'science'. During the IAT, participants were requested to sort words into groups by pressing keys. In congruent blocks, female names (e.g., 'Julia') and stereotypically feminine words (e.g., 'Dutch language') needed to be sorted under the 'Female & Language' category and male names (e.g., 'Daan') and stereotypically masculine words (e.g., 'math') needed to be sorted under

the 'Male & Science' category. In incongruent blocks female names and stereotypically masculine words needed to be sorted under the 'Female & Science' category and male names and stereotypically feminine words needed to be sorted under the 'Male & Languages' category. The reaction time and accuracy scores were recorded by LimeSurvey. Each participant's level of implicit gender-typed associations was determined by calculating differences in reaction time and accuracy in congruent and incongruent blocks resulting in a d-score (using scoring algorithm by Greenwald et al., 2003). A positive d-score represents stronger implicit gender-typed associations (e.g., math is for males, Dutch language for females). Negative d-scores represent counter-gender-typed associations (e.g., math is for females, Dutch language for males). To reduce possible order effects of the presentation of congruent and incongruent blocks, two versions of the IAT were used, one in which the congruent block was administrated first and one in which the incongruent block was administrated first (Nosek et al., 2005).

5.3.4 DATA ANALYSES

We used SPSS Statistics version 27 for data inspection and further analyses. All variables were inspected for possible outliers that were defined as values more than 3.29 SD under or above the mean (Tabachnick & Fidell, 2012). Two outliers were identified and winsorized by giving them the most extreme not outlying value (Tabachnick & Fidell, 2012). All continuous variables were normally distributed.

To examine whether teachers' grading bias is negatively associated with male students we plotted the distribution of the scores of the gender grading bias variable in histograms for Dutch language and math. Additional one sample t-tests were performed to test whether the mean scores were significantly different from zero (hypothesis 1). Differences between male and female teachers' gender grading bias in Dutch language and math were examined using independent t-tests (hypothesis 2).

To test hypotheses 3 and 4, the serial multiple mediation model (Model 6) in PROCESS was run to incorporate bootstrapping techniques for estimating indirect effects (Hayes, 2018). Using this model we estimated the effects of implicit gender-typed associations and expectations on gender grading bias controlled for teachers' gender (hypothesis 3), as well as the indirect effect of teachers' gender on gender grading bias through implicit gender-typed associations and expectations (hypothesis 4). We ran the model for expectations regarding students' talent and effort separately. Indirect effects are significant when the confidence intervals do not include zero.

The main analyses provided insight in the strength of the associations between the main variables, but lacked information about the direction of the effects of implicit gender stereotypes on grading biases against either male or female students. To explore these directions, scatterplots with regression lines were produced and additional post hoc analyses were performed with separate outcome variables for grading bias against male students and grading bias against female students (results are presented in appendix B).

The gender grading bias variable in the main analyses included scores from -3 to +3 with negative scores representing grading bias against females and positive scores representing grading bias against male students. In the post hoc analyses, the gender grading bias variable was separated for a grading bias against male students and a grading bias against female students. In the variable for grading bias against male students, we gave teachers with biases against female students score 0. In the variable for grading bias against female students, we gave teachers with biases against male students score 0. This corresponds with the method of Liben and Bigler (2002) to infer stereotypical attitudes towards others, rather than contra-stereotypical attitudes, using questionnaire data. A higher score on the variables in the post hoc analyses indicated a stronger grading bias against either male or female students.

To test the robustness of our results we conducted the analyses with the control variables age, experience, and test version.

Table 5.1. Mean Scores and Pearson Correlations on Central Variables of the Study

			Mean (SD)	Min, Max	Pearson Correlations			
					1	2	3	4
Dutch	1	Gender Grading Bias	0.023 (0.533)	-1.14, 1.64				
Language	2	Gender Teacher	0.760 (0.427)	0, 1	-0.025			
N = 131	3	Gender-typed Associations	0.488 (0.424)	-0.51, 1.35	0.069	0.512**		
	4	Expectations Talent	0.038 (0.490)	-1.39, 1.55	0.727**	-0.050	0.008	
	5	Expectations Effort	0.114 (0.499)	-1.36, 1.25	0.410**	-0.119	0.004	0.404**
Math	1	Gender Grading Bias	0.019 (0.828)	-2.36, 2.64				
N = 227	2	Gender Teacher	0.560 (0.497)	0, 1	0.012			
	3	Gender-typed Associations	0.386 (0.431)	-0.74, 1.33	-0.031	-0.311**		
	4	Expectations Talent	-0.129 (0.480)	-1.64, 1.09	0.442**	-0.024	-0.119	
	5	Expectations Effort	0.284 (0.517)	-1.27, 1.82	0.116	0.004	0.122	-0.042

Note: Gender, 0 = male, 1 = female, Expectations Talent = Implicit gender-typed expectations regarding students' talent, Expectations Effort = Implicit gender-typed expectations regarding students' effort. ** $p < 0.01$, * $p < 0.05$

5.4 RESULTS

5.4.1 GENDER GRADING BIAS ON THE AGGREGATE LEVEL

Table 5.1 presents the means, standard deviations and range of gender grading bias in Dutch language and math. The distributions of the scores on gender grading bias are depicted in Figure 5.3. Negative scores indicate a grading bias against female students, positive scores indicate a grading bias against male students. In Dutch language, the mean gender bias in grading is 0.023 ($SD = 0.533$), indicating that Dutch language teachers on average did not differentiate in their grading based on students' gender ($t(130) = 0.497$, $p = 0.620$, $d = 0.043$). In math, the average score on the gender grading bias variable is 0.019 ($SD = 0.828$), indicating that math teachers did not differentiate in their grading of male and female students either ($t(226) = 0.338$, $p = 0.736$, $d = 0.022$). Thus, hypothesis 1 should be rejected.

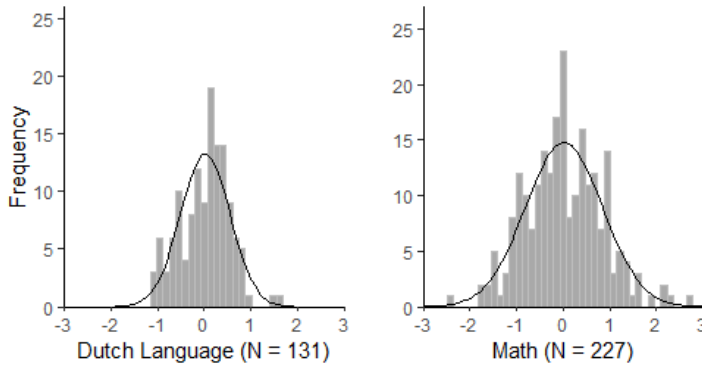


Figure 5.3. Histograms Gender Grading Bias

Nevertheless, Table 5.1 and Figure 5.3 show that there is ample variation in gender grading biases on the individual level. In Dutch language, teachers' grading varies from a bias against female students, rewarding male students up to 1.14 point higher than female students, to a bias against male students, rewarding female students up to 1.64 points higher than female students (on a scale from 1 to 10). In math, in the most extreme cases math teachers graded male students up to 2.36 points higher than female students, or graded female students up to 2.64 points higher than male students (on a scale from 1 to 26).

5.4.2 TEACHERS' GENDER AND GENDER GRADING BIAS

Next, we tested whether the variation in teachers' gender grading bias is associated with teachers' gender. Table 5.1 reveals that teachers' gender grading bias is not correlated with teachers' gender. Independent sample t-tests confirm that gender grading bias among male Dutch language teachers ($M = 0.047$, $SD = 0.519$) is not significantly different from the gender grading bias among female Dutch language teachers ($M = 0.016$, $SD = 0.540$), $t(129) = 0.283$, $p = 0.616$, $d = 0.058$. Also for math, gender grading bias among male teachers ($M = 0.007$, $SD = 0.838$) is not significantly different from gender grading bias among female teachers ($M = 0.027$, $SD = 0.824$), $t(225) = -0.180$, $p = 0.858$, $d = -0.024$. Therefore, hypothesis 2 should be rejected.

5.4.3 TEACHERS' IMPLICIT GENDER STEREOTYPES AND GENDER GRADING BIAS

Table 5.1 presents the mean scores and correlations of teachers' implicit gender-typed associations and implicit gender-typed expectations. On average, teachers in both subjects tend to associate math with masculinity and languages with femininity (Dutch language $t(130) = 13.158$, $p = 0.000$, $d = 1.150$, math $t(226) = 13.487$, $p = 0.000$, $d = 0.895$). Further, both Dutch language and math teachers expect female students to invest more effort in preparing an exam than male students (Dutch language $t(130) = 2.622$, $p = 0.010$, $d = 0.229$, math $t(226) = 8.279$, $p = 0.000$, $d = 0.549$). Additionally, math teachers expect male students to have

more talent in math than female students ($t(226) = -4.043, p = 0.000, d = -0.268$). Table 5.1 shows that teachers' gender-typed associations are not correlated with gender grading bias in either Dutch language or math. However, teachers' gender-typed expectations were positively correlated with gender grading bias. In Dutch language, both gender-typed expectations regarding students' talent and effort were positively associated with teachers' gender bias in grading. In math, only teachers' expectations regarding students' talent were positively associated with gender grading bias.

Models 3 and 6 in Table 5.2 present the results of regressing teachers' implicit gender-typed associations and implicit gender-typed expectations regarding students' talent (top half of Table 5.2) and effort (bottom half of Table 5.2) on gender grading bias controlled for teachers' gender (hypothesis 3). For Dutch language, Model 3 confirms that teachers' implicit gender-typed associations are not related to gender grading bias. The extent to which teachers associate math with masculinity and languages with femininity is not related to gender bias in teachers' grading. This Model also confirms that teachers' gender-typed expectations regarding students' talent and effort are positively associated with gender grading bias. Figure 5.4 indicates that teachers who expect male students to have more talent in Dutch language than female students, give male students more points for their answers than female students, despite the equal quality of the tests. Teachers who expect female students to have more talent in Dutch language than male students, appear to give female students more points for their answers than male students. The same is true for the relation between teachers' gender-typed expectations regarding students' effort and gender grading bias. Post-hoc analyses confirm this image (see appendix B).

For math, Model 6 in Table 5.2 shows a similar pattern. Teachers' implicit gender-typed associations are not related to gender grading bias, neither are teachers' gender-typed expectations regarding students' effort. However, teachers' gender-typed expectations regarding students' talent is positively associated with gender grading bias. Also for math, Figure 5.4 indicates that the positive relation between teachers' expectations and gender grading bias works both ways: teachers who expect male students to have more talent in math compared to female students, give male students more points for their answers compared to female students. Teachers who expect female students to have more talent in math than male students, give female students more point for their answers compared to male students. Post-hoc analyses confirm this pattern regarding the association between teachers' expectations of students' talent and gender grading bias. Furthermore, the post-hoc analyses show that teachers' gender-typed expectations regarding students' effort are significantly positively associated with a grading bias against female students, but not with a grading bias against male students. Teachers who expect male students to put more effort in preparing for an exam than female students give male students more points for their answers than female students, whereas teachers who expect female students to put more effort in preparing for an exam than male students do not give female students more points for their answers than male students (see appendix B).

Table 5.2. Double Mediation of Implicit Gender-typed Associations and Expectations regarding Talent (top-half) and Effort (bottom-half) in the Relation Between Teachers' Gender and Gender Grading Bias

	Dutch Language (N = 131)			Math (N = 227)		
	(1) IA	(2) IE	(3) GGB	(4) IA	(5) IE	(6) GGB
Constant	0.099 (0.066)	0.076 (0.089)	-0.027 (0.067)	0.539 (0.041)	-0.032 (0.064)	0.063 (0.100)
Gender Teacher	0.509** (0.075)	-0.084 (0.118)	-0.036 (0.088)	-0.270** (0.055)	-0.065 (0.067)	0.055 (0.106)
Implicit Associations		0.053 (0.119)	0.098 (0.089)		-0.155* (0.077)	0.063 (0.122)
Expectations Talent			0.789** (0.066)			0.772** (0.105)
R ²	0.262**	0.004	0.534**	0.097**	0.018	0.197**
Constant	0.099 (0.066)	0.210 (0.090)	-0.059 (0.091)	0.539 (0.041)	0.194 (0.069)	0.001 (0.112)
Gender Teacher	0.509** (0.075)	-0.192 (0.119)	-0.018 (0.119)	-0.270** (0.055)	0.048 (0.073)	-0.005 (0.117)
Implicit Associations		0.104 (0.120)	0.094 (0.118)		0.163 (0.084)	-0.089 (0.135)
Expectations Effort			0.436** (0.087)			0.194 (0.108)
R ²	0.262**	0.020	0.173**	0.097**	0.017	0.015

Note: Bootstrap estimates derived from 10 000 bootstrap samples along with 95% CIs are reported. Standard Deviations in parentheses. IA = Implicit Gender-typed Associations, IE = Implicit Gender-typed Expectations, GGB = Gender Grading Bias, ** p<0.01, * p<0.05

In all, hypothesis 3 can only partly be accepted. Implicit gender-typed associations are not directly associated with gender grading bias. However, teachers' gender-typed expectations regarding students' talent is related to gender grading bias. Evidence for a role of gender-typed expectations regarding students' invested effort is mixed.

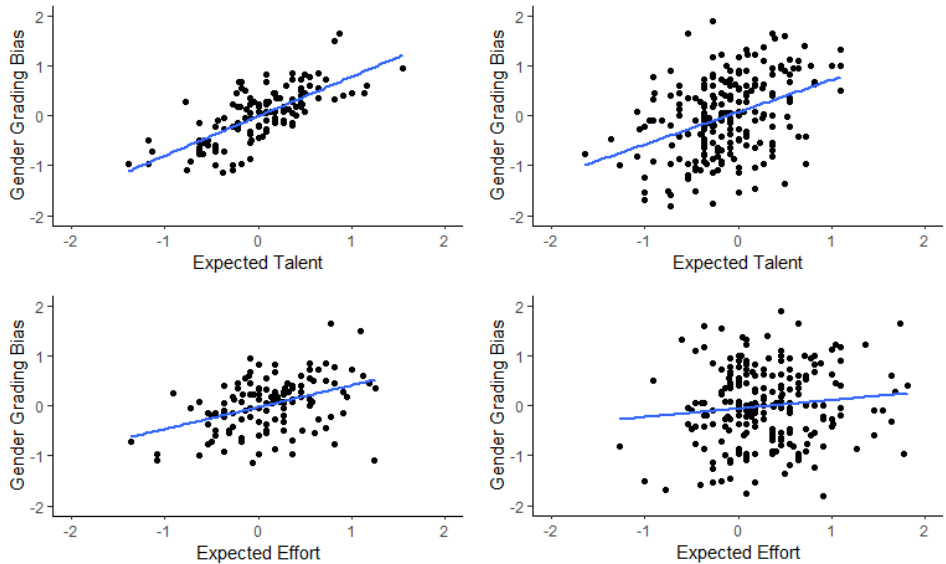


Figure 5.4. Plots of the Relation between Gender-Typed Expectations and Gender Grading Bias
Note: negative scores indicate biases against female students, positive scores indicate biases against male students.

5.4.4 THE INDIRECT RELATION BETWEEN GENDER AND GENDER GRADING BIAS

Table 5.1 reveals that teachers' gender is correlated with teachers' gender-typed associations in Dutch language and math. In Dutch language, gender and gender-typed associations are positively correlated, indicating that male Dutch language teachers have weaker gender-typed associations than female Dutch language teachers. In math, gender and gender-typed associations are negatively correlated, indicating that female math teachers have weaker gender-typed associations than male math teachers.

