Gendered teacher-student classroom interactions in secondary education: perception, reality and professionalism

Els Consuegra

Dissertation submitted to the Faculty of Psychology and Educational Sciences in partial fulfillment of the requirements for the degree of Doctor in Educational Sciences [Doctoraat in de Pedagogische Wetenschappen]

Supervisors: Nadine Engels and Koen Lombaerts

Vrije Universiteit Brussel
Brussels, Belgium

June 2015
Dedico especialmente este trabajo a mi mamá, dr. Lena Maria Martens, y a mi papá, Mario Rigomer Consuegra, por haberme apoyado siempre en todo a lo largo de mi vida y por enseñarme la fuerza y la belleza de la educación.

Supervisors:  
Prof. dr. Nadine Engels  
Department of Educational Sciences  
Prof. dr. Koen Lombaerts  
Department of Educational Sciences

Dean:  
Prof. dr. Caroline Andries

Rector:  
Prof. dr. Paul De Knop
# Table of Contents

*Acknowledgements*

*Preface*

0  **Chapter 0. General Introduction** ................................................................. 13
   0.1  **Problem** ...................................................................................................... 14
   0.1.1  Dropout ........................................................................................................ 14
   0.1.2  Grade retention ............................................................................................ 15
   0.1.3  Tracking ........................................................................................................ 16
   0.1.4  Standardised test scores .............................................................................. 17
   0.2  **Theoretical framework** ................................................................................ 19
   0.2.1  Rise of a research field .............................................................................. 19
   0.2.2  Explanatory framework .............................................................................. 20
   0.3  **Research focus** .......................................................................................... 32
   0.3.1  Procrustes programme .............................................................................. 32
   0.3.2  Dissertation project .................................................................................... 33
   0.4  **Research design** ........................................................................................ 38
   0.4.1  Integration of quantitative and qualitative methods .................................... 38
   0.4.2  Data collection and sampling ...................................................................... 39
   0.4.3  Central variables and methods of analysis ................................................ 43

1  **Chapter 1. Reviewing European Research** ..................................................... 51
   1.1  **Introduction** ............................................................................................... 54
   1.1.1  From quantity to quality ............................................................................. 55
   1.1.2  Other variables than student gender alone ................................................ 56
   1.1.3  Methodological shifts .................................................................................. 56
   1.2  **Positive interactions underexposed** ............................................................ 57
   1.3  **Research focus** .......................................................................................... 58
   1.4  **Results** ....................................................................................................... 65
   1.4.1  Descriptive results ..................................................................................... 65
   1.4.2  Negative teacher-student interaction ........................................................ 65
   1.4.3  Positive teacher-student interaction ............................................................ 66
   1.4.4  Ratio of positive to negative teacher-student interaction ............................ 67
   1.4.5  Methodological shifts .................................................................................. 68
   1.4.6  Breaking the taboo: student sexual maturity ............................................... 70
   1.5  **Discussion** .................................................................................................. 71
   1.5.1  Three new perspectives ............................................................................. 71
   1.5.2  Appreciative Inquiry ................................................................................... 73
   1.6  **Conclusion** .................................................................................................. 73

2  **Chapter 2. Increasing Teacher Awareness** ...................................................... 75
   2.1  **Introduction** ............................................................................................... 78
   2.2  **Research questions** ................................................................................... 82
   2.3  **Method** ........................................................................................................ 82
3 Chapter 3. Gendered Classroom Interactions ..................................................99
  3.1 Introduction .................................................................................................102
  3.1.1 Student gender and quantity of interactions ...........................................103
  3.1.2 Student gender and positive and negative interactions .......................104
  3.1.3 Student behaviour and teacher response ..............................................105
  3.1.4 Teacher characteristics ...........................................................................107
  3.2 Research focus .............................................................................................109
  3.3 Method ..........................................................................................................110
    3.3.1 Sample .......................................................................................................111
    3.3.2 Instruments ...............................................................................................114
    3.3.3 Analysis .....................................................................................................116
  3.4 Results ............................................................................................................118
    3.4.1 Perception of interactions: sense of equity in the classroom ...............118
    3.4.2 Actual interactions: teacher-student classroom interaction ...............120
    3.4.3 Relationship between perceived and actual interactions ....................122
  3.5 Discussion ......................................................................................................124
    3.5.1 Which boys and which girls? .................................................................124
    3.5.2 Teacher bias or disruptive boys? ..............................................................125
    3.5.3 Teaching implications? ...........................................................................126
  3.6 Conclusion .......................................................................................................128

4 Chapter 4. Feminised Teacher Professionalism ...............................................129
  4.1 Introduction ....................................................................................................132
    4.1.1 Gendered professions ..............................................................................132
    4.1.2 Emotionality and rationality .....................................................................134
    4.1.3 Theory and practice ................................................................................135
    4.1.4 Gendered teacher-student relationship ..................................................136
    4.1.5 Research focus .........................................................................................138
  4.2 Method ............................................................................................................139
    4.2.1 Mediation Analysis with Structural Equation Modelling .......................139
    4.2.2 Participants ...............................................................................................139
    4.2.3 Instruments ..............................................................................................140
    4.2.4 Covariates ...............................................................................................142
    4.2.5 Preliminary analyses ................................................................................143
  4.3 Results .............................................................................................................144
    4.3.1 Measurement model ................................................................................144
    4.3.2 Multiple mediation model .......................................................................147
  4.4 Discussion .......................................................................................................153
    4.4.1 Conclusion ...............................................................................................153
7.3.2 Students as active participants rather than passive recipients ..........224
7.4 Recommendations for practitioners and policy makers .....................226
  7.4.1 Tips and tricks for handling boys and girls? ...........................226
  7.4.2 Introducing standardised testing ............................................228
  7.4.3 We are only human, is this realistic? ......................................229

References
Appendices
Academic Track Record
Nederlandse samenvatting
List of Figures

Figure 1 Belgian student performance in PISA combined scales, by gender ........18
Figure 2. Average percentage of male and female teachers in Flanders, from pre-
          school to university education. .................................................................32
Figure 3. Integration of quantitative research base and action research...........39
Figure 4. Disproportional stratified sampling (for five Flemish provinces) ....41
Figure 5. Overview of data collection ................................................................43
Figure 6. Conceptual framework of dissertation research focus. ....................44
Figure 7. Teachers’ explicit and implicit comments on perceived misconduct and
          negative attributions for boys and girls.........................................................89
Figure 8. Hypothesised model of student background variables predicting sense of
          equity and teacher-student classroom interactions .......................................110
Figure 9. Hypothesised model of teacher background variables predicting teacher-
          reported teacher-student relationship .........................................................148
Figure 10. Results structural equation model. ....................................................150
Figure 11. Conceptual framework for assessing the effectiveness of teacher
           professional development .........................................................................164
Figure 12. Percentage of students receiving a C-certificate at the end of the first year
           of secondary education, by gender and SES ..............................................202
Figure 13. Boxplots for being off-task, by certification .....................................203
Figure 14. Histograms for calling out and raising hand ....................................222
List of Tables