Despite the absence of a direct effect of teachers' gender on their gender grading bias, we estimated the indirect effect of teachers' gender on gender grading bias because a correlation between X and Y is not necessarily needed as a precondition for a mediation effect (Hayes, 2018). Table 5.2 presents the findings. Models 1 and 4 include the findings of regressing teachers' gender on their implicit gender-typed associations. Models 2 and 5 present the findings of regressing both teachers' gender and implicit gender-typed associations on their gender-typed expectations regarding students' talent (top half

of Table 5.2) and effort (bottom half of Table 5.2). As mentioned above, Models 3 and 6 show the results of regressing gender, implicit gender-typed associations and gender-typed expectations on gender grading bias. The mediation effects of implicit gender-typed associations and gender-typed expectations, as well as the double mediation effect and the total effect of the model are presented in Table 5.3.

For Dutch language, the findings are presented in the first three Models. Model 1 confirms that implicit gender-typed associations are stronger among female Dutch language teachers compared to male Dutch language teachers. As expected, male Dutch language teachers have less stereotypic associations between gender and math and languages. Model 2 shows that both gender and implicit gender-typed associations are not related to gender-typed expectations regarding students' talent and effort. Contrary to what was expected, strong implicit gender-typed associations are not related to making a stronger distinction between male and female students regarding their talent and effort in Dutch language. As described above, Model 3 shows that teachers' gender-typed expectations are related to gender grading bias.

For math, the findings are presented in the last three Models. Model 4 confirms that implicit gender-typed associations are stronger among male math teachers compared to their female counterparts. As expected, female math teachers have milder gender-typed associations than male math teachers. Model 5 reveals that teachers' gender is not associated with gender-typed associations regarding students' talent and effort in math. Teachers' implicit gender-typed associations are negatively related to their gender-typed expectations regarding students' talent, but not related to teachers' expectations regarding students' effort. As described above, Model 3 shows that only teachers' gender-typed expectations regarding students' talent are positively related to gender grading bias.

Table 5.3 summarizes the effects of teachers' gender on gender grading bias, through implicit gender-typed associations and expectations. For Dutch language, no indirect effects were found for implicit gender-typed associations and gender-typed expectations (or both). For math, implicit gender-typed associations and expectations could not explain differences in male and female teachers' gender grading bias on their own. However, a positive double mediation effect was found: differences in male and female teachers' gender grading bias can be explained through differences in male and female teachers' implicit gender-typed associations and expectations regarding students' talent. Post-hoc analyses reveal that teachers' implicit gender-typed associations and expectations regarding students' talent can explain differences in male and female math teachers' grading biases against male students and female students. Both for the grading bias against male students and the grading bias against female students applies that female math teachers have weaker associations between math and masculinity, that is related to expecting approximately the same of male and female students in terms of talent, which is related to having weaker (closer to zero) gender grading bias. Furthermore, the post-hoc analyses show that implicit gender-typed associations and expectations regarding students' effort can explain differences in male

and female teachers grading biases against female students. Female math teachers have weaker gender-typed associations that lead to expecting male students to put less effort in math which is related to having a stronger grading bias against female students (see appendix B). In all, hypothesis 4 should be accepted in the subject math, but rejected in the subject Dutch language.

Table 5.3. Indirect Effects of Gender on Gender Grading Bias

	Dutch Language			Math		
	Effect Size (SE)	LLCI	ULCI	Effect Size (SE)	LLCI	ULCI
Total Mediated Effect	0.004 (0.098)	-0.192	0.192	-0.035 (0.055)	-0.145	0.073
Mediation through IA	0.050 (0.050)	-0.051	0.146	-0.017 (0.032)	-0.084	0.046
Mediation through IE	-0.066 (0.086)	-0.239	0.098	-0.050 (0.053)	-0.155	0.053
Double mediation	0.021 (0.042)	-0.059	0.110	0.032 (0.019)	0.000	0.073
Total Mediated Effect	-0.013 (0.087)	-0.187	0.155	0.025 (0.042)	-0.055	0.111
Mediation through IA	0.048 (0.068)	-0.090	0.182	0.024 (0.039)	-0.052	0.107
Mediation through IE	-0.084 (0.050)	-0.186	0.011	0.009 (0.017)	-0.022	0.051
Double mediation	0.023 (0.032)	-0.032	0.095	-0.009 (0.007)	-0.026	0.002

Note: Total Mediated Effect is the sum of the indirect effects of $X \rightarrow Y$. Size of bootstrap sample for calculations of the indirect effect = 10 000. LLCI = bootstrapped lower level confidence interval; ULCI = bootstrapped upper level confidence interval, 95% confidence interval, Effect size is significant when LLCI and ULCI do not include 0.

Finally, the pattern of results for the model was similar when controlling for teachers' age, experience (Lavy, 2008), and the version of the tests that were graded by the teachers, though it should be noted that for math the significance level of the effect of implicit gender-typed associations on gender-typed expectations regarding students' talent and the double mediation effect were no longer significant on the 0.05 level, but on the 0.1 level.

5.5 CONCLUSION AND DISCUSSION

By means of a field experiment among teachers in training in the subjects Dutch language and math in secondary education in the Netherlands, this study tested the influence of students' gender on teachers' grading and examined the role of teachers' gender and implicit gender stereotypes in gender grading bias. We found that teachers' grading, on average, was not gender biased. However, we did find that individual variation in gender grading bias is associated with teachers' implicit expectations of male and female students' talent and effort. Further, we found an indirect effect of teachers' gender on gender grading bias in math via teachers' implicit gender-typed associations and expectations.

Contrary to what we expected (e.g., Protivínský & Münich, 2018), in our field experiment we did not find an association between perceived male gender of students and teachers' grading. Instead, we found that teachers' grading, on average, did not differ between male

and female students. This finding is similar to the few studies that also experimentally tested the influence of students' gender on teachers' grading (Hanna & Linden, 2012; Hinnerich et al., 2011), suggesting that students' gender and teachers' grading might not be *causally* related. Previous studies that found grading bias against male students, all used comparative designs to examine gender grading bias, i.e., comparing scores on tests that were blindly graded to scores on tests that were graded by a teacher (e.g., Protivínský & Münich, 2018). The observed bias against male students in teachers' grading in these studies might not be a direct result of students' gender but rather a reflection of teachers' beliefs about male and female students. Indeed, one of these previous studies found that an effect of grading bias against male students disappeared when controlling for teachers' reports on students' attitudes towards learning (Cornwell et al., 2013). Although Cornwell et al. (2013) concluded that differences in male and female students' behavior could fully explain gender bias in teachers' grading, based on the results in our study we argue that it could have been a gender bias in teachers' perception of male and female students' behavior that explained gender grading bias.

Although we did not find a gender bias in teachers' grading on average, we did find ample variation in gender grading bias between teachers. Our study showed that the variation in gender grading bias was associated with variation in teachers' expectations of male and female students' talent and effort despite the equal quality of the tests of male and female students. Teachers with more equal expectations of male and female students displayed less gender bias in grading. Teachers who differed in their expectations of male and female students displayed more grading bias against both male and female students, despite the equal quality of their tests. Our finding confirmed previous research that demonstrated that the extent to which a person distinguishes between males and females based on generalized preconceptions is related to their discriminative behaviors against males and females (e.g., Burgess & Borgida, 1999; Dasgupta, 2004), and that different expectations can lead to different assessments (Muntoni & Retelsdorf, 2018; Powell, 1986).

Additionally, to get a complete picture of the role of implicit gender-typed expectations in gender grading bias our study explored the full spectrum of gender-typed expectations including stereotypical and contra-stereotypical expectations. Based on previous research we expected gender stereotypes to differ mainly in strength, and to lesser extent in direction (e.g., Lenton et al., 2009). However, the post-hoc analyses showed that teachers' implicit gender-typed expectations differed in both strength and direction, and that both stereotypical and contra-stereotypical expectations were related to discriminative behaviors. Dutch language and math teachers' grading biases against male or female students were in line with the content of their biases in expectations. Despite the equal quality of the tests, in both subjects, teachers who expected male students to have more talent, gave male students more points for their answers. Teachers who expected female students to have more talent, gave female students more points for their answers. The same was true for Dutch language teachers' expectations regarding students' effort: expecting more effort

from male students was related to rewarding male students more points for their answers, and expecting more effort from female students was related to rewarding female students more points for their answers. The explorations support previous research in the family context that found that fathers' stereotypical and contra-stereotypical expectations of males and females are related to opposite differential treatments of males and females (Endendijk et al., 2013). It should be noted that, in math, only teachers' contra-stereotypical expectations of students' effort were associated with gender biases against female students. Math teachers who had the contra-stereotypical expectation that male students put more effort in the preparation of a test, gave male students more points for their answers, whereas math teachers who expected female students to put more effort, did not give female students more points for their answers. Putting effort is gender stereotyped as something female students generally do, and male students generally do not do (e.g., Heyder and Kessels, 2007). Therefore, expecting students to have put effort could be praised and result in positive effects in teachers' grading for male students but not for female students. This finding supports literature on double standards for men and women in, among others, parenting roles (Deutsch & Saxon, 1998), on the work floor (Lyness & Heilman, 2006), and in engaging in sexual behaviors (Endendijk et al., 2020; Sagebin Bordini & Sperb, 2013).

Furthermore, similar to previous research (Martin & Dinella, 2012; Smeding, 2012), our results showed that female teachers in a stereotypically masculine subject (math) and male teachers in a stereotypically feminine subject (languages) had weaker associations between math and masculinity and between language and femininity than male math teachers and female Dutch language teachers. Our finding supports the assumption that an individual's gender stereotypes are flexible and that these gender stereotypes can become less strict by experiencing or observing rejection of the gender stereotype (Bussey & Bandura, 1999; Charlesworth & Banaji, 2021; Groeneveld et al., 2021). Though it should be noted that some scholars have argued that gender stereotypes are resistant to change (Fiske, 2017; Haines et al., 2016). Additionally, previous studies showed that implicit gender stereotypes were similar among males and females (Banaji & Greenwald, 1995; Rudman & Kilianski, 2000), or somewhat stronger among females than males (Endendijk et al., 2013; Nosek et al., 2002a). Our study showed that both males and females can have weaker implicit gender stereotypes and that these seem to be dependent on someone's own experiences rather than someone's gender perse.

We found that these weaker implicit gender-typed associations among female math teachers were related to weaker gender-typed expectations of students and that this mechanism could explain weaker gender grading biases among female math teachers than male math teachers. Explorations of the direction of the mediating effect showed that, in math, female teachers' weaker gender stereotypical associations (math is for men and women, languages is for men and women) were related to weaker gender-stereotypical expectations of students' talent (males and females are talented in math). This finding supports previous findings that teachers' stereotypes and expectations are positively

related (Muntoni & Retelsdorf, 2018). Consequently, weaker (more equal) gender-typed expectations regarding students' talent resulted in a weaker grading bias against both males and females. Weaker gender-typed expectations regarding students' effort could only explain differences between male and female math teachers' grading bias against female students, but it should be noted that this effect size was fairly small.

To our knowledge, our study is the first to find evidence for a psychological mechanism underlying gender grading bias. However, we did not find evidence for the mechanism in Dutch language teachers. Male Dutch language teachers did have weaker gender-typed associations than female Dutch language teachers, but these were not related to weaker gender-typed expectations. This may be because gender stereotypes regarding students' talent in language are not as persuasive as in math. Large government campaigns were directed at enhancing female participation in math, but not at enhancing male participation in languages. Furthermore, many famous Dutch writers are male, providing convincing contra-gender stereotypical examples in Dutch language. It thus seems that gender salience is important for effects of implicit gender stereotypes in gender grading bias to occur (Keiser et al., 2002).

The findings of this study should be interpreted in light of its limitations. The balancing between the maximalization of this experiment's ecological validity and practical feasibility resulted in the exclusion of real-life student-teacher interactions that could play an important role in gender grading bias (Falch & Naper, 2013). Moreover, the study design resulted in the selection of two experimental groups of tests that were not of the exact same quality. The group of tests with male names were of approximately the same quality as the group of tests with female names based on the experts' blind grading. However, as the tests in the two groups were not identical in terms of students' answers, differences in the students' answers might have resulted into differences between the two groups that affect gender grading bias. Further, although implicit measures are encouraged to study differential treatment of male and female students (Denessen et al., 2022), in particular the IAT is criticized for its validity and reliability in predicting discriminative behaviors (Blanton et al., 2009). However, for the current study the Arts-Science version of the IAT was highly relevant because of the theoretical link between the content of this test and our outcome measure, i.e., gender grading bias in the subjects math and languages. Despite the limitations of the use of different and real-student answers and the implicit measures of gender stereotypes, these components of this study's design contributed in particular to the participants not being aware of what the research was about. Therefore, social desirability and conscious behaviors, which participants often have when they participate in social research (Wulff & Villadsen, 2019), were probably prevented.

In all, the results in our study question previous research on the role of students' behaviors in gender grading bias that relied on teacher reports (Cornwell et al., 2013). Our study encourages future research to be conscious of gender bias in teacher reports when relying on teacher reports to explain differences between male and female students. Furthermore,

we recommend further exploration of the psychological mechanisms underlying gender grading bias and empirically test the role of gender stereotypes and teachers' gendered beliefs. Future research could tap into the role of teachers' explicit gender stereotypes, or the extent to which teachers' gendered beliefs and gender grading bias are causally related to unravel the processes that can result in gender biases in teachers' grading.

Further, our results have implications for teacher training programs. Teacher training programs might want to contribute to increasing awareness of teachers (in training) about the role of their own gender-typed expectations of male and female students in grading practices. Improving teachers' self-awareness can stimulate their quality of teaching (Schussler et al., 2010). We are working on translating the setup of this study's experiment into an educational tool directed at increasing teachers' awareness of their own gender (grading) bias and improving teachers' agency in equal treatment of male and female students. Teachers achieve agency (i.e., actions that are assumed to contribute to good and meaningful education (Biesta et al., 2015) when they are able to intentionally choose behaviors instead of just behave in a routinized matter without considering alternatives (Priestley et al., 2015). Awareness of one's own gender-typed expectations of male and female students and gender grading bias, might help teachers to tackle automatic thoughts and behaviors and to treat males and females more equally, both in interaction with and evaluation of students. Indeed, previous research demonstrated that awareness of one's own gender bias can contribute to reducing gender bias in an education setting (Carnes et al., 2015; Devine et al., 2017; Forscher et al., 2017; Girod et al., 2016). It should be noted that it is crucial that gender equality is valued by the individual teacher as well as by the (school) collective (Biesta et al., 2015). This can be a complicating factor as the extent to which gender inequality is perceived (see for instance Kinias & Kim, 2011), gender equality in education is valued, and ideas about the routes towards equal treatment can differ between individual teachers and across cultures (see Cardona López et al., 2018). Besides the importance of discussions between teachers, teacher trainers, and policy makers 'to get everyone on the same page' (Cardona López et al., 2018), we believe that courses that increase awareness of teachers' own gender biases are critical as gender bias often happens unconsciously and its consequences are therefore often overlooked or underestimated. Teacher training programs could be a suitable place to increase teachers' (in training) self-awareness of gender stereotypes and agency in equal treatment of boys and girls (De Boer et al., 2019).

To conclude, our study pointed out that rather than a causal relation between students' gender and teachers' grading, the role of teachers' implicit gendered beliefs about male and female students should be considered. Not only were teachers' implicit gender-typed expectations related to gender grading bias, implicit gender-typed associations and expectations together could explain differences in male and female math teachers' grading biases. Students' gender and teachers' gender do not directly translate in differential grading practices, instead, teachers' beliefs about males and females can affect their perception and assessment of male and female students.

5.6 ACKNOWLEDGEMENTS

The authors would like to thank Arthur Pormes, Anneke Wurth and Peter Kop for their assistance in the design of the field experiment. Further, the authors would like to thank Anne Floor Lubbers for her effort during the experiment's data collection.

5.7 APPENDIX

5.6.1 APPENDIX A – MANIPULATION

	Dutch male names	Dutch female names	Turkish male names	Turkish female names
1	Bastiaan Driessen	Sara Wolters	Yusuf Koc	Leyla Bulut
2	Thomas van der Horst	Meike Groen	Murat Aslan	Sevda Avcı
3	Daniël van der Pol	Juliette Kramer	Hakan Yilmaz	Kubra Demir
4	Benjamin de Lange	Madelief Scholten	Ahmet Cicek	Merve Günes
5	Pieter Smeets	Fien de Ruiter		
6	Teun Bakker	Emma Hoekstra		
7	Lars Willemsen	Tess van Veen		
8	Florian van Dijk	Olivia Maas		

5.6.2 APPENDIX B – BIAS AGAINST MALE OR FEMALE STUDENTS

Table: Gender Grading Bias Against Male students.

Double Mediation of Implicit Gender-typed Associations and Gender-typed Expectations in the Relation Between Teachers' Gender and Gender Grading Bias against Male Students

	Dutch Language (N = 131)			Math (N = 227)		
	(1) IA	(2) IE	(3) GGB	(4) IA	(5) IE	(6) GGB
Constant	0.099 (0.066)	0.076 (0.089)	0.191 (0.045)	0.539 (0.041)	-0.032 (0.064)	0.398 (0.062)
Gender	0.509** (0.075)	-0.084 (0.118)	-0.028 (0.059)	-0.270** (0.055)	-0.065 (0.067)	0.011 (0.066)
Teacher						
Implicit		0.053 (0.119)	0.086 (0.060)		-0.155* (0.077)	-0.036 (0.076)
Associations						
Expectations			0.363** (0.044)			0.430** (0.065)
Talent						
R ²	0.262**	0.004	0.356**	0.097**	0.018	0.170**
Constant	0.099 (0.066)	0.210 (0.090)	0.167 (0.052)	0.539 (0.041)	0.194 (0.069)	0.370 (0.069)
Gender	0.509** (0.075)	-0.192 (0.119)	-0.012 (0.068)	-0.270** (0.055)	0.048 (0.073)	-0.020 (0.071)
Teacher						
Implicit		0.104 (0.120)	0.080 (0.068)		0.163 (0.084)	-0.114 (0.083)
Associations						
Expectations			0.243** (0.050)			0.072 (0.066)
Effort						
R ²	0.262**	0.020	0.171**	0.097**	0.017	0.012

Note: Standard Deviations in parentheses. IA = Implicit Gender-typed Associations, IE = Implicit Gender-typed Expectations, GGB = Gender Grading Bias, ** p<0.01, * p<0.05

Table: Gender Grading Bias against Males

Indirect Effects of Gender on Gender Grading Bias against Male Students

	Dutch Language			Math		
	Effect Size (SE)	LLCI	ULCI	Effect Size (SE)	LLCI	ULCI
Total Mediated Effect	0.023 (0.050)	-0.079	0.117	0.000 (0.033)	-0.065	0.065
Mediation through IA	0.044 (0.031)	-0.018	0.103	0.010 (0.019)	-0.028	0.048
Mediation through IT	-0.031 (0.041)	-0.116	0.044	-0.028 (0.031)	-0.093	0.030
Double mediation	0.010 (0.020)	-0.027	0.052	0.018 (0.011)	0.000	0.043
Total Mediated Effect	0.007 (0.045)	-0.085	0.093	0.031 (0.025)	-0.014	0.084
Mediation through IA	0.041 (0.035)	-0.029	0.107	0.031 (0.024)	-0.013	0.083
Mediation through IT	-0.047 (0.029)	-0.111	0.004	0.004 (0.008)	-0.010	0.023
Double mediation	0.013 (0.017)	-0.017	0.052	-0.003 (0.004)	-0.012	0.003

Note: Standard Deviations in parentheses. IA = Implicit Gender-typed Associations, IE = Implicit Gender-typed Expectations, GGB = Gender Grading Bias, ** p<0.01, * p<0.05

Table: Gender Grading Bias Against Female students.

Double Mediation of Implicit Gender-typed Associations and Gender-typed Expectations in the Relation Between Teachers' Gender and Gender Grading Bias against Female Students

	Dutch Language (N = 131)			Math (N = 227)		
	(1) IA	(2) IE	(3) GGB	(4) IA	(5) IE	(6) GGB
Constant	0.099 (0.066)	0.076 (0.089)	-0.218 (0.044)	0.539 (0.041)	-0.032 (0.064)	-0.335 (0.059)
Gender Teacher	0.509** (0.075)	-0.084 (0.118)	-0.008 (0.058)	-0.270** (0.055)	-0.065 (0.067)	0.044 (0.062)
Implicit Associations		0.052 (0.119)	0.012 (0.058)		-0.155* (0.077)	0.098 (0.072)
Expectations Talent			0.425** (0.043)			0.342** (0.062)
R ²	0.262**	0.004	0.437**	0.097**	0.018	0.123**
Constant	0.099 (0.066)	0.210 (0.090)	-0.226 (0.056)	0.539 (0.041)	0.194 (0.069)	-0.370 (0.063)
Gender Teacher	0.509** (0.075)	-0.192 (0.119)	-0.007 (0.074)	-0.270** (0.055)	0.048 (0.073)	0.015 (0.066)
Implicit Associations		0.104 (0.120)	0.014 (0.073)		0.163 (0.084)	0.025 (0.076)
Expectations Effort			0.193** (0.054)			0.122* (0.060)
R ²	0.262**	0.020	0.094**	0.097**	0.017	0.020

Note: Standard Deviations in parentheses. IA = Implicit Gender-typed Associations, IE = Implicit Gender-typed Expectations, GGB = Gender Grading Bias, ** p<0.01, * p<0.05

Table: Gender Grading Bias against Females

Indirect Effects of Gender on Gender Grading Bias against Female Students

	Dutch Language			Math		
	Effect Size (SE)	LLCI	ULCI	Effect Size (SE)	LLCI	ULCI
Total Mediated Effect	-0.018 (0.055)	-0.124	0.090	-0.035 (0.030)	-0.092	0.022
Mediation through IA	0.006 (0.031)	-0.055	0.069	-0.027 (0.022)	-0.072	0.016
Mediation through IT	-0.036 (0.046)	-0.128	0.055	-0.022 (0.023)	-0.066	0.026
Double mediation	0.011 (0.023)	-0.033	0.059	0.014 (0.008)	0.000	0.032
Total Mediated Effect	-0.020 (0.052)	-0.126	0.082	-0.006 (0.025)	-0.057	0.042
Mediation through IA	0.007 (0.044)	-0.080	0.093	-0.007 (0.023)	-0.054	0.038
Mediation through IT	-0.037 (0.024)	-0.088	0.005	0.006 (0.011)	-0.012	0.031
Double mediation	0.010 (0.015)	-0.016	0.046	-0.005 (0.004)	-0.016	0.000

Note: Standard Deviations in parentheses. IA = Implicit Gender-typed Associations, IE = Implicit Gender-typed Expectations, GGB = Gender Grading Bias, ** p<0.01, * p<0.05



CHAPTER **SIX**

Conclusion

6.1 THE ROLE OF GENDER AND GENDER STEREOTYPES IN EDUCATION

Even though the relation between students' and teachers' gender and students' performance have often been studied in different scientific disciplines (e.g., Dee, 2005; Riccucci & Van Ryzin, 2017), still little is known about the mechanisms explaining these relations (Volman, Ten Dam, & Van Eck, 1995). Gender stereotypes have often been assumed to play a role in these explanatory mechanisms, but empirical tests are rare and solely situated in the 'females in math' context (Xu & Meier, 2021; Zhang, 2019). This dissertation aimed to provide more insight into the relation between student and teacher gender and student performance through theorizing and examining the role of gender stereotypes. By complementing insights from studies on representative bureaucracy in public administration with social psychological literature on gender stereotypes and social identity the relation between student-teacher gender congruence and student performance was scrutinized. In four empirical studies I specifically examined the contexts in which relations between student-teacher gender congruence and student performance occur as well as the mechanisms underlying these relations, from both the student and the teacher perspective.

This concluding chapter is structured as follows. Section 6.2 summarizes the findings of each empirical study. Then, Section 6.3 discusses the lessons learned from this research. Thereafter, Section 6.4 describes the theoretical and methodological implications, followed by Section 6.5 in which this research's limitations are discussed. Finally, section 6.6 concludes this dissertation and formulates the implications for education practice.

6.2 SUMMARY OF FINDINGS

6.2.1 WHEN AND WHERE: THE ROLE OF CONTEXT

Chapter 2 focused on the role of context in the relation between student-teacher gender congruence and students' performance. Based on literature on gender salience and gender stereotypes (e.g., Hilliard & Liben, 2010; Keiser et al., 2002), I suggested that the relation between student-teacher gender congruence and student performance would occur for female students in math and physics, and for male students in Dutch language and French language, because these are the subjects in which performance of these groups of students is negatively gender stereotyped. Negative gender stereotypes (i.e., men/women are bad at languages/STEM) are generally easier to internalize and more resistant to change than positive gender stereotypes (Baumeister et al., 2001). I therefore expected gender to be a salient characteristic for women in STEM and men in languages, enabling the relation between gender congruence and performance to occur in these settings. Indeed, analyses of administrative secondary school data showed that student-teacher gender congruence was positively related to female students' performance in math and physics but not in Dutch

language and French language, and to male students' performance in Dutch language and French language but not in math and physics.

Further, I argued that apart from the type of subject, the extent to which the gender stereotypes connected to STEM and languages play a role, may vary in different educational settings. I expected the relation between student-teacher gender congruence and student performance to be stronger in lower educational levels, in religious schools, and in schools located in less populated areas, because these are the educational contexts in which gender stereotypes may be relatively strong (De Vries et al., 2022; Istenič, 2007; Turner et al., 2019). However, the results showed distinct relations for the different educational contexts for male and female students across the subjects which did not correspond with the suggested role of gender stereotypes in these contexts.

6.2.2 HOW: THE STUDENT PERSPECTIVE

While no or contradictory effects of context were found, Chapter 2 also showed that, regardless of context, a relation between student-teacher gender congruence and student performance exists in subjects with gender stereotypes connected to it. More specifically, associations between student-teacher gender congruence and student performance were observed for groups of students in subjects in which the gender of a teacher can serve as a rejection of a negative gender stereotype connected to the subject. In Chapter 3 and 4, I proposed a mechanism that may underly these associations from the perspective of the student based on insights from social cognitive theory (Bussey & Bandura, 1999). The mechanism explains effects of the presence of a male or female teacher on students' performance through students' gender stereotypes and self-efficacy beliefs. In Chapter 3, we tested this association between student-teacher gender congruence, gender stereotypes, and self-efficacy beliefs. A vignette experiment among Dutch high school students showed significant gender differences in stereotypical beliefs and academic self-concepts of male and female students and a significant effect of female math teachers on both female and male students' self-beliefs in math. However, the experiment did not provide evidence for the hypothesized mechanism as students' gender stereotypes were not affected by teachers' gender.

In Chapter 4, I built on this study by including the association with student performance. We tested the full proposed mechanism in both math and Dutch language based on multisource data from a longitudinal research project. The results showed that student-teacher gender congruence was not related to female students' gender stereotypes, self-beliefs or performance in math or Dutch language. For male students, student-teacher gender congruence was not related to gender stereotypes, self-beliefs or performance in math. However, in Dutch language, student-teacher gender congruence was unexpectedly negatively associated with male students' performance. Although male students' gender stereotypes and self-beliefs were related to their Dutch language performance, teachers' gender was not associated with male students' gender stereotypes or self-beliefs.

6.2.3 HOW: THE TEACHER PERSPECTIVE

Finally, in Chapter 5, I studied the perspective of the teacher. I suggested that teachers' grading is gender biased because of teachers' gender-typed associations and expectations influencing students' administered performance (grades). Based on insights from gender schema theory (Bem, 1981), I argued that male teachers in Dutch language and female teachers in math may display less gender grading bias, because of weaker gender-typed associations and expectations. By means of a field experiment among teachers in training we tested the hypothesized mechanism. The experiment showed that teachers' grading, on average, was not gender biased. However, individual variation in gender grading bias was associated with teachers' implicit expectations of male and female students. Further, the experiment showed that female math teachers and male Dutch language teachers had weaker implicit gender stereotypes than their counterparts. In math, these weaker implicit gender stereotypes translated into less gender bias in expectations of students, which resulted in less gender grading bias.

6.3 LESSONS LEARNED

This dissertation provided three important lessons with regard to the role of gender and gender stereotypes in education. First, this dissertation showed that student and teacher gender are related to student performance. Each study within this dissertation provided evidence for an association between students' and teachers' gender and students' performance. Second, this dissertation showed that gender stereotypes are an important part of the puzzle that explains the role of gender in education. Each study within this dissertation showed that gender stereotypes in students and teachers still exist and that these gender stereotypes are associated with students' performance through students' beliefs of self and teachers' beliefs of others. The gender stereotypes of teachers were found to explain effects of teachers' gender on student performance through teachers' gendered associations, expectations, and gender grading bias.

However, and this is the third lesson with regard to the role of gender and gender stereotypes in education, context matters. Student and teacher gender do not always relate to student performance and when they do the magnitude and direction of the associations can vary in different contexts. This dissertation showed positive (Chapter 2 and 3), as well as negative (Chapter 4), and no associations (Chapter 4) between student-teacher gender congruence and students' performance. Further, gender stereotypes were found to explain effects of teacher gender on student performance, but only from the perspective of the teacher and only in the subject math. These findings stress the importance of contextual factors.

Although this dissertation examined the role of contextual factors (Chapter 2), a conclusive answer with regard to what contextual conditions strengthen or weaken the relation between student-teacher gender congruence and student performance could

not be given. On the macro level, we found that student-teacher gender congruence was related to better performance of male and female students in subjects in which students' performance is negatively gender stereotyped. We tested the existence of these relations in three educational contexts that were assumed to differ in the extent to which gender stereotypes play a role (i.e., students' educational level, schools' religiousness, and schools' location). I expected the associations between student-teacher gender congruence and student performance to be the strongest in lower educational levels, in religious schools, and in less-populated areas because these are the contexts in which gender stereotypes are assumed to be relatively strong. However, the data showed distinct relations between student-teacher gender congruence and performance in different educational contexts that did not correspond with the suggested role of gender stereotypes. For instance, on the macro level, student-teacher gender congruence was positively related to female students' performance and was not related to male students' performance in physics. Zooming in on the three different educational levels, student-teacher gender congruence was both positively and negatively related to female and male students' physics performance. Student-teacher gender congruence was positively related to female students' physics performance in the lower vocational track and the scientific track (lowest and highest educational level) and negatively related in the higher vocational track (moderate educational level). For male students, student-teacher gender congruence had a positive association with performance in the higher vocational track (moderate level) and a negative association with performance in the lower vocational track and scientific track (lowest and highest level). These distinct findings, that do not correspond with the hypothesized role of gender stereotypes, leave important questions and avenues for future research on the topic of gender and gender stereotypes in education which I will discuss in the next section.

6.4 THEORETICAL AND METHODOLOGICAL IMPLICATIONS

The role of gender and gender stereotypes in student performance has been the topic of study in different scientific disciplines. The theoretical framework in this dissertation was built based on studies on representative bureaucracy in public administration complemented with insights from social psychological theories. Therefore, in the following sections, the theoretical implications for the study of representative bureaucracy are discussed first, followed by the theoretical implications for the study of gender and gender stereotypes in education. In the last section I elaborate on the methodological implications of this research.