Table 1. Dropout in Flanders, by gender .................................................................15
Table 2. Grade retention of at least one year in regular secondary education, by gender.........................................................................................................................16
Table 3. Qualification at the age of 22 in different birth cohorts, by gender ..........17
Table 4. Central variables in the dissertation research ........................................44
Table 5. Research methodology and analyses ......................................................49
Table 6. Reviews on gendered teacher-student classroom interaction ...............54
Table 7. Summary of studies included in the systematic literature review ..........60
Table 8. Intercoder reliability (Krippendorff’s alpha, kalpha) for key codes in content analysis...........................................................................................................85
Table 9. Overview of nature of themes of teachers’ comments (thematic analysis).91
Table 10. Survey and observation sample, by gender, SES, foreign language background and migration background .........................................................113
Table 11. Observation codebook and intercoder reliability (Krippendorff’s alpha, kalpha) for each code .......................................................................................................................116
Table 12. Results multiple regression models predicting sense of equity ............119
Table 13. Results Mann-Whitney U test group comparisons for teacher-student interactions ..............................................................................................................121
Table 14. Spearman rho correlation matrix for boys and girls ...........................123
Table 15. Goodness-of-fit indicators for the latent variables .................................145
Table 16. Standardised and unstandardised coefficients for confirmatory factor analyses .........................................................................................................................146
Table 17. Correlation table for the independent variables .................................148
Table 18. Results from structural equation model: direct, indirect and total effects .........................................................................................................................151
Table 19. Control and intervention group samples, by gender .........................168
Table 20. Descriptive results pre-test and post-test for teacher-student interactions and student sense of equity .................................................................173
Table 21. Results generalized linear mixed modelling analyses (best models) .....179
Table 22. Descriptive statistics for post-test survey and observation sample .......200
Table 23. Odds of grade retention (C-certificate) results from logistic regression analyses ...............................................................................................................202
Acknowledgements

Zonder de juiste condities, geen kennisproductie en dus ook geen doctoraatsthesis. Ik wens verschillende collega’s, vrienden en familie te bedanken voor hun rechtstreekse en onrechtstreekse bijdrage in het schrijven van dit proefschrift.


marginale muziek, Vicky en Colruyt en ik mijn duiven), het ventileren van frustraties en het samen vieren van feest.

In vergelijking met Myriam en Vicky verbleken mijn andere collega’s spijtiggenoeg in het ... Hahaha. Mopje! Ook mijn andere PE- en IDLO-collega’s verdienen een vermelding in mijn dankwoord. Jullie allemaal bij naam noemen, lijkt me niet opportuun (omdat jullie zo talrijk zijn, maar ook omdat het laat is en ik geen risico wil lopen op het vergeten van mensen). Samenvattend: bedankt voor de filosofische gesprekken tijdens de vele lunches in de resto. Bedankt voor de bij momenten censureerbare maar hilarische gesprekken bij taart en drank. Bedankt voor de kunstfilms voor volwassenen, de zomerbars, de tekeningen, de avonturen van Ronny de Ponny, en de crazy party in Leiden daags voor het indienen van mijn proefschrift.

Dit onderzoek had niet mogelijk geweest zonder de bijdrage van zovele honderden leerkrachten, leerlingen, ouders en directies die deelnamen aan het Procrustes onderzoek. Ik wens hen uitdrukkelijk te bedanken voor hun vertrouwen. Een heel bijzondere dankjewel is voor de ‘interventieleerkrachten’ die deelnamen aan het professionaliseringstraject en bij wie ik herhaaldelijk over de vloer mocht komen met mijn camera’s. Dank voor jullie lef om met mij in zee te gaan en dank voor de toegang die jullie me gaven tot jullie immens rijke praktijk. Ik hoop dat ik jullie op de verdediging van dit proefschrift mag ontvangen en dat jullie me op jullie beurt vragen kunnen stellen over de vragen die wij jullie de voorbije jaren hebben gesteld (Doordenkertje!). Ook wens ik te bedanken: Joseph Kessels, Niel Van Meeuwen en Peter Beschuyt. De laatste twee heren voor hun facilitatie van de professionaliseringstrajecten en Joseph voor het geslaagd modelleren van de waarderende benadering. Mijn tijd bij Kessels & Smit, nog voor ik onderzoeker was aan de VUB, was fundamenteel belangrijk in mijn professionele vorming. Ik draag jullie voor altijd mee in mijn praktijk.

Ook mijn vrienden van het Mysterie van Onderwijs hebben zich diep in mijn professionele hart genesteld: Bert Smits, Pat Vandewiele, Bavo Wouters, Stijn Dhert, Ann Martin, Saskia Vandeputte, Peter Beschuyt, Kristof Bleus, Maaike Eggermont, Noura El-Jafoufi, Amina Bouzakoura, Bahattin Kocak,
An Lanssens, Myriam Halimi, Ellen Huyge en Anouk Van Der Wildt. Wanneer ik als academicus dreig op te stijgen naar de hogere étages van de ivoren toren, houden jullie mij met beide voeten in de praktijk. Dank voor de inspirerende avonden en heerlijke weekends, bedankt voor de radicale dromen en de wilde utopie.


Ten slotte rest mij nog te bedanken: de meest cruciale figuren in mijn omgeving.

Mijn beste vrienden van het atheneum
Mijn beste vrienden van de VUB
Mijn ouders, mijn zussen en mijn liefste
Verpletterend belangrijk is jullie bijdrage aan mijn ‘academische successen’. De superkracht die jullie me geven is van zo’n crimineel onschatbare waarde dat ze gelukkig niet belast wordt in dit land. Dank voor jullie onvoorwaardelijke liefde, dank voor de rust in mijn soms woelige leven. Dank voor de toekomst die straalt.

Els Consuegra
23 juni 2015
Preface

This doctoral dissertation is the result of four years of research on the topic of gendered teacher-student classroom interactions in secondary education in Flanders. The research was funded by the government Agency for Innovation by Science and Technology [Agentschap voor Innovatie door Wetenschap en Techniek, IWT, Grant IWT110020]. The dissertation is composed of eight chapters. Chapter 0 provides a general introduction to gendered achievement in compulsory education and presents an explanatory framework for boys’ scholastic underachievement. The chapter also introduces the central research focus and research questions of the dissertation, the overall methodological approach and the studies that have been performed. Chapters 1 to 6 constitute the main body of the dissertation. Six papers are presented which have either been presented at international conferences or which have been published in, or are in review for, journals included in the Social Citation Index of the Web of Science. This means that the papers have gone through, or are going through, a process of anonymous peer-review. The dissertation ends with Chapter 7 in which the main research findings and the limitations of the reseach and directions for further research and educational policy and practice are discussed.
Chapter 0
General Introduction
0.1 Problem

Education in Flanders ranks among the top for academic performance in reading, mathematical and scientific literacy. Unfortunately, various studies show that significant differences exist between different groups of students, such as socio-economically advantaged or disadvantaged students and students from different ethnic or cultural traditions. Another gap that has received increasing attention is the gap between boys and girls. Significant disparities exist and as this introduction will reveal, even when boys and girls are equally proficient, their school achievement is markedly different – and this has a significant impact on their decision to pursue further education and their choice of career.

This introduction provides a broad overview of gender disparities in educational achievement in compulsory education. Gaps in the existing evidence-base will be emphasised and the research focus and research design of this dissertation will be presented.