6.4.1 REPRESENTATIVE BUREAUCRACY

The study of representative bureaucracy within public administration has intensified in the past decade and focuses on the outcomes of representation within public organizations.

Representation within public organizations is assumed to benefit democratic values such as inclusiveness and equity (e.g., Sowa & Selden, 2003). Despite the increase of studies focusing on individual level effects of representation (e.g., Nicholson-Crotty et al., 2016; Song, 2018; Zhang, 2019), still little is known about the mechanisms underlying representation effects. When expanding on the mechanisms underlying representation it can be helpful to make use of theories from other disciplines. This dissertation made use of social psychological theories to propose and empirically test psychological mechanisms in students and teachers that may explain effects of gender representation.

The explanatory psychological mechanisms were centered around the role of gender stereotypes. The explanatory role of gender stereotypes in processes of gender representation have been suggested by public administration scholars (e.g., Keiser et al., 2002; Riccucci & Van Ryzin, 2017). However, theorization and empirical tests of such mechanisms are limited (Xu & Meier, 2021; Zhang, 2019). This dissertation hypothesized that the presence of representative teachers can influence students' gender stereotypes, self-efficacy beliefs, and performance explaining effects of representation from the student perspective. Further, this dissertation hypothesized that the presence of representative teachers can influence teachers' own gender stereotypes, expectations of male and female students, and evaluations of student performance explaining effects of representation from the perspective of the teacher. Although the studies within this dissertation did not provide evidence for these mechanisms to explain effects of gender representation (i.e., student-teacher gender congruence), these mechanisms can still be a starting point for representative bureaucracy scholars to further examine the role of gender stereotypes. In addition, this dissertation showed that gender stereotypes play out for a large part on an implicit level indicating that gender representation may unconsciously and unintentionally result in performance effects. Although representative bureaucracy scholars recognized these unconscious and unintentional processes in the perspective of the citizen/client, their role in the perspective of the bureaucrat are rarely acknowledged (Akram, 2018). Representative bureaucracy scholars should further explore the role of unconscious and unintentional gendered processes that may underly symbolic as well as active gender representation effects.

Further, the findings of the studies within this dissertation stress the importance of context in gender representation effects. Representative bureaucracy scholars have been suggesting that context is an important condition for processes of (gender) representation to occur (e.g., Keiser et al., 2002; Groeneveld et al., 2015; Meier, 2019). Hence, research on this topic has increased in the past decade and contributed to insights on contextual factors that facilitate or impede gender representation on the national (e.g., An et al., 2021) and organizational level (e.g., Song, 2018). This dissertation showed that contextual factors in the vicinity of an organization (e.g., educational level, religiousness, and location) are related to the occurrence and the direction (positive or negative) of representation effects. However, the suggested role of gender stereotypes (i.e., differences in the relative strength

of gender stereotypes in the different contexts) was unable to explain the distinct findings of student-teacher gender congruence. As a consequence, it remains unclear why effects of student-teacher gender congruence occur and result in positive or negative effects in particular educational contexts. Future research should try to unravel why student-teacher gender congruence has distinct effects in different contexts. Further, future research should further explore contextual factors in the vicinity of public organizations that may facilitate or impede positive as well as negative effects of gender representation to receive a better understanding of where, when, and why positive and negative effects of gender representation occur.

Lastly, similar to much research within the field of representative bureaucracy, this research was situated in the education sector (e.g., Keiser et al., 2002). Consequently, the proposed mechanisms and contextual factors explored within this dissertation apply to the student-teacher relationship. Although the suggested mechanisms and contextual factors may apply to other citizen-bureaucrat relationships, the relation between teachers and students (that is generally characterized by frequent and enduring interactions) is different from the relation between, for instance, police officers and citizens (that is generally characterized by infrequent and short interactions). Public administration scholars should further theorize and examine whether the explanatory mechanisms and contextual factors as proposed in this dissertation apply to other settings within the public sector.

6.4.2 THE STUDY OF GENDER AND GENDER STEREOTYPES IN EDUCATION

Apart from public administration, the role of student and teacher gender in student performance is studied in the fields of economics and education (e.g., Dee, 2005; Muntoni & Retelsdorf, 2018). Although several theories in the field of education studies elaborate on the role of gender in education (e.g., Thompson, 2003), little attention has been paid to the combination of student and teacher gender and its potential unique influence on student performance. Therefore, hypotheses on the effects of student-teacher gender congruence are often based on results of previous empirical studies rather than on gender or educational theories explaining why student and teacher gender can be related to student performance (e.g., Cho, 2012; Holmlund & Sund, 2008). Consequently, recent research on the topic of student and teacher gender and student performance focused on the existence of student-teacher gender congruence effects rather than on the explanation of these effects. To receive an understanding of the role of student-teacher gender congruence in student performance, and to be able to do something about them, knowledge on why and when effects of student-teacher gender congruence occur is necessary. This dissertation brought together theoretical insights from studies on gender representation (e.g., Keiser et al., 2002), social cognitive theory (Bussey & Bandura, 1999), and gender schema theory (Bem, 1981) to develop hypotheses on the relations between student and teacher gender, gender stereotypes, and student performance. In doing so, this dissertation showed the applicability and significance of different streams in the literature to build a theoretical framework for

understanding why and when effects of student-teacher gender congruence in education might occur. Future research might want to follow this example to further theorize the role of student-teacher gender congruence in education, and possibly try to generate and specify middle range theories on gender congruence in education, to move the study and understanding of student-teacher gender congruence on student performance beyond empirical observations.

The studies within this dissertation are among the first to empirically test mechanisms explaining effects of teacher gender on student performance. Teacher gender was proposed to influence student performance through the role of gender stereotypes and gender-typed expectations of self and others, from both the student and teacher perspective. The empirical tests of the hypothesized mechanisms only provided evidence for the explanatory role of teachers' gender stereotypes in the relation between teachers' gender and students' performance (through teachers' gender bias in expectations and grading). Although previous research demonstrated that teacher gender can affect students' gender stereotypes (e.g., Beilock et al., 2012), the studies within this dissertation provided no evidence for this relation. Based on this dissertation it seems more likely that teachers' gender stereotypes explain effects of teacher gender on student performance than students' gender stereotypes. At the same time, teachers' gender stereotypes and gender bias in expectations and grading were unable to explain effects of student-teacher gender congruence as teachers' grading was, on average, not gender biased. Future research should further explore the role of teachers' gender stereotypes as an explanation for effects of teacher gender on student performance. It might be interesting to explore interactions between teachers' gender stereotypes and other teacher characteristics such as effortful control, awareness of own gender bias, or gender ideology to receive a better understanding when teachers' gender stereotypes translate into student performance.

6.4.3 METHODOLOGICAL IMPLICATIONS

This dissertation showed the value and suitability of a variety in research angles, methods, and populations to study the role of gender and gender stereotypes in education. Most research on gender in education used international or national databases on student performance (e.g., TIMMS) to study effects of student and teacher gender (e.g., Cho, 2012). These databases rely on assessments and questionnaires to measure student performance and other relevant student, teacher, and school characteristics. Measurements through assessments and questionnaires are prone to bias, social desirability, and lack of motivation or fatigue which might result in invalid statements about the role of gender and gender stereotypes in education (Wulff & Villadsen, 2019). This dissertation showed the possibility and value of using register data to study real-life effects of student-teacher gender congruence on students' performance. Schools often use student administration systems to register information on students' demographic characteristics, school attendance, school performance, subject choice, teacher characteristics, and much more. These rich, large, and

longitudinal datasets are only rarely used for research purposes while the data is available and is not prone to response bias. When studying (causal) relations between gender and student performance, future research should make use of (longitudinal) register data more often.

In addition to longitudinal (register) data, experimental designs can provide insights into the causal effects of gender. Previous research often compared average grades of groups of students for the student-teacher gender dyad to study effects of student and teacher gender on student performance (e.g., Protivínský & Munich, 2018; Zhang, 2019). As a result, still little is known about causality in the relation between gender and student performance. This dissertation used a vignette experiment (Chapter 3) and a field experiment (Chapter 5) to study the causal effects of student and teacher gender on student performance. Future research should make use of experimental designs (or longitudinal designs) more often to explore the causality in relations between student and teacher gender and student performance.

Lastly, this dissertation demonstrated the role of *implicit* gender stereotypes in education indicating that gender and gender stereotypes play an unconscious and unintentional role in student performance. Especially in countries in which gender equality is considered important, such as the Netherlands where this research was conducted, implicit measures may provide a better understanding of the working of gender and gender stereotypes in education. Future research should consider the unconscious and unintentional role of gender and gender stereotypes and should use research designs that allow for measuring implicit biases to unravel effects of gender in education such as qualitative content analyses or classroom observations.

6.5 LIMITATIONS

This research also has some limitations that need to be discussed. First, the variation in research angles, methods, and populations in each of the studies within this dissertation is a strength that comes with a disadvantage. While the variety in studies offers a rich picture of the roles that gender and gender stereotypes play within secondary education, the use of different angles, methods, and populations limits the extent to which the findings can be synthesized. For instance, it would have been valuable if the students from the individual schools included in the macro-level study also participated in the micro-level studies, so that these studies could have been linked in the empirical analyses. Hence, more studies with data at multiple levels are needed to be able to connect the dots.

Further, this research solely focused on gender. Although the focus on gender alone is relevant, the contextualization and understanding of effects of gender in education might have benefitted from the inclusion of other student and teacher characteristics. For instance, characteristics such as migration background and socioeconomic status in combination

with students' and teachers' gender (i.e., intersectionality) may play an important role in processes of representation (Meier, 2019) and discrimination (Rieggle-Crumb & Humphries, 2012). Intersectionality refers to characteristics such as gender, migration background, and socioeconomic status jointly shaping individuals' experiences and reinforcing or neutralizing one another (Collins et al., 2021; Shields, 2008). Future research should consider the role of intersectionality to further contextualize and understand effects of gender in education.

Lastly, this research only used quantitative research methods and thereby did not make use of the benefits of qualitative research methods. Qualitative research methods can provide other types of insights that may have contributed to the understanding of the role of gender and gender stereotypes in education. Following the implications of this dissertation, it may be valuable to use qualitative research methods from other scientific disciplines that enable the study of the effects of gender and gender stereotypes due to implicit psychological and social processes. Hence, future research should consider ethnographic research methods, such as participant observation. Participants observation may provide new insights on how gender and gender stereotypes unconsciously and unintentionally shape the interaction between students and their teachers and how these processes can translate into student performance.

6.6 CONCLUSION

In all, this dissertation brought together literature from different scientific disciplines and used a variety of methods and research angles to theorize about and empirically test the role of gender and gender stereotypes in student performance. This dissertation showed that both gender and gender stereotypes have an impact on student performance but stresses the importance of context. On the macro level, students' and teachers' gender were found to be related to students' performance in subjects in which students' performance is negatively gender stereotyped (i.e., women in STEM and men in languages). However, when zooming in on educational contexts, the relations between student and teacher gender and student performance differed for men and women across subjects in the different contexts. Further, teachers' gender stereotypes, and not students' gender stereotypes, were found to play an explanatory role in effects of teachers' gender on student performance through gender bias in expectations and grading, but only in the subject math. Hence, follow-up questions about the contexts in which gender and gender stereotypes affect student performance were raised resulting in avenues for future research on the theorization and examination of the role of student and teacher gender and gender stereotypes in education.

The research within this dissertation has implications for education practice. First, this dissertation showed that student and teacher gender are relevant for student performance. The findings within this dissertation indicate that improving gender representation in STEM and language subjects in Dutch secondary education may be desirable as gender

representation benefits the performance of the groups of students that are confronted with negative gender stereotypes in STEM and language subjects. For this reason, schools should try to attract and maintain female STEM teachers and male language teachers.

Second, this dissertation showed that gender stereotypes still exist in students and teachers and that implicit gender stereotypes can translate into student performance through students' beliefs of self and teachers' beliefs of others. So, improving students' and teachers' awareness of implicit gender stereotypes may contribute to reducing negative effects of gender stereotypes on beliefs of self and others and student performance.

However, this dissertation also showed that the role of gender and gender stereotypes can differ for male and female students across subjects and in different educational contexts. Thus, improving gender representation and awareness of implicit gender stereotypes does not necessarily benefit student performance in every educational context (i.e., every group of students in every school). Hence, schools should explore if gender and gender stereotypes may impact beliefs of self, beliefs of others, grading, and student performance in their school, and for which groups of students and teachers gender and gender stereotypes are relevant. This requires schools to gain knowledge on the possible role of gender in education. The studies within this dissertation can help to recognize when, where, and how gender and gender stereotypes affect the education process. Further, this dissertation laid the foundation for the development of an education tool intended to create awareness about teachers' own gender bias in expectations and grading. The development of this tool is one of the projects that I will be working on after this research. In doing so, I aim to contribute to shedding a light on the hidden effects of gender and gender stereotypes in education so that gendered beliefs or lack of representation no longer hinder students to see their capabilities or, in other words, *butterflies to see themselves*.





References

A

- Agyapong, E. (2018). Representative bureaucracy: Examining the effects of female teachers on girls' education in Ghana. *International Journal of Public Administration*, 14(16), 1338-1350.
- Akram, S. (2018). Representative bureaucracy and unconscious bias: Exploring the unconscious dimension of active representation. *Public Administration*, 96(1), 119-133. <https://doi.org/10.1111/padm.12376>.
- Alan, S., Ertac, S., & Mumcu, I. (2018). Gender stereotypes in the classroom and effects on achievement. *The Review of Economics and Statistics*, 100(5), 876-890. https://doi.org/10.1162/rest_a_00756.
- An, S. H., Song, M., & Meier, K. J. (2021). Representative bureaucracy and the policy environment: Gender representation in forty-four countries. *Public Administration*, October. <https://doi.org/10.1111/padm.12789>.
- Arens, A.K., Marsh, W.M., Pekrun, R., Lichtenfeld, S., Murayama, K. & vom Hofe, R. (2017). Math selfconcept, grades, and achievement test scores: Longterm reciprocal effects across five waves and three achievement tracks. *Journal of Educational Psychology*, 109(5), 623-634.
- Annot, M., Gray, J., James, M., Rudduck, J., & Duveen, G. (1998). Recent research on gender and educational performance. London: Ofsted.
- Aslam, M., & Kingdon, G. (2011). What can teachers do to raise pupil achievement? *Economics of Education Review*, 30(3), 559–574. <https://doi.org/10.1016/j.econedurev.2011.01.001>.
- Autin, F. & Croizet, J. (2012). Improving memory efficiency by reframing metacognitive interpretation of task difficulty. *Journal of Experimental Psychology*, 141(4), 610-618.

B

- Banaji, M. R., & Greenwald, A. G. (1995). Implicit gender stereotyping in judgments of fame. *Journal of Personality and Social Psychology*, 68(2), 181–198. <https://doi.org/10.1037/0022-3514.68.2.181>.
- Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1). <https://doi.org/10.18637/jss.v067.i01>.
- Bardbury, M. & Kellough, J.E. (2011). Representative bureaucracy: Assessing the evidence on active representation. *The American Review of Public Administration*, 41(2), 157-167.
- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, 5(4), 323–370. <https://doi.org/10.1037/1089-2680.5.4.323>.

- Beilock, S. L., Gunderson, E. A., Ramirez, G., & Levine, S. C. (2010). Female teachers' math anxiety affects girls' math achievement. *Proceedings of the National Academy of Sciences of the United States of America*, 107(5), 1860–1863. <https://doi.org/10.1073/pnas.0910967107>.
- Bem, S. L. (1981). Gender schema theory: A cognitive account of sex typing. *Psychological Review*, 88, 354–364.
- Berg, P., Palmgren, O., & Tyrefors, B. (2020). Gender grading bias in junior high school mathematics. *Applied Economics Letters*, 27(11), 915–919. <https://doi.org/10.1080/13504851.2019.1646862>.
- Biesta, G., Priestley, M., & Robinson, S. (2015). The role of beliefs in teacher agency. *Teachers and teaching*, 21(6), 624–640. <https://doi.org/10.1080/13540602.2015.1044325>
- Blakemore, J.E.O., Berenbaum, S.A. & Liben, L.S. (2009). *Gender Development*. Londen: Psychology Press.
- Blakemore, J.E.O & Hill, C.A. (2008). The child gender socialization scale: A measure to compare traditional and feminist parents, *Sex Roles*, 58, 192–207.
- Blanton, H., Jaccard, J., Klick, J., Mellers, B., Mitchell, G., & Tetlock, P. E. (2009). Strong claims and weak evidence: Reassessing the predictive validity of the IAT. *Journal of Applied Psychology*, 94(3), 567–582. <https://doi.org/10.1037/a0014665>.
- Breda, T., Jouini, E., Napp, C., & Thebault, G. (2020). Gender stereotypes can explain the gender-equality paradox. *Proceedings of the National Academy of Sciences*, 117(49), 31063–31069. <https://doi.org/10.1073/pnas.2008704117>
- Breda, T., & Ly, S. T. (2015). Professors in core science fields are biased in favor of women: evidence from France. *American Economic Journal: Applied Economics*, 7(4), 53–75. <http://www.parisschoolofeconomics.eu/docs/ly-son-thierry/gendergapulm.pdf>.
- Brown, C. S., & Stone, E. A. (2016). Gender stereotypes and discrimination. How sexism impacts development. *Advances in Child Development and Behavior*, 50, 105–133. <https://doi.org/10.1016/bs.acdb.2015.11.001>.
- Bügel, K., Alberts, R.V.J., & Zwitser, R. J. (2011). Hebben jongens een steeds groter wordende onderwijs achterstand? Sekseverschillen in onderwijsprestaties vanaf vijftien jaar. *Tijdschrift Voor Genderstudies*, 4, 63–76. <https://ugp.rug.nl/genderstudies/article/view/1447/1441>.
- Burgess, D., & Borgida, E. (1999). Who women are, who women should be: Descriptive and prescriptive gender stereotyping in sex discrimination. *Psychology, Public Policy, and Law*, 5(3), 665–692. <https://doi.org/10.1037/1076-8971.5.3.665>.
- Burgess, S., & Greaves, E. (2013). Test scores, subjective assessment, and stereotyping of ethnic minorities. *Journal of Labor Economics*, 31(3), 535–576. <https://doi.org/10.1086/669340>.
- Bussey, K., & Bandura, A. (1999). Social cognitive theory of gender development and differentiation. *Psychological Review*, 106(4), 676–713.

C

- Caldwell, E., & Hartnett, R. (1967). Sex bias in college grading? *Journal of Educational Measurement*, 4(3), 129–132.
- Cardona López, J.A., Nordfjell, O.B., Gaini, F., Heikkinen, M. (2018). Promising Nordic practices in gender equality promotion: Developing teacher education dialogue, practice, and policy cycles on-line. *Policy Futures in Education*, 16(5), 605-619. <https://doi.org/10.1177/1478210317722286>
- Carnes, M., Devine, P. G., Baier Manwell, L., Byars-Winston, A., Fine, E., Ford, C. E., Forscher, P., Isaac, C., Kaatz, A., Magua, W., Palta, M., & Sheridan, J. (2015). The effect of an intervention to break the gender bias habit for faculty at one institution: A cluster randomized, controlled trial. *Academic medicine journal of the Association of American Medical Colleges*, 90(2), 221–230. <https://doi.org/10.1097/ACM.0000000000000552>
- CBS. (2022). Stedelijkheid. Consulted from. <https://www.cbs.nl/nl-nl/nieuws/2019/44/meeste-afval-per-inwoner-in-minst-stedelijke-gemeenten/stedelijkheid>.
- CBS (2020). Hoeveel mensen met een migratieachtergrond wonen in Nederland? Consulted from: <https://www.cbs.nl/nl-nl/dossier/dossier-asiel-migratie-en-integratie/hoeveel-mensen-met-een-migratieachtergrond-wonen-in-nederland-> .
- Charlesworth, T. E. S., & Banaji, M. R. (2021). Patterns of implicit and explicit stereotypes III: Long-term change in gender stereotypes. *Social Psychological and Personality Science*, 13(1), 14-26. <https://doi.org/10.1177/1948550620988425>.
- Chamorro-Premuzic, T., Harlaar, N., Greven, C.U. & Plomin, R. (2010). More than just IQ: A longitudinal examination of self-perceived abilities as predictors for academic performance in a large sample of UK twins. *Intelligence*, 38(4), 385-392.
- Cho, I. (2012). The effect of teacher-student gender matching: Evidence from OECD countries. *Economics of Education Review*, 31(3), 54–67. <https://doi.org/10.1016/j.econedurev.2012.02.002>.
- Cole, M. S., Feild, H. S., & Giles, W. F. (2004). Interaction of recruiter and applicant gender in resume evaluation: A field study. *Sex Roles*, 51(9–10), 597–608. <https://doi.org/10.1007/s11199-004-5469-1>.
- Collins, P.H., Da Silva, E.C.G., Ergun, E., Furseth, I., Bond, K.D., & Martínez-Palacios, J. (2021). Intersectionality as critical social theory. *Contemporary Political Theory*, 20, 690–725. <https://doi.org/10.1057/s41296-021-00490-0>.
- Cornwell, C., Mustard, D. B., & Van Parys, J. (2013). Noncognitive skills and the gender disparities in test scores and teacher assessments: Evidence from primary school. *Journal of Human Resources*, 48(1), 236–264. <https://doi.org/10.3368/jhr.48.1.236>.
- Crisp, R. J., Bache, L. M., & Maitner, A. T. (2009). Dynamics of social comparison in counter-stereotypic domains: Stereotype boost, not stereotype threat, for women engineering majors. *Social Influence*, 4(3), 171–184. <https://doi.org/10.1080/15534510802607953>.

D

- Daalmans, S., Kleemans, M. & Sadza, A. (2017). Gender representation on gender-targeted television channels: A comparison of female- and male-targeted TV Channels in the Netherlands. *Sex Roles*, 77, 366-378.
- Dasgupta, N. (2004). Implicit ingroup favoritism, outgroup favoritism, and their behavioral manifestations. *Social Justice Research*, 17(2), 143–169. <https://doi.org/10.1023/B:SORE.0000027407.70241.15>.
- De Boer, E., Janssen, F.J.J.M, Dam, M., Van Driel, J.H. (2019). Ontwikkeling van agency van docenten in opleiding: Een ecologisch perspectief. *Pedagogische Studiën*, 96(6), 354-377.
- Dee, T. S. (2005). A teacher like me: Does race, ethnicity, or gender matter? *The American Economic Review*, 95(2), 158-165. <https://www.jstor.org/stable/4132809>
- Denessen, E., Hornstra, L., Van den Bergh, L., & Bijlstra, G. (2022). Implicit measures of teachers' attitudes and stereotypes, and their effects on teacher practice and student outcomes: A review. *Learning and Instruction*, 78, 101437. <https://doi.org/10.1016/j.learninstruc.2020.101437>.
- De Vries, E. E., van der Pol, L. D., Toshkov, D. D., Groeneveld, M. G., & Mesman, J. (2022). Fathers, faith, and family gender messages: Are religiosity and gender talk related to children's gender attitudes and preferences? *Early Childhood Research Quarterly*, 59, 21–31. <https://doi.org/10.1016/j.ecresq.2021.10.002>.
- Deutsch, F. M., & Saxon, S. E. (1998). The double standard of praise and criticism for mothers and fathers. *Psychology of Women Quarterly*, 22(4), 665–683. <https://doi.org/10.1111/j.1471-6402.1998.tb00184.x>.
- Devine, P.G., Forscher, P.S., Cox, W.T.L., Kaatz, A., Sheridan, J., & Carnes, M. (2017). A gender bias habit-breaking intervention led to increased hiring of female faculty in STEMM departments. *Journal of Experimental Social Psychology*, 73, 211-215. <https://doi.org/10.1016/j.jesp.2017.07.002>
- Dhillon, A., & Meier, K. J. (2022). Representative bureaucracy in challenging environments: Gender representation, education, and India. *International Public Management Journal*, 25(1), 43–64. <https://doi.org/10.1080/10967494.2020.1802633>.
- Ding, F., Lu, J., & Riccucci, N.M. (2021). How bureaucratic representation affects public organizational performance: A meta-analysis. *Public Administration Review*, 81(6), 1003-1018. <https://doi.org/10.1111/puar.13361>.
- Disatnik, D. & Sivan, L. (2016). Multicollinearity illusion in moderated regression analysis. *Marketing Letters*, 27, 403-408. DOI 10.1007/s11002-014-9339-5.
- Doornkamp, L., Groeneveld, S., Groeneveld, M. G., Van der Pol, L. D., & Mesman, J. (2021). Understanding the symbolic effects of gender representation: A multi-source study in education. *International Public Management Journal*. <https://doi.org/10.1080/10967494.2021.1991534>.

- Doornkamp, L. Van den Bekerom, P., & Groeneveld, S. (2019). The individual level effect of symbolic representation: An experimental study on teacher-student gender congruence and students' perceived abilities in math. *Journal of Behavioral Public Administration*, 2(2), 1-11. <https://doi.org/10.30636/jbpa.22.64>.
- Doornkamp, L., Van der Pol, L.D., Groeneveld, S., Mesman, J., Endendijk, J.J., & Groeneveld, M.G. (2022). Understanding gender bias in teachers' grading: The role of gender stereotypical beliefs. *Teaching and Teacher Education*, 118, 1-14. <https://doi.org/10.1016/j.tate.2022.103826>.
- DUO (2020). 03 Lesgevend personeel voortgezet onderwijs (in personen). Consulted from: https://duo.nl/open_onderwijsdata/databestanden/vo/onderwijspersoneel/vo-personeel3.jsp.
- Dutch Report Education Inspectorate (2019). De staat van het onderwijs 2019. Utrecht: Inspectie van onderwijs.

E

- Eagly, H. A., Wood, W. & Diekmann, A. B. (2000). Social role theory of sex differences and similarities: A current appraisal. In T. Eckes and H. M. Trautner (Eds.), *The developmental social psychology of gender* (pp. 123-174). Mahwah, NJ: Lawrence Erlbaum Associates.
- Ellemers, N. (2018). Gender stereotypes. *Annual Review of Psychology*, 69, 275–298. <https://doi.org/10.4324/9781003041870-5>.
- Endendijk, J. J., Groeneveld, M. G., van Berckel, S. R., Hallers-Haalboom, E. T., Mesman, J., & Bakermans-Kranenburg, M. J. (2013). Gender stereotypes in the family context: Mothers, fathers, and siblings. *Sex Roles*, 68(9–10), 577–590. <https://doi.org/10.1007/s11199-013-0265-4>.
- Endendijk, J. J., Van Baar, A. L., & Deković, M. (2020). He is a stud, she is a slut! A meta-analysis on the continued existence of sexual double standards. *Personality and Social Psychology Review*, 24(2), 163-190.
- Espinoza, P., Arêas da Luz Fontes, A. B., & Arms-Chavez, C. J. (2014). Attributional gender bias: Teachers' ability and effort explanations for students' math performance. *Social Psychology of Education*, 17(1), 105–126. <https://doi.org/10.1007/s11218-013-9226-6>.
- Evans, D.K, Akmal, M. and Jakiela, P. (2020). Gender gaps in education: The long view. Center for Global Development Working Paper 523, consulted from: <https://www.edu-links.org/sites/default/files/media/file/gender-gaps-education-long-view.pdf>.
- Education in Numbers (2020). Opleidingsniveau van de bevolking, consulted from: <https://www.onderwijsin cijfers.nl/kengetallen/internationaal/opleidingsniveau-bevolking>.

F

- Falch, T., & Naper, L. R. (2013). Educational evaluation schemes and gender gaps in student achievement. *Economics of Education Review*, 36, 12–25. <https://doi.org/10.1016/j.econedurev.2013.05.002>.
- Fay, D.L., Hicklin Fryar, A., Meier, K.J., & Wilkins, V. (2020). Intersectionality and equity: Dynamic bureaucratic representation in higher education. *Public Administration*, 1-18.
- Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology*, 54, 297–327. <https://doi.org/10.1146/annurev.psych.54.101601.145225>.
- Fiske, S. T. (2017). Prejudices in cultural contexts: Shared stereotypes (gender, age) versus variable stereotypes (race, ethnicity, religion). *Perspectives on Psychological Science*, 12(5), 791–799. <https://doi.org/10.1177/1745691617708204>.
- Forscher, P.S., Mitamura, C., Dix, E.L., Cox, W.T.L., & Devine, P.G. (2017). Breaking the prejudice habit: Mechanisms, timecourse, and longevity. *Journal of Experimental Social Psychology*, 72, 133-146. <https://doi.org/10.1016/j.jesp.2017.04.009>
- Fredriksen, K., & Rhodes, J. (2004). The role of teacher relationships in the lives of students. *New directions for youth development*, 2004(103), 45-54.

G

- Gade, D.M. & Wilkins, V.M. (2013). Where did you serve? Veteran identity, representative bureaucracy, and vocational rehabilitation. *Journal of Public Administration Research and Theory*, 23(2), 267-288.
- Gansen, H. M. (2019). Push-ups versus clean-up: Preschool teachers' gendered beliefs, expectations for behavior, and disciplinary practices. *Sex Roles*, 80, 393-408. <https://doi.org/10.1007/s11199-018-0944-2>.
- Garcia-Retamero, R., Müller, S. M., & López-Zafra, E. (2011). The malleability of gender stereotypes: Influence of population size on perceptions of men and women in the past, present, and future. *Journal of Social Psychology*, 151(5), 635–656. <https://doi.org/10.1080/00224545.2010.522616>.
- Gibbons, S., & Chevalier, A. (2008). Assessment and age 16+ education participation. *Research Papers in Education*, 23(2), 113–123. <https://doi.org/10.1080/02671520802048638>.
- Girod, S. , Fassiotto, M. , Grewal, D. , Ku, M. , Sriram, N. , Nosek, B. & Valentine, H. (2016). Reducing implicit gender leadership bias in academic medicine with an educational intervention. *Academic Medicine*, 91 (8), 1143-1150. doi: 10.1097/ACM.0000000000001099.
- Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, 85(2), 197–216. <https://doi.org/10.1037/0022-3514.85.2.197>.

- Greven, C.U., Harlaar, N., Kovas, Y., Chamorro-Premuzic, T. & Plomin, R. (2009). More than just IQ: School achievement is predicted by self-perceived abilities – But for genetic rather than environmental reasons. *Psychological Science*, 20(6), 753-762.
- Grissom, J. A., Kern, E. C., & Rodriguez, L. A. (2015). The “representative bureaucracy” in education: Educator workforce diversity, policy outputs, and outcomes for disadvantaged students. *Educational Researcher*, 44(3), 185–192. <https://doi.org/10.3102/0013189X15580102>.
- Groeneveld, S., Andrews, R., Meier, K.J., & Schröter, E. (2015). Representative bureaucracy and public service performance: Where, why and how does representativeness work? <http://dx.doi.org/10.2139/ssrn.3982606>.
- Groeneveld, S. & Meier, K.J. (2022). Theorizing status distance: Rethinking the micro theories of representation and diversity in public organizations. *Administration & Society*, 54(2), 248-276. <https://doi.org/10.1177/00953997211028825>.
- Groeneveld, S. & Van de Walle, S. (2010). A contingency approach to representative bureaucracy: Power, equal opportunities and diversity. *International Review of Administrative Sciences*, 76(2), 239-258.
- Groeneveld, M. G., Van der Pol, L. D., De Vries, E. E., & Mesman, J. (2021). Triadic family conversations about gender: Children as driving forces and messiness in messages they receive. *Journal of Family Psychology*. <https://doi.org/10.1037/fam0000931>.
- Guimond, S., & Roussel, L. (2000). Bragging About One's School Grades: Gender Stereotyping and Students' Perception of Their Abilities in Science, Mathematics, and Language. *Social Psychology of Education*, 4(3-4), 275-293. <https://doi.org/10.1023/A:1011332704215>
- Guul, T.S. (2018). The individual-level effect of gender matching in representative bureaucracy. *Public Administration Review*, 78(3), 398-408.