0.1.1 Dropout

Van Landeghem, De Fraine, Gielen, and Van Damme (2013) studied the dropout of 17 to 22 year olds in Flanders. Table 1 shows the evolution of dropout rates for boys and girls. The numbers up until 2010 are retrieved from Van Landeghem et al. (2013) and data from the last official report *Vroegtijdig Schoolverlaten Vlaams Secundair Onderwijs* for the years 2010 to 2013 (Flemish Department of Education and Training, 2013) are added. An increase of dropout rates can be observed for both boys and girls since 1999 up until 2007. After 2007 a downward trend can be observed. This plunge has received ample media coverage recently (Verstraete & Vermeersch, 2015). No attention was paid, however, to the gap between boys and girls which remains unchanged: more boys (16%) than girls (10%) leave our educational system before qualifying.
Table 1. Dropout in Flanders, by gender

<table>
<thead>
<tr>
<th>Year</th>
<th>Boys %</th>
<th>Girls %</th>
<th>Total %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1999</td>
<td>14.5</td>
<td>9.4</td>
<td>12.0</td>
</tr>
<tr>
<td>2000</td>
<td>15.6</td>
<td>9.9</td>
<td>12.8</td>
</tr>
<tr>
<td>2001</td>
<td>15.3</td>
<td>9.4</td>
<td>12.4</td>
</tr>
<tr>
<td>2002</td>
<td>15.7</td>
<td>10.0</td>
<td>12.9</td>
</tr>
<tr>
<td>2003</td>
<td>16.6</td>
<td>10.1</td>
<td>13.4</td>
</tr>
<tr>
<td>2004</td>
<td>16.9</td>
<td>10.2</td>
<td>13.6</td>
</tr>
<tr>
<td>2005</td>
<td>16.7</td>
<td>10.3</td>
<td>13.6</td>
</tr>
<tr>
<td>2006</td>
<td>17.6</td>
<td>10.5</td>
<td>14.1</td>
</tr>
<tr>
<td>2007</td>
<td>17.9</td>
<td>10.6</td>
<td>14.4</td>
</tr>
<tr>
<td>2008</td>
<td>17.6</td>
<td>10.4</td>
<td>14.1</td>
</tr>
<tr>
<td>2009</td>
<td>16.4</td>
<td>10.2</td>
<td>13.4</td>
</tr>
<tr>
<td>2010</td>
<td>15.5</td>
<td>9.9</td>
<td>12.7</td>
</tr>
<tr>
<td>2010/11</td>
<td>15.5</td>
<td>9.7</td>
<td>12.7</td>
</tr>
<tr>
<td>2011/12</td>
<td>14.2</td>
<td>9.7</td>
<td>12.0</td>
</tr>
<tr>
<td>2012/13</td>
<td>13.8</td>
<td>9.4</td>
<td>11.7</td>
</tr>
</tbody>
</table>

Note. Van Landeghem et al. (2013)(for the years 1999-2010) and Flemish Department of Education and Training (2013)(for the school years 2010-2013)

0.1.2 Grade retention

Figures on grade retention indicate a similar lower performance of boys in comparison to girls (Datawarehouse Flemish Education and Training, s.d.). Table 2 shows the percentage of boys and girls that have repeated one or more years in regular secondary education. In the last grade of secondary education 30% of girls and 39% of boys have doubled a year or more. These numbers receive our attention, since grade retention has proven to have negative effects on academic self-concept, early school leaving and participation and success in higher education (Lamote, Van Damme, Van Den Noortgate, 2013; Lamote, Pinxten, Van Den Noortgate, & Van Damme, 2014).
Boys are not only repeating a year more often, they are also overrepresented in what is often perceived as the ‘lower’ tracks of secondary education (Van Landeghem, Goos, & Van Damme, 2010). Secondary education in Flanders is divided into four branches: GSE (General Secondary Education), ASE (Artistic Secondary Education), TSE (Technical Secondary Education), and VSE (Vocational Secondary Education). These branches are introduced in the second grade of secondary education. In the first grade only the regular and a V-track exist. The latter prepares for the vocational education track. At the end of each school year, students receive not only a school report, but also a certificate stating in which year and track they are allowed to enrol next year. An A-certificate grants access to the next year without restrictions. A B-certificate implies two options: repeating the current year or promotion with restrictions from certain tracks. A C-certificate forces the student to repeat the current year. The Flemish educational system is often referred to as a ‘waterfall’ in which students gradually descend to a ‘lower’ track rather than opting for a track based on the content of the programme. GSE is perceived to be the most academically demanding track and we see a clear underrepresentation of boys in this branche (see Table 3).

**Table 2. Grade retention of at least one year in regular secondary education, by gender**

<table>
<thead>
<tr>
<th>School year</th>
<th>1&lt;sup&gt;st&lt;/sup&gt; grade</th>
<th>2&lt;sup&gt;nd&lt;/sup&gt; grade</th>
<th>3&lt;sup&gt;rd&lt;/sup&gt; grade</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Boys %</td>
<td>Girls %</td>
<td>Total %</td>
</tr>
<tr>
<td>2009-2010</td>
<td>24.0</td>
<td>20.8</td>
<td>22.4</td>
</tr>
<tr>
<td>2010-2011</td>
<td>23.9</td>
<td>20.5</td>
<td>22.2</td>
</tr>
<tr>
<td>2011-2012</td>
<td>24.2</td>
<td>20.7</td>
<td>22.5</td>
</tr>
<tr>
<td>2012-2013</td>
<td>23.8</td>
<td>20.4</td>
<td>22.2</td>
</tr>
<tr>
<td>2013-2014</td>
<td>24.5</td>
<td>20.4</td>
<td>22.1</td>
</tr>
</tbody>
</table>

*Note. Datawarehouse Flemish Education and Training (s.d.)*
Again, these gender disparities call for concern, since the track followed by students in secondary education predicts their success in higher education. Students who obtained their qualification in VSE, ASE, or TSE have a success rate of 12% in their first year of university while the success rate of students in GSE ranges from 25% up to 77% (Declercq & Verboven, 2010). In total, only 55% of boys (versus 67% of girls) decide to start higher education after secondary education (Van Woensel, 2007). Buchmann (2009) analysed gender inequalities in the transition to college and recommends tackling gender inequalities in higher education by examining gender inequalities early in the educational life course.

### 0.1.4 Standardised test scores

Figure 1 presents the PISA scores of 15-year-old boys and girls in Belgium on the combined reading, mathematical and scientific literacy scales. We see that girls outperform boys in reading, but the opposite is true for mathematics. For science the difference was not statistically significant in 2000 and 2003. In 2009 a small underperformance of girls was observed.
In the large majority of countries that participate in PISA, among high-performing students, this pattern of girls doing worse than boys in mathematics and *visa versa* for reading is observed (OECD, 2015). However, in top-performing economies in PISA, such as Shanghai-China, Singapore, Hong Kong-China and Chinese Taipei, girls perform equally well as boys in mathematics and attain higher scores in mathematics than boys in most other countries and economies around the world.

OECD in their extensive report “The ABC of Gender Equality in Education” (OECD, 2015) conclude that gender disparities in PISA scores do not stem from innate differences in ability. Rather, they stem from differences in attitudes and behaviours of students, and teachers and parents are known to play a significant role in shaping these. Also, the gender gap that is observed in PISA scores is not the same as the gender gap observed in school marks in both language-of-instruction and mathematics classes. The gap in school performance is much bigger than would be expected based on the PISA scores. Boys and girls that score equally well in the PISA test are not performing equally well at school (OECD, 2015). Something thus is happening in schools which impedes boys to perform to their full potential. In
the following sections we will review several explanations and hypotheses concerning boys’ underperformance in school.