H

- Haines, E. L., Deaux, K., & Lofaro, N. (2016). The times they are a-changing ... or are they not? A comparison of gender stereotypes, 1983–2014. *Psychology of Women Quarterly*, 40(3), 353–363. <https://doi.org/10.1177/0361684316634081>.
- Hanna, R. N., & Linden, L. L. (2012). Discrimination in grading. *American Economic Journal: Economic Policy*, 4(4), 146–168. <https://doi.org/10.1257/pol.4.4.146>.
- Hausmann, M. (2014). Arts versus science – Academic background implicitly activated gender stereotypes on cognitive abilities with threat raising men's (but lowering women's) performance, *Intelligence*, 46, 235-245.
- Hayes, A.F. (2018). Introduction to mediation, moderation, and conditional process analysis: A regression-based approach. London: The Guilford Press.

- Heller, K., Finsterwald, M., & Ziegler, A. (2001). Implicit theories of German mathematics and physics teachers on gender specific giftedness and motivation. *Psychologische Beiträge*, 43(1), 172.
- Herrmann, S.D., Adelman, R.M., Bodford, J.E., Graudejus, O., Okun, M.A., & Kwan, V.S.Y. (2016). The effects of a female role model on academic performance and persistence of women in STEM courses. *Basic and Applied Social Psychology*, 38(5), 258-268. <https://dx.doi.org/10.1080/01973533.2016.1209757>.
- Heyder, A., & Kessels, U. (2017). Boys don't work? On the psychological benefits of showing low effort in high school. *Sex Roles*, 77, 72-85. <https://doi.org/10.1007/s11199-016-0683-1>.
- Hilliard, L. J., & Liben, L. S. (2010). Differing levels of gender salience in preschool classrooms: Effects on children's gender attitudes and intergroup bias. *Child Development*, 81(6), 1787-1798. <https://doi.org/10.1111/j.1467-8624.2010.01510.x>.
- Hinnerich, B. T., Högl, E., & Johannesson, M. (2011). Are boys discriminated in Swedish high schools? *Economics of Education Review*, 30(4), 682-690. <https://doi.org/10.1016/j.econedurev.2011.02.007>.
- Holmlund, H., & Sund, K. (2008). Is the gender gap in school performance affected by the sex of the teacher? *Labour Economics*, 15(1), 37-53. <https://doi.org/10.1016/j.labeco.2006.12.002>.
- Hyde, J. S., Bigler, R. S., Joel, D., Tate, C. C., & Van Anders, S. M. (2019). The future of sex and gender in psychology: Five challenges to the gender binary. *American Psychologist*, 74(2), 171-193. <https://doi.org/10.1037/amp0000307>.

I

- Istenič, M. Č. (2007). Attitudes towards gender roles and gender role behaviour among urban, rural, and farm populations in Slovenia. *Journal of Comparative Family Studies*, 38(3), 477-496. <http://www.jstor.org/stable/41604170>.

J

- Jeynes, W. H. (2002). Why religious schools positively impact the academic achievement of children. *International Journal of Education & Religion*, 3, 16-32. <http://dx.doi.org/10.1163/157006202760182418>.
- Jones, S., & Myhill, D. (2004). "Troublesome boys" and "compliant girls": Gender identity and perceptions of achievement and underachievement. *British Journal of Sociology of Education*, 25(5). <https://doi.org/10.1080/0142569042000252044>.

Jungbluth, P. (1982). Docenten over onderwijs aan meisjes. Positieve diskriminatie met een dubbele bodem [Teachers on girls' education. Positive discrimination with a double bottom] (dissertation). ITS, Nijmegen.

K

- Kapitanoff S. & Pandey C. (2017). Stereotype threat, anxiety, instructor gender, and underperformance in women. *Active Learning in Higher Education*, 18(3), 213-229. <https://doi.org/10.1177/1469787417715202>.
- Keiser, L. R., Wilkins, V. M., Meier, K. J., & Holland, C. A. (2002). Lipstick and logarithms: Gender, institutional context, and representative bureaucracy. *The American Political Science Review*, 96(3), 553–564. <https://doi.org/10.1017/S0003055402000321>.
- Kennedy, B. (2014). Unraveling representative bureaucracy: A systematic analysis of the literature. *Administration & Society*, 46(4), 395-421.
- Khalilzadeh, J., & Tasci, A. D. A. (2017). Large sample size, significance level, and the effect size: Solutions to perils of using big data for academic research. *Tourism Management*, 62, 89–96. <https://doi.org/10.1016/j.tourman.2017.03.026>.
- Kiefer, A. K., & Sekaquaptewa, D. (2007). Implicit stereotypes and women's math performance: How implicit gender-math stereotypes influence women's susceptibility to stereotype threat. *Journal of Experimental Social Psychology*, 43(5), 825–832. <https://doi.org/10.1016/j.jesp.2006.08.004>.
- Kingsley, J. D. (1944). Representative bureaucracy. Yellow Springs, OH: Antioch Press.

L

- Lavy, V. (2008). Do gender stereotypes reduce girls' or boys' human capital outcomes? Evidence from a natural experiment. *Journal of Public Economics*, 92(10–11), 2083–2105. <https://doi.org/10.1016/j.jpubeco.2008.02.009>.
- Lavy, V., & Sand, E. (2018). On the origins of gender gaps in human capital: Short- and long-term consequences of teachers' biases. *Journal of Public Economics*, 167, 263–279. <https://doi.org/10.1016/j.jpubeco.2018.09.007>.
- Lee, J., Rhee, D. E., & Rudolf, R. (2019). Teacher gender, student gender, and primary school achievement: Evidence from ten francophone african countries. *Journal of Development Studies*, 55(4), 661–679. <https://doi.org/10.1080/00220388.2018.1453604>.
- Lenton, A. P., Bruder, M., & Sedikides, C. (2009). A meta-analysis on the malleability of automatic gender stereotypes. *Psychology of Women Quarterly*, 33(2), 183–196. <https://doi.org/10.1111/j.1471-6402.2009.01488.x>.

- Liben, L. S., & Bigler, R. S. (2002). The developmental course of gender differentiation: Conceptualizing, measuring, and evaluating constructs and pathways. *Monographs of the Society for Research in Child Development*, 67(2), i-viii+1-183. <https://www.jstor.org/stable/3181530>.
- Lindahl, E. (2016). Are teacher assessments biased? Evidence from Sweden. *Education Economics*, 24(2), 224–238. <https://doi.org/10.1080/09645292.2015.1014882>.
- Lyness, K. S., & Heilman, M. E. (2006). When fit is fundamental: Performance evaluations and promotions of upper-level female and male managers. *Journal of Applied Psychology*, 91(4), 777–785. <https://doi.org/10.1037/0021-9010.91.4.777>.

M

- Maries, A., Karim, N.I., & Singh, C. (2018). Is agreeing with a gender stereotype correlated with the performance of female students in introductory physics? *Physical Review Physics Education Research*, 14(2), 1-10. DOI: 10.1103/PhysRevPhysEducRes.14.020119.
- Maries, A., Karim, N.I., & Singh, C. (2020). Active learning in an inequitable learning environment can increase the gender performance gap: The negative impact of stereotype threat. *The Physics Teacher*, 58, 430-433. <https://doi.org/10.1119/10.0001844>.
- Marsh, H.W., Byrne, B.M. & Shavelson, R.J. (1988). A multifaceted academic self-concept: Its hierarchical structure and its relation to academic achievement. *Journal of Educational Psychology*, 80(3), 366-380.
- Marsh, H.W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects models of causal ordering. *Child Development*, 76(2), 397–416. <https://doi.org/10.1111/j.1467-8624.2005.00853.x>.
- Martin, C. L., & Dinella, L. M. (2012). Congruence between gender stereotypes and activity preference in self-identified tomboys and non-tomboys. *Archives of Sexual Behavior*, 41(3), 599–610. <https://doi.org/10.1007/s10508-011-9786-5>.
- Marx, D. M., & Roman, J. S. (2002). Female role models: Protecting women's math test performance. *Personality and Social Psychology Bulletin*, 28(9), 1183–1193. <https://doi.org/10.1177/01461672022812004>.
- Matějů, P., & Smith, M. L. (2015). Are boys that bad? Gender gaps in measured skills, grades and aspirations in Czech elementary schools. *British Journal of Sociology of Education*, 36(6), 871–895. <https://doi.org/10.1080/01425692.2013.874278>.
- McFarland, L., Murray, E. & Phillipson, S. (2016). Student-teacher relationships and student self-concept: Relations with teacher and student gender. *Australian Journal of Education* 60(1), 5-25.
- Meier, K.J. (2019). Theoretical frontiers in representative bureaucracy: New directions for research. *Perspectives on Public Management and Governance*, 2(1), 39-56.

- Meier, K.J. & Bohte, J. (2001). Structure and discretion: Missing links in representative bureaucracy. *Journal of Public Administration Research and Theory*, 11(4), 455-470.
- Meier, K.J., Doerfler, C., Hawes, D., Hicklin, A.K. & Rocha, R.R. (2006). The role of management and representation in improving performance of disadvantaged students: An application of Bum Phillips's "don shula rule". *Review of Policy Research*, 23(5), 1095-1110.
- Meier, K.J., Pennington, M.S. & Eller, W.S. (2005). Race, sex and Clarendo Thomas: Representation change in the EEOC. *Public Administration Review*, 65(2), 171-179.
- Meier, K.J. & Stewart, J. (1992). The impact of representative bureaucracies: Educational systems and public policies. *American Review of Public Administration*, 22(3), 157-171.
- Meier, K.J., Wrinkle, R.D. & Polinard, J.L. (1999). Representative bureaucracy and distributional equity: Addressing the hard question. *The Journal of Politics*, 61(4), 1025-1039.
- Muntoni, F., & Retelsdorf, J. (2018). Gender-specific teacher expectations in reading—The role of teachers' gender stereotypes. *Contemporary Educational Psychology*, 54, 212–220. <https://doi.org/10.1016/j.cedpsych.2018.06.012>.
- Muralidharan, K., & Sheth, K. (2016). Bridging education gender gaps in developing countries: The role of female teachers. *Journal of Human Resources*, 51(2), 269–297. doi: 10.3368/jhr.51.2.0813-5901R1.

N

- Neugebauer, M., Helbig, M., & Landmann, A. (2011). Unmasking the myth of the same-sex teacher advantage. *European Sociological Review*, 27(5), 669–689. <https://doi.org/10.1093/esr/jcq038>.
- Nicholson-Crotty, S., Grissom, J.A., Nicholson-Crotty, J. & Redding, C. (2016). Disentangling the causal mechanisms of representative bureaucracy: Evidence from assignment of students to gifted programs. *Journal of Public Administration Research and Theory*, 26(4), 745-757.
- Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002a). Harvesting implicit group attitudes and beliefs from a demonstration website. *Group Dynamics*, 6(1), 101–115. <https://doi.org/10.1037/1089-2699.6.1.101>.
- Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002b). Math = male, me = female, therefore math ≠ me. *Journal of Personality and Social Psychology*, 83(1), 44–59. <https://doi.org/10.1037/0022-3514.83.1.44>.
- Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the implicit association test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin*, 31(2), 166–80. <https://doi.org/10.1177/0146167204271418>.

- Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., et al. (2009). National differences in gender-science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences*, 106 (26), 10593–10597. <https://doi.org/10.1073/pnas.0809921106>
- Nürnberg, M., Nerb, J., Schmitz, F., Keller, J., & Sütterlin, S. (2016). Implicit gender stereotypes and essentialist beliefs predict preservice teachers tracking recommendations. *Journal of Experimental Education*, 84(1), 152–174. <https://doi.org/10.1080/00220973.2015.1027807>.

O

- Obioma, I.F., Hentschel, T., & Hernandez Bark, A.S. (2021). Gender stereotypes and self-characterizations in Germany and Nigeria: A cross-cultural comparison. *Journal of Applied Social Psychology*, 00, 1-17. <https://doi.org/10.1111/jasp.12801>
- OESO (2019). Part-time and partly equal: gender and work in the Netherlands. Parijs: OECD Publishing.
- Onderwijsraad (2019). Een verkenning van sekse verschillen in het onderwijs, consulted from: <https://www.onderwijsraad.nl/publicaties/adviezen/2020/10/07/verkenning-sekse-verschillen-onderwijs>.

P

- Pajares, F. & Miller, D. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis. *Journal of Educational Psychology*, 86(2), 193-203.
- Palan, K. M. (2001). Gender identity in consumer behavior research: A literature review and research agenda. *Academy of Marketing Science Review*, 2001(10), 1–24.
- Park, S. (2012). Does gender matter? The effect of gender representation of public bureaucracy on governmental performance. *American Review of Public Administration*, 43(2), 221-242.
- Peugh, J.L. (2010). A practical guide to multilevel modeling. *Journal of School Psychology*, 48, 85-112.
- Pitts, D.W. (2007). Representative bureaucracy, ethnicity and public schools: Examining the link between representation and performance. *Administration & Society*, 39(4), 497-526.
- Plante, I., De la Sablonnière, R., Aronson, J.M. & Théorêt, M. (2013). Gender stereotype endorsement and achievement-related outcomes: The role of competence beliefs and task values. *Contemporary Educational Psychology*, 38, 225-235.

- Powell, G. N. (1986). Applicants' qualifications on recruiters' evaluations. *Psychological Reports*, 58, 1003–1010. <https://doi.org/10.2466/pr0.1986.58.3.1003>.
- Prentice, D.A. & Carranza, E. (2002). What women and men should be, shouldn't be, are allowed to be, and don't have to be: The contents of prescriptive gender stereotypes. *Psychology of Women Quarterly*, 26, 269-281.
- Priestley, M., Biesta, G.J.J. & Robinson, S. (2015). Teacher agency: What is it and why does it matter? In R. Kneyber & J. Evers (eds.), *Flip the System: Changing Education from the Bottom Up*. London: Routledge.
- Protivínský, T., & Münich, D. (2018). Gender bias in teachers' grading: What is in the grade. *Studies in Educational Evaluation*, 59, 141–149. <https://doi.org/10.1016/j.stueduc.2018.07.006>.