0.2 Theoretical framework

0.2.1 Rise of a research field

In the problem statement we presented evidence for the academic underachievement of boys in Flanders: they tend to dropout of school early, they repeat a year more often, they are overrepresented in the academically less challenging tracks of secondary education, and they are less likely to attend higher education. These Flemish data confirm a phenomenon that has received increasing policy and media attention since the 1990s in North America, Australia, the UK and many parts of Western Europe (Arnot, David, & Weiner, 1999; Mendick, 2013; Moureau, 2009). Research on the ‘underachievement of boys’ has clearly been in vogue and the number of studies on the topic has risen accordingly (Gorard, Rees, & Salisbury, 2001). However, debate exists on whether the underachievement of boys is new or not (Smith, 2003). In the UK Cohen (1998) and Mahony and Smedley (1998) analysed the failing of boys from a historical perspective and found that from the 1950s throughout the 1980s girls were outperforming boys. “The ‘problem’ is not new even though it has been given a new significance” (Mahony & Smedley, 1998, p. 41). In addition, Gorard (2001) claims that boys have never attained higher grades than girls in compulsory education at any time since the 1970s. Not only the evolution but also the magnitude of the gender gap in educational attainment is disagreed upon. Moureau (2009) analysed the issue from a state level perspective, and shows that the UK and France share similar patterns in terms of gender disparities in achievement but the UK panic about boys’ underperformance is not shared in France.

In Flanders the underperformance of boys first came into central focus in the early 2000s. Derks and Vermeersch in 2001 performed a study on grade retention in secondary education and this work was followed by many other researchers such as Siongers (2002), Van de Gaer (2006) and Bossaert (2009). In 2012 the large scale strategic research project ‘Teaching in the bed
of Procrustes’ was launched. It is within this Procrustes-programme that this dissertation research is situated. This dissertation presents one of the many studies that have been performed over the course of the last four years. The underperformance of boys has clearly been demonstrated, however, research has been inconclusive about the extent and evolution of the gap and empirical evidence has been inconclusive about possible explanations underlying the problem (Laevers, Van Houtte, Engels, Lombaerts and Vermeersch, 2011). Salisbury, Rees and Gorard (2010) review recent findings of research and policies attempting to explain and overcome differences in educational outcomes in boys and girls but discover that there is no totally convincing explanation of the phenomenon and therefore little hope of an effective strategy to deal with it.

The Procrustes project is aimed at doing away with shortcomings in the existing literature and is aimed at gaining insight into the processes and factors that play a part in gendered school achievement. In order to avoid oversimplification, both differences within and between gender groups were taken into account. Also, the project sought to assess the effectiveness of strategies to make sure all boys and girls can participate successfully in education.

0.2.2 Explanatory framework

In this paragraph a series of explanations and theories for boys’ underperformance will be presented. We organise this overview along the lines of three levels: the individual, the interactional and the institutional level. This classification is based on Risman and Davis’ (2013, p. 747) conceptualisation of gender as a social structure which states that “gender inequality is produced, maintained and reproduced at each level of social analysis (individual, interactional and institutional)”. The state of the art that will be presented is a general exploration of the body of literature regarding gendered academic achievement. In depth literature reviews will follow in the upcoming chapters, when introducing the different empirical studies that have been performed.

1 The name ‘Teaching in the bed of Procrustes’ refers to the Greek myth of Procrustes, son of Poseidon, who had an iron bed, in which he invited guests to spend the night. Guests
0.2.2.1 Individual level

Biological-genetic explanations

“There is no reason to deny the influence of bodies on how selves develop.” (Risman & Davis, 2013, p. 747).

Although some researchers use ‘sex’ and ‘gender’ interchangeably, a clear distinction can be made between both terms. Sex refers to a biological categorisation and gender refers to social, psychological and cultural components related to sex (Archer & Lloyd, 2002). Plenty of research has been conducted to biological and genetic differences between the sexes. Traditionally, however, there has been little concern in this field with the relationship between such differences to inequalities between men and women (Risman & Davis, 2013). Recently, there has been renewed attention to biological-genetic explanations when trying to explain the divide between boys and girls concerning their academic performances (Arnot et al., 1999). Studies in this field are concentrated in the neuro- and cognitive sciences, mostly focusing on the study of brain structure, brain development and hormonal characteristics (Fischer, 2007; Risman & Davis, 2013). Several critical reviews of literature however have proven these studies to be inconsistent and methodologically flawed (Jordan-Young, 2010). In 2005 Hyde reviewed 46 meta-analyses on gendered cognition and found that in half of the studies the sex differences were small and in another third they were virtually nonexistent. Based on her own work in the field of neuroplasticity and based on an extensive review of literature, Neuroscientist Eliot (2010) argues that boys and girls differ to some small extent (e.g. higher levels of testosterone for boys), but these small differences do not account for the behavioural differences in, for example, aggressiveness. It is our socialisation, she argues, that blows minor biological differences up to extremes. Spelke (2005) in her review of 111 studies concludes somewhat in the same direction that cognitive systems that emerge in early childhood are indeed gendered, but at birth men and women on the whole possess equal aptitude. For example, boy and girl infants at 6 months were found to perform equally well on tasks that underlie mathematical abilities.
**Sex role attitudes**

Serious attempts to study sex and gender inequalities emerged during the second wave of feminism\(^2\) and were introduced by the first female scientists (Risman & Davis, 2013). One concept that has emerged and is still studied in relation to educational attainment are sex role attitudes. Early definitions assumed that ‘masculinity’ and ‘femininity’ were opposite ends of one dimension and that an individual was either high on femininity and low on masculinity, or the opposite. Researchers studying personality traits such as efficacy and empathy however soon discarded the use of a unidimensional masculinity and femininity scale. It was generally assumed that different and multiple masculinities and femininities exist (Epstein, 1998; Connell, 1995; Schippers, 2007). In 1973 Constantinople stated that femininity and masculinity are probably two of the muddiest concepts in social sciences. In 2009 Davis and Greenstein define the sex role attitudes independent from masculinity and femininity as “individuals’ levels of support for a division of paid work and family responsibilities that is based on the belief in gendered separate spheres” (p. 88). Individuals with a traditional view support a gendered division of family labour, regarding women as homemakers who are responsible for parenting and men as wage earners. Egalitarian or liberal sex role attitudes include a more equal view on participation in labour and domesticities (Davis & Pearce, 2007). More traditional sex role attitudes are observed to be negatively related to school achievement (Updegraff, McHale, & Crouter, 2004), school motivation (Leaper, Farkas & Brown, 2012), and expectations and desire of attending a college education (Davis & Pearce, 2007). A recent review of research by Halimi, Consuegra and Engels (2015) shows that boys are more likely to be traditional-minded than girls, however, there is evidence pointing to a narrowing of the gap. The negative effects that boys experience as a consequence of their more traditional sex role attitudes is what O’Neil (1981) would identify as ‘gender role conflicts’, which surface when the socialised gender roles generate negative effects either for the individual enacting the role or for (significant) others. We pick up on this

---

\(^2\) Whereas first-wave feminism was focused on tackling legal obstacles to gender equality (e.g. voting rights, property rights), second-wave feminism broadened the debate to a wide range of issues such as sexuality, family, work, reproductive rights. Not only ‘de jure’ but also ‘de facto’ inequalities were addressed.
topic later on in the section 2.2.2. when discussing boys’ and girls’ interactions with peers.

*Gender identity*

Another approach to studying sex and gender and their effects on academic achievement has been provided by gender identity theory. The ‘identity’ concept is argued to provide great benefits in studying gendered educational outcomes (Kessels, Heyder, Latsch, & Hannover, 2014) since it provides a more dynamic approach than the traditional sex dichotomy (Gee, 2000). The term ‘identity’ has taken on many different meanings in literature, and Gee (2000) suggests the following broad but straightforward definition: the kind of person one is recognised as being, at a given time and place. In this sense, each person has multiple identities connected not to their internal states, but to their roles and positions in society. This is another approach to identity than the concept of a “core identity” that is kept constant for oneself and others across contexts. Gee does not deny that such concept of identity is useful, however he prefers to emphasise that identity can be ambiguous and unstable, that it can change from moment to moment and from context to context. Gender identity is believed to be one of many components in one’s greater identity. In their literature review on gender identity Vantieghem, Vermeersch and Van Houtte (2014a) conclude that too little research has attempted to operationalise gender in a way that genuinely transcends the sex dichotomy. Hence, Vantieghem et al. (2014a) suggest a gender identity theory which integrates the sociological theory of ‘doing gender’ which focuses on interpersonal interactions and symbolic behaviour in the social sphere, and the social psychology theory that departs from the intrapersonal level as a self-evaluation of masculinity and femininity. Gender identity is then defined as the degree to which a person perceives the self to be masculine or feminine, given what it means to be masculine or feminine in a given society in a certain time and setting. Vantieghem, Vermeersch and Van Houtte (2014b) show that girls who score higher on this self-perceived femininity score highest on academic self-efficacy. Boys and girls who feel atypical, score lowest. In school (sub)cultures where the perception dominates that displaying effort and engagement at school is feminine a misfit between boys’ gender identity and academic engagement can occur
(Kessels et al., 2014). This mechanism is very similar to the ‘gender role conflict’ described by O’Neil (1981). We continue the discussion of this conflict in section 2.2.2. since it involves the interactional-level rather than the individual-level explanations for school underperformance.

**Gendered attitudes**

Boys and girls absorb society’s or subcultures’ notions of behaviour which are appropriate for their gender from a very early age on. This results in gendered attitudinal differences towards school and learning. We have already discussed that a dichotomous approach to studying gender has its limitations, however, much research still departs from such a point of view and we do not want to fail in presenting their results. The OECD (2015), for example, in their PISA studies surveyed pupils on their attitudes towards school and learning. Across most countries and economies that participated in PISA 2012, boys were more likely than girls to report that school is a waste of time and boys were less likely to agree that trying hard at school is important or that they enjoy receiving good marks (OECD, 2015). These results are in line with previous research. Warrington, Younger & Williams (2000) found that boys’ study motivation is considerably lower than that of girls and Wang, Willet and Eccles (2011) show that boys’ emotional school engagement (valuing of school education and school belonging) is lower compared to that of girls. Wang et al. (2011) also found boys’ behavioural school engagement (attentiveness and school compliance) to be lower. Behavioural disengagement such as arriving late, skipping classes, not caring about homework and spending less time completing it, results in boys missing out on learning opportunities, falling behind in class and earning low marks – all of which feed back into new disengagement, completing a vicious cycle (Browne & Mitsos, 1998; Clark, Oakley, & Adams, 2006; OECD, 2015). With boys being trapped in this self-fulfilling prophecy, it should not surprise that their school well-being is found to be significantly lower than girls’ (Engels, Aelterman, Van Petegem, & Schepens, 2004).
0.2.2.2 Interactional level

We already hinted to the process of socialisation as a mechanism for production, maintenance and reproduction of inequalities. Socialisation can be defined as the (un)conscious process by which people learn to behave in a certain way through social interaction with several actors (parents, peers, teachers ...) and through socialization agents such as the media, textbooks etc. (UNICEF, 2007). Socialisation is thus a key process at the interactional level. We discuss the effects of three key socialisation agents (peers, parents, and teachers) on students’ gendered academic performance (and related concepts).

Peers

Peers are of key importance in (gender) socialisation processes which in turn influence students’ attitudes and behaviours, especially in adolescence (Coleman, 2011; Brenchwald & Prinstein, 2011; Baumeister & Leary, 1995). Warrington and Younger (2000) found boys in comparison to girls to be more concerned with the way they are perceived by their peers. Flemish research confirms this increased sensitivity to peer pressure of boys, showing that boys are more susceptible towards dominant school related attitudes of their classmates than girls (Van de Gaer, 2006). Research also shows that boys and girls prefer the company of peers of the same gender and identify to a greater extent with same-sex peers (Kandel, 1978; Warrington & Younger, 2000). Given the preference for same-sex friendships and the importance of acceptance by peers during adolescence, gendered subcultures emerge in secondary education. These subcultures are identified as an important mechanism for transmission and reinforcement of boys’ negative attitudes towards school and learning (Van Houtte, 2004). Especially among boys, anti-schoolish subcultures develop in which academic achievement and putting effort into schoolwork is perceived to be ‘not cool’ (Kleinfeld, 1997). For girls it is easier to combine being ‘cool’ with a visible status as high achiever (Van Houtte, 2004). Nevertheless, research shows that these anti-school attitudes can also be present in girls (Lyng, 2009; Jackson, 2006; Renold, 2001) and also in high achieving populations. Renold (2001) interviewed girls expressing a gender role conflict between ‘being clever’ and being feminine and a study of Watts & Borders (2005) found that more
academically motivated boys most frequently express a conflict between spending time with friends and doing schoolwork. Myhill (2002) as well documents that high achieving boys can behave in an anti-schoolish way in class (not actively engaging in positive interactions with teachers). She observes an evolution of high performing boys taking over the behavioural patterns of low achievers as the years progress. In girls Myhill does not observe this evolution.

When taking into account the gendered effects of peers on academic achievement, debates about single-sex and coeducational schooling come into focus. The gendered class and school composition is a recurring topic in policy debates. Debates are sharp since international research on the topic appears to be inconclusive or neutral pointing to no benefits nor disadvantages (Buchmann, DiPrete, & McDaniel, 2008). Assessing the effects of a specific educational approach is never an easy and clear-cut business. Pahlke, Hyde and Allison (2014) meta-analysed 184 studies on single-sex schooling and they found that much studies are methodologically weak, for example not controlling for selection effects. The highest quality studies (involving controls or using random assignment) show no benefits for students’ achievement or academic interest for single-sex when compared to coeducational schooling. As to the ratio of boys and girls in secondary schools, Flemish research shows that a slight majority of girls promotes a more study-oriented culture among both girls and boys (Van Houtte, 2004).