R

- Riccucci, N. M., & Van Ryzin, G. G. (2017). Representative bureaucracy: A lever to enhance social equity, coproduction, and democracy. *Public Administration Review*, 77(1), 21–30. <https://doi.org/10.1111/puar.12649>.
- Riccucci, N.M., Van Ryzin, G.G. & Jackson, K. (2018). Representative bureaucracy, race and policing: A survey experiment. *Journal of Public Administration Research and Theory*, 28(4), 506-518.
- Riccucci, N.M., Van Ryzin, G.G. & Lavena, C.F. (2014). Representative bureaucracy in policing: Does it increase perceived legitimacy? *Journal of Public Administration Research and Theory*, 24(3), 537-551.
- Riegle-Crumb, C., & Humphries, M. (2012). Exploring bias in math teachers' perceptions of students' ability by gender and race/ethnicity. *Gender and Society*, 26(2), 290-322. <https://doi.org/10.1177/0891243211434614>.
- Roch, C.H., Pitts, D.W. & Navarro, I. (2010). Representative bureaucracy and policy tools: Ethnicity, student discipline, and representation in public schools. *Administration & Society*, 42(1), 38-65.
- Rudman, L. A., & Kilianski, S. E. (2000). Implicit and explicit attitudes toward female authority. *Personality and Social Psychology Bulletin*, 26(11), 1315–1328. <https://doi.org/10.1177/0146167200263001>.

S

- Sagebin Bordini, G., & Sperb, T. M. (2013). Sexual double standard: A review of the literature between 2001 and 2010. *Sexuality and Culture*, 17(4), 686–704. <https://doi.org/10.1007/s12119-012-9163-0>.

- Saifi, S., & Mehmood, T. (2011). Effects of socioeconomic status on student achievement. *International Journal of Social Sciences and Education*, 1(2), 119–128.
- Schmenk, B. (2004). Language learning: A feminine domain? The role of stereotyping in constructing gendered learner identities. *TESOL Quarterly*, 38(3), 514–524. <https://doi.org/10.2307/3588352>.
- Schulze, E., & Tomal, A. (2005). Perceptions of gender competence: Are christian colleges different. *Growth*, 5(5). https://pillars.taylor.edu/acsd_growth/vol5/iss5/2.
- Schussler, D.L., Stooksberry, L.M., & Bercaw, L.A. (2010). Understanding teacher candidate dispositions: Reflecting to build self-awareness. *Journal of Teacher Education*, 61(4), 350–363. <https://doi.org/10.1177/0022487110371377>
- Shieh, G. (2010). On the misconception of multicollinearity in detection of moderating effects: Multicollinearity is not always detrimental. *Multivariate Behavioral Research*, 45(3), 483–507. <https://doi.org/10.1080/00273171.2010.483393>.
- Shields, S.A., (2008). Gender: An intersectionality perspective. *Sex Roles*, 59, 301–311. DOI 10.1007/s11199-008-9501-8.
- Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science*, 10(1), 80–83. <https://doi.org/10.1111/1467-9280.00111>.
- Siegle, D., & Reis, S. M. (1998). Gender differences in teacher and student perceptions of gifted students' ability and effort. *Gifted Child Quarterly*, 42(1), 39–47. <https://doi.org/10.1177/001698629804200105>.
- Smeding, A. (2012). Women in science, technology, engineering, and mathematics (STEM): An investigation of their implicit gender stereotypes and stereotypes' connectedness to math performance. *Sex Roles*, 67(11–12), 617–629. <https://doi.org/10.1007/s11199-012-0209-4>.
- Smetackova, I. (2015). Gender stereotypes, performance and identification with math. *Procedia- Social and Behavioral Science*, 190, 211–219.
- Solanki, S.M. & Xu, D. (2018). Looking beyond academic performance: The influence of instructor gender on student motivation in STEM fields. *American Educational Research Journal*, 55(4), 801–835. <https://doi.org/10.3102/0002831218759034>.
- Song, M. (2018). Gender representation and student performance: Representative bureaucracy goes to Korea. *American Review of Public Administration*, 48(4), 346–358. <https://doi.org/10.1177/0275074016676888>.
- Sowa, J.E., & Selden, S.C. (2003). Administrative discretion and active representation: An expansion of the theory on representative bureaucracy. *Public Administration Review*, 63(6), 700–710.
- Spencer, S. J., Logel, C., & Davies, P. G. (2016). Stereotype threat. *Annual Review of Psychology*, 67, 415–437. <https://doi.org/10.1146/annurev-psych-073115-103235>.

- Spinath, B., Spinath F.M., Harlaar, N. and Plomin, R. (2006). Predicting school achievement from general cognitive ability, self-perceived ability, and intrinsic value. *Intelligence*, 34, 363-374.
- Steele, J. (2003). Children's gender stereotypes about math: The role of stereotype stratification. *Journal of Applied Social Psychology*, 33(12), 2587-2606.
- Steffens, M.C. & Jelenec, P. (2011). Separating implicit gender stereotypes regarding math and language: Implicit ability stereotypes are self-serving for boys and men, but not for girls and woman. *Sex Roles*, 64(5), 324-335.
- Stobart, G., Elwood, J., & Quinlan, M. (1992). Gender bias in examinations: How equal are the opportunities? *British Educational Research*, 18(3), 261–276. <https://www.jstor.org/stable/1500831>
- Song, M. (2018). Gender representation and student performance: Representative bureaucracy goes to Korea. *The American Review of Public Administration*, 48(4), 346-358.
- Szymanowicz, A. & Furnham, A. (2011). Do intelligent women stay single? Cultural stereotypes concerning the intellectual abilities of men and women. *Journal of Gender Studies*, 20(1), 43-54.

T

- Tabachnick, B. G., & Fidell, L. S. (2012). Using multivariate statistics (6th ed.). Harper Collins.
- Ten Dam, G.T.M., Van Eck, E., Volman, M.L.L. (1997). Research Programmes on Gender and Education: results and conceptualisations. *European Journal of Education*, 32(4), 411-425. <https://www.jstor.org/stable/1503475>
- Ten Dam, G.T.M. & Volman, M.M. (1991). Conceptualising Gender Differences in Educational Research: The Case of the Netherlands. *British Journal of Sociology of Education*, 12(3), 309-321. <https://www.jstor.org/stable/1393040>
- Tenenbaum, H.R. & Leaper, C. (2002). Are parents' gender schema's related to their children's gender-related cognitions. *Developmental Psychology*, 38(4), 615-630.
- Theobald, N.A., & Haider-Markel, D.P. (2008). Race, bureaucracy, and symbolic representation: Interactions between citizens and police. *Journal of Public Administration Research and Theory*, 19(2), 409-426.
- Thompson, A. (2003). Caring in Context: Four Feminist Theories on Gender and Education, *Curriculum Inquiry*, 33(1), 9-65. <http://www.jstor.org/stable/3202137>
- TIMMS & PIRLS (2015). About TIMMS 2015. Retrieved from: <https://timssandpirls.bc.edu/timss2015/>.

- Trusty, J., Robinson, C. R., Plata, M., & Ng, K. (2000). Effects of gender, socioeconomic status, and early academic performance on postsecondary educational choice. *Journal of Counseling & Development*, 78, 87–96. <https://doi.org/10.1002/j.1556-6676.2000.tb01930.x>.
- Turner, S. L., Joeng, J. R., Sims, M. D., Dade, S. N., & Reid, M. F. (2019). SES, gender, and STEM career interests, goals, and actions: A test of SCCT. *Journal of Career Assessment*, 27(1), 134–150. <https://doi.org/10.1177/1069072717748665>.

V

- Van den Bergh, L., Denessen, E., Hornstra, L., Voeten, M., & Holland, R. W. (2010). The implicit prejudiced attitudes of teachers: Relations to teacher expectations and the ethnic achievement gap. *American Educational Research Journal*, 47(2), 497–527. <https://doi.org/10.3102/0002831209353594>.
- Vinopal, K. (2018). Understanding individual and organizational level representation: The case of parental involvement in schools. *Journal of Public Administration Research and Theory*, 28(1), 1-15.
- Volman, M.L.L., Ten Dam, G.T.M., Van Eck, E. (1995). Girls in science and technology: the development of a discourse, *Gender and Education*, 7(3), 283-292. <https://doi.org/10.1080/09540259550039004>

W

- Wach, F. S., Spengler, M., Gottschling, J., & Spinath, F. M. (2015). Sex differences in secondary school achievement - The contribution of self-perceived abilities and fear of failure. *Learning and Instruction*, 36, 104–112. <https://doi.org/10.1016/j.learninstruc.2015.01.005>.
- Walton, G.M. & Spencer, S.J. (2009). Latent ability: Grades and test scores systematically underestimate the intellectual ability of negatively stereotyped students. *Psychological Science*, 20(9), 1132-1139. <https://doi.org/10.1111/j.1467-9280.2009.02417>.
- Wang, Y. A., Sparks, J., Gonzales, J. E., Hess, Y. D., & Ledgerwood, A. (2017). Using independent covariates in experimental designs: Quantifying the tradeoff between power boost and Type I error inflation. *Journal of Experimental Social Psychology*, 72, 118-124.
- Weber, H.S., Lu, L., Shi, J. & Spinath, F.M. (2013). The roles of cognitive and motivational predictors in explaining school achievement in elementary school. *Learning and Individual Differences*, 25, 85-92.

- Wilkins, V.M. (2007). Exploring the causal story: Gender, active representation, and bureaucratic priorities. *Journal of Public Administration Research and Theory*, 17(1), 77-94.
- Wilkins, V.M. & Keiser, L.R. (2006). Linking passive and active representation by gender: The case of child support agencies. *Journal of Public Administration Research and Theory*, 16(1), 87-102.
- Winters, M.A., Haight, R.C., Swaim, T.T., & Pickering, K.A. (2013). The effect of same-gender teacher assignment on student achievement in the elementary and secondary grades: Evidence from panel data. *Economics of Education Review*, 34, 69-75. <http://dx.doi.org/10.1016/j.econedurev.2013.01.007>.
- Wulff, J. N., & Villadsen, A. R. (2019). Are survey experiments as valid as field experiments in management research? An empirical comparison using the case of ethnic employment discrimination. *European Management Review*, 17(1), 347-356. <https://doi.org/10.1111/emre.12342>.

X

- Xu, X., & Meier, K. J. (2021). Separating symbolic and active representation: A mixed methods study of gender and education in China. *Public Management Review*, 00(00), 1–23. <https://doi.org/10.1080/14719037.2021.1900352>.

Z

- Zhang, Y. (2019). Representative bureaucracy, gender congruence, and student performance in China. *International Public Management Journal*, 22(2), 321–342. <https://doi.org/10.1080/10967494.2018.1428702>.



DUTCH SUMMARY

Nederlandse samenvatting

De Nederlandse samenleving wordt gekenmerkt door segregatie op basis van gender (Daalmans et al., 2017). Zo werken mannen bijvoorbeeld vaker fulltime en zijn daardoor kostwinner, terwijl vrouwen vaker parttime werken en een groter aandeel hebben in de zorg voor kinderen (OESO, 2019). Ook de school- en beroepsloopbanen van mannen en vrouwen zijn gesegregeerd: mannen zijn oververtegenwoordigd in de bèta en techniek, terwijl vrouwen oververtegenwoordigd zijn in de zorg en in talen (Onderwijsraad, 2019). Deze segregatie op basis van gender bevestigt en versterkt genderstereotypen. Die genderstereotypen kunnen allerlei gevolgen hebben voor onder andere interesses, zelfvertrouwen en prestaties van jongens en meiden, mannen en vrouwen (e.g., Bem, 1981) en segregatie op die manier bestendigen.

Ook in Nederlandse middelbare scholen is segregatie op basis van gender zichtbaar. Mannelijke docenten zijn oververtegenwoordigd in vakken zoals wiskunde en natuurkunde terwijl vrouwelijke docenten oververtegenwoordigd zijn in vakken zoals Nederlands en Frans (DUO, 2020). De ongelijke verdeling van mannelijke en vrouwelijke docenten over exacte vakken en talen bevestigt genderstereotypen die gelinkt zijn aan deze vakken (zoals: mannen zijn van nature goed in wiskunde en slecht in talen en vrouwen zijn van nature goed in talen en slecht in wiskunde). In deze vakken kan gender van een docent invloed hebben op de genderstereotypen van leerlingen. Gender van een docent in wiskunde of Nederlands kan genderstereotypen gelinkt aan deze vakken namelijk bevestigen of tegenspreken. Blootstelling aan een bevestiging van een genderstereotiepe (bijvoorbeeld een mannelijke wiskundedocent) kan de genderstereotypen van leerlingen versterken terwijl blootstelling aan een contra- genderstereotiepe (bijvoorbeeld een vrouwelijke wiskundedocent) de genderstereotypen van leerlingen kan verzwakken. Gender van een docent kan ook de genderstereotypen van docenten beïnvloeden (Martin & Dinella, 2012; Smeding, 2012). Een vrouwelijke wiskundedocent kan bijvoorbeeld zwakkere genderstereotypen hebben dan een mannelijke wiskundedocent door haar persoonlijke ervaring die indruist tegen het genderstereotiepe. Deze effecten van gender van docenten kunnen vooral een rol spelen in vakken waarin leerlingen en docenten geconfronteerd worden met *negatieve* genderstereotypen (met andere woorden: mannen of vrouwen zijn slecht in...), omdat negatieve genderstereotypen over het algemeen sterker en gemakkelijker te internaliseren zijn dan positieve genderstereotypen (Baumeister et al., 2001). De rol van gender en genderstereotypen kunnen dus vooral belangrijk zijn voor vrouwen in exacte vakken en voor mannen in talen.

De effecten van gender van docenten evenals de effecten van genderstereotypen op de prestaties van leerlingen zijn aangetoond in eerder onderzoek. Veel onderzoek laat zien dat gender van docenten samenhangt met de prestaties van leerlingen, in het bijzonder wanneer leerling en docent hetzelfde gender hebben (e.g., Dee, 2005). We spreken dan van gendercongruentie. Daarnaast laat onderzoek zien dat leerlingen met relatief sterke genderstereotypen slechter presteren dan leerlingen met relatief zwakke genderstereotypen (e.g., Maries et al., 2018). Verder blijkt uit onderzoek dat genderstereotypen van docenten

invloed hebben op prestaties van leerlingen door gender getypeerde verwachtingen van docenten (Muntoni & Retelsdorf, 2018).

Op basis van dit wetenschappelijk onderzoek is vaak verondersteld dat genderstereotypen effecten van gender van leerlingen en docenten op prestaties van leerlingen kunnen verklaren, maar empirisch onderzoek naar dit soort mechanismen is beperkt en enkel uitgevoerd in de context van wiskundeonderwijs en de prestaties van meiden (Xu & Meier, 2021; Zhang, 2019). Er is daardoor nog maar weinig bekend over de contexten waarin effecten van gender op prestaties zich voordoen en voor welke leerlingen. Er is tevens nog maar weinig bekend over hoe effecten van gender op prestaties van leerlingen tot stand komen. Meer inzicht in de rol van gender en genderstereotypen in de prestaties van leerlingen is belangrijk om negatieve consequenties van gender en genderstereotypen in het onderwijs in de toekomst te kunnen verminderen (Onderwijsraad, 2019).

In dit proefschrift staat daarom de volgende onderzoeksvraag centraal: *wat is de samenhang tussen gender van leerling en docent en de prestaties van leerlingen en wat is de rol van genderstereotypen in deze relatie?* Het theoretisch raamwerk in dit proefschrift integreert inzichten uit verschillende disciplines en is daarmee interdisciplinair. Inzichten uit voorgaand onderzoek naar genderrepresentatie (op basis van bestuurskundige literatuur, e.g., Keiser et al., 2002) is gecombineerd met elementen uit theorieën over genderstereotypen (op basis van sociaal psychologische literatuur (Bem, 1981; Bussey & Bandura, 1999)). De hypothesen zijn getoetst in vier empirische studies met uiteenlopende onderzoeksmethoden en invalshoeken. In de vier studies onderzocht ik in welke contexten gender van leerlingen en docenten effect heeft op prestaties van leerlingen alsook de mechanismen die ten grondslag kunnen liggen aan deze effecten, vanuit zowel het perspectief van de leerling als het perspectief van de docent.