**Parents**

Parents are indicated as one of the primary socialising agents. Halimi et al. (2015) provide evidence for a parent-child congruence in sex role attitudes, meaning that children of parents with traditional views hold more traditional views themselves and visa versa. Parents’ sex roles are not only ‘transmitted’ to their children, they also influence parents’ educational expectations towards their children. This socialisation starts at a very early age with for example parents letting baby boys cry longer than girls before holding them or parents estimating that baby boys can crawl up steeper ramps than baby girls (Eliot, 2010) while no difference was present in actual crawling performances. This is the start of boys learning that showing their emotions will not pay. At preschool age, Baker and Milligan (2013) found that parents
spend more time with girls for ‘teaching activities’ such as reading and the use of numbers and letters. In middle school, parents are found to hold lower educational expectations for boys than for girls (Halimi, Consuegra, Engels, & Struyven, 2014). These lower parental educational expectations are nontrivial, since high parental expectations are positively associated with study outcomes (Bodovski, 2014; Froiland & Davison, 2014) and adolescents educational aspirations (Kirk, Lewis-Moss, Nilsen, & Colvin, 2011). Froiland and Davison (2014) found parental expectations to have even stronger effects than SES in middle and high school. Zhang, Haddad, Torres and Chen (2010) in a large-scale longitudinal study confirm that parents’ expectations influence adolescents’ academic achievement. Their analyses also revealed parents’ expectations and adolescents’ expectations mutually influence each other with the effects of parents’ expectations on students’ own expectations being stronger among males than among females.

More structural elements of family context (such as SES, migration background or foreign language background) are widely acknowledged to be an important predictor of students’ academic achievement. With regard to gender disparities in education, it is interesting to note that family context can generate different effects for boys and girls. Mensah & Kiernan (2010) for example found that boys in families where mothers are young or lack qualifications are more disadvantaged compared to girls in similar circumstances. Gender gaps in education are also shown to evolve differently in different ethnic groups (e.g. McDaniel, DiPrete, Buchmann, & Shwed, 2011). These results advocate for a more systematic evaluation of interaction effects between the long-standing axes of social inequality predicting the achievement gaps. Relatively little research has, however, evaluated interactions between these variables (Strand, 2014). An intersectional-informed approach (Collins, 1998) is therefore suggested, acknowledging the individual as irreducible to a single characteristic (Hankivsky, 2014). Intersectionality argues for researching the intersects of multiple dimensions of diversity as opposed to examining each distinctive axe since several dimensions of oppression mutually construct one another.
Teachers

Not only the expectations of parents but also those of teachers can affect student outcomes. The existence of teacher expectation effects is long-established. The gendered character and strength of these effects have been discussed heavily. First, we need to distinguish between direct and indirect teacher expectation effects. An example of direct effects is the practice of teachers providing girls with easier grading than boys (Bonesronning, 2008; Lavy, 2004; Marcenaro-Gutierrez & Vignoles, 2015). Lavy (2004) however does note that in comparison to blind grading or standardised tests grading practices of teachers could reflect a value of student efforts and girls could systematically provide more effort than boys. It can be argued that in some cases it is desirable to take into account aspects such as effort or neatness. Teacher expectations can also have indirect effects on student outcomes. In the 1960s and 1970s it is hypothesised that teacher expectation effects are mostly mediated through teacher classroom interactions (Rosenthal & Jacobsen, 1968; Brophy & Good, 1970). A complex but very informative and detailed model of teacher expectation effects is suggested by Brophy and Good (Brophy, 1985a, pp. 307-308):

2. Consistent with these differential expectations, teachers behave differently toward different students.
3. This differential teacher behaviour communicates to each individual student something about how he or she is expected to behave in the classroom and perform on academic tasks.
4. If teacher treatment is consistent over time, and if students do not actively resist or change it, it will likely affect student self-concept, achievement motivation, level of aspiration, classroom conduct, and interactions with the teacher.
5. These effects generally will complement and reinforce the teacher’s expectations, so that students will conform to these expectations more than they might have otherwise.
6. **Ultimately, this will make a difference in student achievement and other outcomes, indicating that teacher expectations can function as self-fulfilling prophecies.**

In 1985 Good and Findley reviewed the literature on gendered teacher expectations. They suggest that the majority of teachers does not hold explicit gendered expectations but that teacher treatment is gendered, however a decrease of the differences between boys’ and girls’ interaction patterns can be observed since the 1960s. A meta-analysis of Dusek & Joseph in 1985 and the work of Cooper and Moore in 1995 support the lack of evidence for gender differentiated teacher expectations. Based on these findings, in the 1980s and 1990s it was claimed that differences in teacher-student interactions are rather due to differences in student behaviour than to teacher bias. It is proposed that boys bring criticism upon themselves, as a consequence of their greater levels of (visible) disruptive behaviours (Howe, 1997). Classroom behaviours thus need to be assessed.

Research on gendered classroom interactions has a long tradition with the majority of studies being performed in the USA (Howe & Abedin, 2013). Several review studies learn that boys in general receive more negative feedback from their teachers than girls (Olivares & Rosenthal, 1992; Howe, 1997; Jones & Dindia, 2004; Beaman, Wheldall & Kemp, 2006). Research is however not conclusive about the degree to which this is caused by student-initiated unwanted behaviours. Also, conflicting results exist for the amount of positive feedback towards boys and girls. In their methodological review Harrop & Swinson (2007) ascribe these inconsistencies concerning positive classroom interactions to differences in the operationalisation of ‘positive’ interactions (sometimes also including neutral feedback but not always). We add to that by pointing to the tradition of deficiency-based research showing more interest in negative interactions than in positive ones (Seligman, Steen, Park, & Peterson, 2005). Howe in 1997 suggests a focus on the ratio of positive to negative interactions rather than focusing on the absolute amounts of each separate category of feedback, since it may be more informative in terms of impact on students. Some studies show correlations between the ratio of positive to negative interactions and pupil on-task behaviour which is in turn believed to be related to higher academic achievement (Myhill, 2002; White, 2010). This correlation however does not necessarily imply a causal
relationship. The complex model of teacher expectation effects described earlier illustrates that a specific pattern of teacher feedback will not automatically translate into specific student outcomes. Predictions are risky since one can hypothesise wrong relationships in any one of the steps in the model.

Another distinction we need to make when discussing expectation effects is distinguishing between negative (‘Golem’) and positive (‘Galatea’) teacher expectation effects (Brophy, 1985a). Self-fulfilling prophecy effects can generate positive as well as negative effects and are thus not the problem as such. One could even theorise that systematically projecting positive expectations on students should produce more positive outcomes than comparable instruction presented more neutrally. Brophy is however not an advocate of projection of inaccurate expectations and treating students as if they are brighter or more studious than they actually are. He also fears that it would be too difficult to do this consistently, since we are accustomed to responding to our real expectations. If the expectations teacher hold and believe to be real are in fact inaccurate and biased, this present a great challenge. Evidence exists for a ‘gender blindness’ of teachers who are not aware of the fact that they take student gender into account when interacting (Garrahy, 2001; Raider-Roth, Albert, Bircann-Barkey, Gidseg, & Murray, 2008). Recently, the work of Wellborn, Huebner and Hills (2012) obtained significant findings on how teacher expectations can be turned more optimistic for students in underachieving groups (non-native speakers) without necessarily projecting inaccurate expectations. Teachers were provided with strength-based information on their students on a variety of subjects and this positively affected teachers’ expectations.

The final recurring debate when discussing teacher gender effects on student achievement concerns the question whether boys achieve better when taught by male teachers. Relatively little is in fact known about the effects of ‘gender matching’ of teachers and pupils in schools (Carrington et al., 2007). Recently, a large-scale study of Helbig (2012) analysed the 2007 TIMMS and 2006 PIRLS data for over 100,000 primary education students and found no significant teacher gender effects for math nor reading. Gender of teachers is also found to have little or no effect on the academic motivation and engagement of either boys or girls (Carrington et al., 2007). Howe’s 1997 review concludes that there are few effects of teacher gender on interaction
patterns with boys and girls. Still, some studies do find some specific teacher gender effects. Students themselves for example report that teacher gender matters in terms of the construction of their own gender identities and the definition of ‘proper masculinity’ and ‘proper femininity’ (Skelton et al., 2009). Also, Einarsson and Granström (2002) found that male teachers in higher secondary education pay more attention to female pupils. The authors relate this to romantic and libidinous aspects in teacher-student interactions, which are often considered taboo. In 2010 Andrew, Marsh, Cheng, and Ginnis (2010) conclude that male teacher certainly can have a positive impact on boys’ academic achievement, however, it is not a function of being male.

0.2.2.3 Institutional level

The institutional perspective departs from the idea that contextual settings shape students’ attitude, behaviour and motivation. Hardly any research on gender differential achievement is situated at the institutional level and this is a clear gap in the empirical evidence base. Plenty of questions are still to be answered, like what is the relationship between gendered achievement and classroom and school composition and educational policies implemented by governments, labour market characteristics or (national) gendered cultures. The idea that schools and education as institutions can be part of the dropout problem is quite recent in literature (De Witte & Rogge, 2013). Lamote et al. (2013) criticize that past research on dropout is characterized by methodological shortcomings failing to account for the hierarchical structure of education. They find that school effects are stronger than usually assumed and suggest future research to take this into account. De Witte and Rogge (2013) found that the characteristics of students’ class in the first year of secondary education such as the mean ability of students, the class size and ethnic composition of the class have an impact on later choice to dropout. One topic at the institutions level that has received plenty of research attention is the ‘feminisation’ of schools. It is often proposed that boys’ relatively lower academic achievement is the result of a conflict or mismatch between boys’ and the school’s culture. We do not elaborate too much on this topic in this general introduction since the topic is discussed in the subsequent chapters. Recent studies however do not find much support for the statement that the feminised character of schools has a negative impact on
boys’ performance. Heyder & Kessels (2013) found that boys’ and girls’ academic achievement in math was unrelated to the extent to which they perceive the school as feminine and themselves as masculine. Also, Skelton (2012) critiques the research on feminised schools as being driven by the implicit assumption that feminised schools are deficient. This perception is fed by the distribution of male teachers throughout the educational system, with an increasing number of men in higher levels of education (see Figure 2, OECD, 2012; VLIR, 2015).

![Figure 2. Average percentage of male and female teachers in Flanders, from pre-school to university education.](image)


### 0.3 Research focus

#### 0.3.1 Procrustes programme

This dissertation is part of the larger research programme ‘Teaching in the Bed of Procrustes’. The central research questions of the programme are:

1. What are the gender-related factors that predict low achievement, retention and dropout of youngsters in secondary education?
2. What are the active ingredients of interventions in view of prevention of low achievement, retention and school dropout in boys and girls?
The programme thus aims at a gendered analysis of and intervention in school achievement. This ‘gendered analysis’ is not limited to the study of differences between boys and girls. It also includes investigating the variability that exists between the genders because not all boys are underachievers and not all girls are high achievers. The same variables and mechanisms may hold for explaining underachievement in boys and girls, with boys scoring higher than girls on characteristics that put them at risk. However, it is also possible that explanations for underperformance are different for boys and girls, with boys and girls being vulnerable to different risk factors (Laevers et al., 2011). Three promising pathways that may help explain (and intervene in) intra- and inter-gender differences in academic achievement are identified in the Procrustes programme:

A. Developing positive, broad and integrated identities: an identity-centred approach to gender-differential achievement which focuses on gender related constructs (gender ideology, gender identity) and identity development.

B. Countering anti-school cultures in secondary education and changing teachers’ attitudes and teacher-pupil relationships: aimed at a better comprehension of gender-specific study cultures and gender-specific aspects of teacher-student relationships.

C. Enhancing well-being and involvement through innovative methods as the key to prevention of dropout: explores open art and broad school projects and self-regulated learning activities.

The dissertation is situated within Focus B and is set out to investigate gender-specific aspects of teacher-pupil relationships.

0.3.2 Dissertation project

Based on the theoretical framework we identify six limitations or gaps in the existing research concerning gendered teacher-student relationships and its effects on student achievement. We conclude each limitation with a key research question.

Limitation 1. The supremacy of USA research.
Most previous research on gendered teacher-student classroom interactions has been performed in the USA (Howe & Abedin, 2013). It is regularly
suggested to explore whether results are reproduced in other cultures and nations, especially since significant differences have been observed in the so-called ‘panic’ with regards to boys’ underachievement in different nations (Moureau, 2009). Understanding differences between nations and cultures could help identify relevant macro-level effects in explaining gendered school achievement.

*What do European studies learn with regard to gendered teacher-student classroom interactions and promising pathways to reducing possible teacher gender bias?*

**Limitation 2. Doubt about the existence of teacher gender bias.**

Conflicting research results exist with regard to teacher gender bias. Some studies report no gender differentiated teacher expectations (Dusek & Joseph, 1985; Cooper & Moore, 1995) and suggest that teacher gender bias is overrated, with boys bringing scholastic underachievement upon themselves by being more anti-schoolish and by being more disruptive (Howe, 1997). Other studies find that teacher gender bias does exist (Garrahy, 2001; Younger, Warrington, & Williams, 1999). These studies however do not agree on the extent to which teachers are aware of their gendered thoughts. Garrahy (2001) suggests teachers to be ‘gender-blind’ while Younger et al. (1999) find that the majority of teachers report being aware of treating boys and girls differently. Teacher awareness of gendered interactions deserves special attention, since it is pivotal in reducing negative teacher expectancy effects (Brophy & Good, 1970). The recognition of problematic situations is needed to initiate and drive the reflective process of professional learning (Hoban & Hastings, 2006; Loughran, 2002). Hoban and Hastings (2006) suggest teachers to view their own practice through another lens, in order to stimulate reflection. Videography offers powerful and increasingly popular possibilities to switch perspectives about one’s teaching (Seidel & Strümer, 2014). We suggest video-stimulated recall interviewing as a measurement and intervention instrument to describe implicit and explicit gendered thoughts of teachers and to increase teacher awareness of possible gender bias.
To what extent are teachers’ thoughts, images and emotions that evoke (re)actions to students gendered?
To what extent can video-stimulated recall interviewing raise teachers’ awareness of their possible gender bias?

Limitation 3. Relationship between teacher and student behaviours.
Research on gendered classroom interactions has generated inconsistent findings with regards to several topics. Negative feedback is found to be higher for boys than girls, but it is still unclear which boys (and which girls) account for these increased levels of negative feedback (Bailey, 1993). Howe and Abedin (2013) conclude their review that mostly attainment and ethnicity have been studied alongside gender and that other variables should be investigated. We suggest an intersectionality-informed approach (Collins, 1998) to meet the comments of many authors that classroom interaction patterns should be understood along the lines of more than student gender alone (Beaman et al., 2006). Another inconsistency in previous research is the extent to which boys bring criticism upon themselves due to higher levels of misbehaviour (see previous section). The reason why so few studies have been able to investigate this question is because most studies use teacher-focal coding (e.g. Younger & Warrington, 2002; Jones & Dindia, 2004; Swinson & Harrop, 2009). Teacher-initiated behaviours are registered and it is recorded if they are directed toward boys or girls. Few studies monitor students and their behaviours as well as teachers and their behaviours. We propose a student-focal coding with recording of both student and teacher behaviours so that student and teacher behaviours can be related. Also, we survey all the observed students in order to link student thinking and a wide range of student background variables to the interaction patterns.