DE ROL VAN GENDER IN DE PRESTATIES VAN LEERLINGEN

WAAR EN WANNEER: DE ROL VAN CONTEXT

Hoofdstuk 2 van dit proefschrift focust op de rol van context in de samenhang tussen leerling-docent gendercongruentie en de prestaties van leerlingen. Ik verwachtte positieve relaties tussen leerling-docent gendercongruentie en de prestaties van vrouwelijke leerlingen in wiskunde en natuurkunde, en van mannelijke leerlingen in Nederlands en Frans. In deze vakken worden deze groepen leerlingen geconfronteerd met negatieve genderstereotypen over hun prestaties. Als gevolg van deze negatieve genderstereotypen kan gender van leerlingen en docenten een rol spelen in de prestaties van deze groepen leerlingen in deze specifieke vakken. Effecten van gender van docenten zouden tot stand kunnen komen doordat een vrouwelijke wiskundedocent het negatieve genderstereotiepe

dat vrouwen over het algemeen slecht zijn in wiskunde tegensprekt. Daardoor kunnen de negatieve genderstereotypen van vrouwelijke leerlingen afnemen, het zelfvertrouwen in wiskunde toenemen, en daardoor de prestaties van vrouwelijke leerlingen in wiskunde verbeteren. Daarnaast kan een vrouwelijke wiskundedocent ervaring hebben gehad met negatieve sociale barrières op basis van genderstereotypen. Als gevolg daarvan zou het kunnen dat een vrouwelijke wiskundedocent deze barrières voor vrouwelijke leerlingen wil verminderen door vrouwelijke leerlingen meer positieve feedback te geven of bij beoordeling eerder het voordeel van de twijfel te geven. Dezelfde processen zouden een rol kunnen spelen voor mannelijke leerlingen en docenten in talen.

Daarnaast kunnen effecten van gender niet alleen variëren voor vrouwelijke en mannelijke leerlingen in verschillende vakken, maar ook tussen verschillende onderwijscontexten. Effecten van leerling-docent gendercongruentie zouden sterker kunnen zijn op lagere onderwijsniveaus, op religieuze scholen en op scholen in minder bevolkte gebieden, omdat dit onderwijscontexten zijn waarin genderstereotypen relatief sterk kunnen zijn (De Vries et al., 2022; Istenic, 2007; Turner et al., 2019).

Samen met collega's verzamelde ik informatie uit leerling-administratiesystemen van middelbare scholen in heel Nederland. We analyseerden de gemiddelde cijfers van meer dan 50.000 leerlingen en hun docenten voor de vakken wiskunde, natuurkunde, Nederlands en Frans. De resultaten toonden aan dat het hebben van een docent van hetzelfde gender over het algemeen samenhangt met betere prestaties voor vrouwelijke leerlingen in wiskunde en natuurkunde, en voor mannelijke leerlingen in Nederlandse en Frans. De analyse van de verschillende onderwijscontexten wees echter uit dat deze associaties niet hoeven te gelden in iedere context en dus ook niet voor iedere groep leerlingen. De associaties tussen gendercongruentie en prestaties verschilden in sterkte en in richting voor mannelijke en vrouwelijke leerlingen over de verschillende vakken in de verschillende onderwijscontexten (e.g., voor onderwijsniveau, religiositeit en locatie van de school). Deze verschillende associaties kwamen niet overeen met de verwachte rol van genderstereotypen. Onderwijsniveau, religiositeit en de locatie van een school kunnen dus een rol spelen in de relatie tussen gender van docent en leerling en de prestaties van leerlingen, maar het is nog niet duidelijk op welke wijze.

HOE: HET PERSPECTIEF VAN DE LEERLING

Onlangs dat hoofdstuk 2 laat zien dat associaties tussen leerling-docent gendercongruentie en prestaties verschillen in verschillende onderwijscontexten, laat hoofdstuk 2 ook zien dat gender samen kan hangen met de prestaties van leerlingen in vakken waarin de prestaties van mannen of vrouwen negatief gestereotypeerd worden. In hoofdstuk 3 en 4 van dit proefschrift onderzoek ik hoe deze associaties tot stand zouden kunnen komen vanuit het perspectief van de leerling. Op basis van elementen uit de sociaal cognitieve leertheorie (Bussey & Bandura, 1999) toetste ik of genderstereotypen en zelfvertrouwen ten grondslag zouden kunnen liggen aan de effecten van gender op de prestaties van

leerlingen. In hoofdstuk 3 toetste ik de relatie tussen gendercongruentie, expliciete (i.e., bewuste) genderstereotypen en zelfvertrouwen van leerlingen door middel van een vignetexperiment onder Nederlandse middelbare scholieren. In het vignetexperiment lieten we 410 leerlingen een korte tekst lezen over een mannelijke of vrouwelijke wiskundedocent en vroegen we leerlingen daarna te reageren op een aantal stellingen over genderstereotypen en zelfvertrouwen in wiskunde. De analyses lieten zien dat de expliciete genderstereotypen en het zelfvertrouwen in wiskunde van mannelijke leerlingen sterker zijn dan die van vrouwelijke leerlingen. Daarnaast rapporteerden zowel mannelijke als vrouwelijke leerlingen meer zelfvertrouwen na het lezen van een tekstje over een vrouwelijke wiskundedocent. Toch leverden de resultaten geen bewijs voor het op basis van de theorie verwachte mechanisme, omdat de genderstereotypen van leerlingen niet werden beïnvloed door gender van de docent.

In hoofdstuk 4 bouwde ik voort op deze studie door de associaties tussen gendercongruentie, impliciete (i.e., onbewuste) genderstereotypen, zelfvertrouwen en de prestaties van leerlingen te toetsen. We maakten gebruik van data van een longitudinaal onderzoek naar gendersocialisatie in de gezins- en schoolcontext. Informatie over docenten, en over de genderstereotypen, zelfvertrouwen en schoolprestaties van leerlingen in de vakken Nederlands en wiskunde werden verzameld door middel van interviews, vragenlijsten en een impliciete associaties test. De analyses toonden aan dat leerling-docent gendercongruentie niet gerelateerd was aan de impliciete genderstereotypen, het zelfvertrouwen of de prestaties van vrouwelijke leerlingen in wiskunde of Nederlands. Ook voor mannelijke leerlingen in wiskunde was gender van hun docenten niet gerelateerd aan impliciete genderstereotypen, zelfvertrouwen of prestaties. Tegenovergesteld aan de verwachting waren de prestaties van mannelijke leerlingen met een mannelijke docent Nederlands slechter dan de prestaties van mannelijke leerlingen met een vrouwelijke docent Nederlands. De analyses toonden aan dat het hebben van sterkere impliciete genderstereotypen gerelateerd was aan het hebben van minder zelfvertrouwen en het hebben van slechtere prestaties in het vak Nederlands voor mannelijke studenten, maar gender van de docent was niet gerelateerd aan de genderstereotypen en zelfvertrouwen van mannelijke leerlingen. De resultaten leverden dus geen bewijs voor het op basis van de theorie verwachte mechanisme.

HOE: HET PERSPECTIEF VAN DE DOCENT

In hoofdstuk 5 bestudeerde ik het perspectief van de docent en onderzocht ik gender bias in het nakijkgedrag van docenten in opleiding. Gender bias in nakijken betekent dat mannelijke of vrouwelijke leerlingen anders beoordeeld worden terwijl de kwaliteit van de antwoorden op toetsen van mannelijke en vrouwelijke leerlingen gelijk is. Met andere woorden: het nakijkgedrag van docenten zou beïnvloed kunnen worden door gender van leerlingen. Ik onderzocht of deze gender bias in nakijken gerelateerd zou kunnen zijn aan gender, impliciete genderstereotypen en verwachtingen van jongens en meiden van

docenten. Samen met collega's ontwierp ik een nakijkexperiment met echte antwoorden van leerlingen op een eerder gemaakte toets. We manipuleerden de namen van de leerlingen zodat de antwoorden van vrouwelijke en mannelijke leerlingen in het experiment van dezelfde kwaliteit waren. Docenten in opleiding in het vak wiskunde en Nederlands (N = 358) namen deel aan het nakijkexperiment. De bevindingen toonden aan dat er, gemiddeld genomen, geen sprake was van gender bias in het nakijken van de docenten in opleiding. Maar het experiment toonde ook aan dat er grote individuele verschillen waren in gender bias in nakijken van docenten in opleiding. Sommige docenten gaven mannelijke leerlingen tot wel 1.5 punt (op een schaal van 1 tot 10) meer voor hun antwoorden dan vrouwelijke leerlingen en vice versa. Deze individuele variatie hing samen met de verwachtingen die docenten van mannelijke en vrouwelijke leerlingen hadden. Een docent die verwachtte dat mannelijke leerlingen meer talent hebben dan vrouwelijke leerlingen voor het vak gaf mannelijke leerlingen meer punten voor antwoorden. Verder vonden we dat vrouwelijke wiskundedocenten en mannelijke docenten Nederlands minder impliciete genderstereotypen hadden dan mannelijke wiskundedocenten en vrouwelijke Nederlands docenten. Voor wiskunde vertaalden deze minder sterke impliciete genderstereotypen zich in minder gender-getypeerde verwachtingen en in minder gender bias in nakijken.

CONCLUSIE

Dit proefschrift heeft literatuur uit verschillende wetenschappelijke disciplines samengebracht en verschillende invalshoeken en onderzoeksmethoden gebruikt om te theoretiseren over de rol van gender en genderstereotypen in de prestaties van leerlingen en de theoretische verwachtingen vervolgens empirisch te toetsen. De bevindingen in dit proefschrift leiden tot een aantal belangrijke lessen met betrekking tot de rol van gender en genderstereotypen in het onderwijs. Allereerst toonde iedere studie in dit proefschrift aan dat gender van docenten effect heeft op de prestaties van leerlingen. Daarnaast liet dit proefschrift zien dat leerlingen en docenten nog steeds genderstereotiepe ideeën hebben over de prestaties van leerlingen. Dit proefschrift liet zien dat deze genderstereotiepe ideeën bewust aanwezig kunnen zijn, maar ook onbewust (of impliciet) aanwezig kunnen zijn. Op dit impliciete niveau kunnen genderstereotiepe ideeën onbewust, onbedoeld en oncontroleerbaar resulteren in verwachtingen die geassocieerd zijn met de prestaties van leerlingen. Dit proefschrift liet zien dat de onbewuste genderstereotypen van docenten effecten van gender van docenten op prestaties van leerlingen kunnen verklaren. Gender van docenten is gerelateerd aan onbewuste genderstereotypen die kunnen resulteren in gender getypeerde verwachtingen die vervolgens kunnen resulteren in bias in de beoordeling van prestaties van mannelijke en vrouwelijke leerlingen.

Maar, en dit is de laatste les met betrekking tot de rol van gender en genderstereotypen in het onderwijs op basis van dit proefschrift, context speelt een belangrijke rol. Gender van

leerlingen en docenten hebben niet altijd effect op de prestaties van leerlingen en als die effecten er wel zijn, kunnen die effecten verschillen in sterkte en richting in verschillende contexten. Dit proefschrift liet zien dat het hebben van een docent van hetzelfde gender kan resulteren in positieve effecten en in negatieve effecten op de prestaties van leerlingen, maar dat het hebben van een docent van hetzelfde gender ook geen effect kan hebben op de prestaties van leerlingen. Zo toonde hoofdstuk 2 aan dat, over het algemeen, vrouwelijke leerlingen beter presteren met een vrouwelijke docent in de vakken wiskunde en natuurkunde, maar niet in Nederlands en Frans, en dat mannelijke leerlingen beter presteren met een mannelijke docent in de vakken Nederlands en Frans, maar niet in wiskunde en natuurkunde. Hoofdstuk 2 liet echter ook zien dat de sterkte en richting van associaties tussen het hebben van een docent van hetzelfde gender en prestaties verschilt voor onderwijsniveau, religiositeit en de locatie van een school. Deze verschillende associaties kwamen niet overeen met de verwachte rol van genderstereotypen en daardoor is het nog niet duidelijk op welke wijze onderwijscontexten effecten van gender beïnvloeden. Ook voor de rol van genderstereotypen lijkt context van belang. Zo blijkt uit hoofdstuk 5 dat genderstereotypen van docenten effecten van gender van docenten op de prestaties van leerlingen kunnen verklaren, maar alleen in het vak wiskunde, niet in het vak Nederlands.

De lessen uit dit proefschrift resulteren in drie implicaties voor de onderwijspraktijk. Ten eerste, doordat leerlingen, over het algemeen, beter presteren met een docent van hetzelfde gender in vakken waarin zij geconfronteerd worden met negatieve genderstereotypen (e.g., vrouwelijke leerlingen in wiskundige vakken, mannelijke leerlingen in talige vakken) is het wenselijk dat scholen vrouwelijke docenten in exacte vakken en mannelijke docenten in talen aantrekken en behouden. Ten tweede, doordat genderstereotypen een onbewuste en onbedoelde rol in prestaties van leerlingen kunnen spelen is het wenselijk dat bewustwording over eigen genderstereotiepe ideeën van leerlingen en docenten wordt vergroot. En als laatste, omdat de rol van gender en genderstereotypen afhankelijk is van de onderwijscontext zullen scholen moeten verkennen of gender en genderstereotypen een rol spelen in hun school en voor welke groepen leerlingen en docenten gender en genderstereotypen relevant kunnen zijn. Dit proefschrift biedt handvatten aan scholen door inzicht te geven in de mogelijke rol van gender en genderstereotypen van leerlingen en docenten in de onderwijsprestaties van leerlingen. Deze inzichten kunnen helpen om de vaak verborgen rol van gender en genderstereotypen bloot te leggen en aan te pakken. Zo kan gewerkt worden aan een toekomst waarin verwachtingen van leerlingen en docenten niet langer gekleurd worden door stereotiepe ideeën zodat leerlingen in staat worden gesteld naar hun beste kunnen te presteren.



ABOUT THE AUTHOR

Curriculum vitae

Laura Doornkamp werd op 31 augustus 1994 geboren in Groningen. In 2006 startte Laura op het VWO op het Maartens College te Haren, waar zij in 2012 haar vwo-diploma behaalde. In 2013 startte Laura met de bacheloropleiding Bestuurskunde: Beleid, Bestuur en Organisatie aan de Universiteit Leiden. Tijdens deze bacheloropleiding voltooide zij de minor Lerarenopleiding Maatschappijleer en behaalde daarmee een tweedegraads lesbevoegdheid voor het voortgezet onderwijs. In 2017 rondde Laura de bachelor bestuurskunde cum laude af. Een jaar later behaalde zij de master Public Administration: Public Management eveneens cum laude. Gedurende haar bachelor- en masteropleiding gaf Laura huiswerkbegeleiding, gaf zij les op zomerscholen en was zij student-assistent bij onder andere het Leiden Leadership Centre (Universiteit Leiden). Na haar afstuderen startte Laura een promotieonderzoek bij het interdisciplinaire onderzoeksproject Girls in Science onder leiding van Judi Mesman. Gedurende dit promotietraject verdiepte zij zich in bestuurskundige literatuur over genderrepresentatie en sociaal psychologische literatuur over genderstereotypen en sociale identiteit. Op de scheidslijn tussen bestuurskunde, sociale psychologie en onderwijskunde onderzocht Laura de rol van gender en genderstereotypen in het voortgezet onderwijs door middel van vier empirische studies met uiteenlopende invalshoeken, designs en meetmethoden. De resultaten van dit onderzoek zijn in dit proefschrift beschreven. Momenteel is Laura werkzaam als projectmanager bij VHTO: Expertisecentrum genderdiversiteit in bèta, techniek en IT waar zij haar onderzoek naar genderrepresentatie en genderstereotypen in het onderwijs voortzet en vertaalt naar de praktijk.