Which boys and which girls are receiving higher levels of negative feedback from their teachers?
Are these students bringing negative feedback upon themselves due to higher levels of misbehaviour?
How do students perceive the equity of classroom interactions and what is the relationship to actual classroom interactions?
Limitation 4. Disconnection between gender and professionalism theory.

Several remarks can be made concerning previous research on the ‘feminisation’ of teaching. Research indicates that teacher gender does not significantly affect classroom interactions (Howe, 1997) or boys’ learning (Helbig, 2012). Acker (1990) critiques however that studying gendered professions and organisation is much more than merely describing differences between men and women. The gendering of professions is deeply embedded in the work itself. Many studies on gender differences in professions analyse gender ‘in addition’ to other research questions and do not substantiate their analyses with theories on gendered organisations and professions (Acker, 1990). A disconnection exists between theories on professionalism and theories on gendered work. We will investigate both fields of research and develop a multidisciplinary framework for assessing the feminisation of teaching. Also, the feminization of teaching is mostly studied in primary education (e.g. Skelton, 2012) which should not surprise since the number of female teachers in primary education is higher than in secondary education (OECD, 2012). It might be very interesting however to study gendered teaching in secondary education in particular since secondary education is a crucial phase in the gender identity formation of adolescents, especially since students themselves identify teacher gender as important and influencing their gender identity construction (Skelton et al., 2009). It is striking that observational research claims no teacher gender effects exist, while students themselves report teacher gender does matter. We hypothesise that the gendered character of teaching is subtle and deeply rooted in teacher professionalism. A fine-grained analysis of teacher self-perceptions is suggested.

To what extent is teacher professionalism gendered?
To what extent are female and male teachers’ perceptions of their relationship with students different?
What role does teacher professionalism play in teacher perceptions of teacher-student relationships?
Limitation 5. Mostly descriptive research, lack of intervention research.
One recurrent critique in review works on gendered classroom interactions is that research has predominantly been of a descriptive nature without an attempt at improving teaching practice (Howe, 1997; Beaman et al., 2006). We identified one study of Swinson and Harrop (2005) who trained teachers to increase their amount of positive feedback to pupils. They argue that praise is the key to tackling bad behaviour. This study is not aimed at reducing boys’ anti-schoolish behaviour in specific, but Swinson and Harrop have conducted much research into boys’ and girls’ gendered classroom behaviour. Other authors as well suggest positive feedback and a positive teaching approach to be the keys to tackling (incorrect) negative teacher expectations (Wellborn et al., 2012) and increasing student on-task behaviour (White, 2010). We designed a professional development programme aimed at improving teachers’ feedback patterns to boys and girls by emphasising the power of positive feedback and the importance of information-based differentiated instruction (rather than differentiation based on stereotypes). We assess the effectiveness of the intervention on several levels, as proposed by van Veen, Zwart, Meirink and Verloop (2010): teacher behaviour, student behaviour, students’ perceptions.

What are the effects of a professional development programme aimed at increasing teachers’ feedback toward boys and girls on students’ on-task behaviour and students’ sense of equity?

Limitation 6. Gendered underachievement, which boys and girls?
The analysis of gendered underachievement is not limited to the study of differences between boys and girls. It also includes investigating the variability in predicting underachievement in boys and girls since boys and girls might be vulnerable to different risk factors (Laevers et al., 2011). Intra- and inter-gender differences in grade retention are analysed. A rich set of predictor variables are assessed: initial ability math, initial ability reading, socio-economic status, language background, migration background, sense of equity, on-task behaviour.
How well do student individual background characteristics, students’ sense of equity and student on-task behaviour predict student grade retention? To what extent do prediction models differ for boys and girls?

0.4 Research design

In depth presentations of the methodology are included in the subsequent chapters when introducing the empirical studies. In this section we will provide a brief and general overview of the central methodological choices and the type of data which are collected to address the above described research questions.

0.4.1 Integration of quantitative and qualitative methods

Central within the research programme is a quantitative base of longitudinal data (student and teacher surveys) which has been collected from a representative sample of Flemish schools. One cohort of first year secondary students and first grade teachers has been followed for three years. At the first measurement parents and school leaders were also surveyed. The longitudinal surveying is only one component of the data collection. Several smaller-scale qualitative, action research and quasi-experimental studies were performed parallel to the quantitative baseline research. In this dissertation two cycles of quasi-experimental research were designed to be performed in a purposeful subsample of schools, teachers and their students. Within the dissertation project two cycles of experiments would serve to assess two types of interventions in order to be able to draw conclusions about the effectiveness of an intervention A in comparison to an intervention B (as opposed to statements about the effects of an intervention in comparison to no intervention). The control group in the first cycle would actually be the wait group since they served as the intervention group in the second cycle. The intervention group of the first cycle would then become a retention group.

---

3 The Flemish secondary education system consists out of 3 grades which each count 2 years of education.
Qualitative techniques such as interviews as well as quantitative analysis of survey data and video-recorded classroom interactions were used to assess the effectiveness of the experiment. Effectiveness was assessed by investigating effects at the level of teacher perceptions, teacher behaviours, student perceptions and student behaviours. Figure 3 represents the integration of quantitative data and action research in the Procrustes programme.

Figure 3. Integration of quantitative research base and action research

*Note.* Laevers et al. (2011)

### 0.4.2 Data collection and sampling

Central to the Procrustes programme is the start of secondary education, since national and international research suggests that the gender gap in academic achievement is relatively small or nonexistent in primary education (Laevers et al., 2011). When students enter secondary education the gap manifests itself (more) clearly and widens year by year.
**Procrustes programme**

Data for the large-scale surveying was collected by all researchers involved in the Procrustes programme (a team of seven pre-doc researchers and one post-doc researcher). At the first measurement students, teachers, parents and school leaders were surveyed. The first measurement was performed at the beginning of the first year of secondary education (September/November). A second measurement in teachers and students was performed at the end of that year (April/May) and a third and forth data collection in the same population was performed at the end of the second and third year.

The random school sample consists of 59 secondary school campuses in Flanders and counts over 6000 pupils. Disproportionate stratified sampling was used on the entire Flemish population of secondary school campuses that offer a first grade. The strata that were used are: geographical, rural and urban location and public and private education. Geographical location in Flanders was operationalised by selecting schools in the five Flemish provinces and the capital region of Brussels. Rural and urban location was operationalised by using official urbanisation data (Mérenne-Schuemaker, Van der Haegen, & Van Hecke, 1991). Schools located in cities or municipalities in the most important agglomerations (Category A) and schools located in places with strong morphological and functional urbanisation (Category B) are categorised as “urban”. Schools in places from the other categories (B2-E) are labelled “rural”. A realistic distribution of public and private schools was achieved. Data from the Flemish Ministery of Education and Training for the year 2011-2012 show that 69% of school campuses that offer regular secondary education are private (Vrij Gesubsidieerd Onderwijs) and 31% of school campuses are public schools (Gemeenschapsonderwijs en Officieel Gesubsidieerd Onderwijs) (Vlaams Ministerie van Onderwijs en Vorming, 2012).

The programme aimed at sampling 63 school campuses: 12 in each province and 3 in Brussels. For the provinces 6 rural and 6 urban schools would be sampled, with 4 private and 2 public schools. The schools in Brussels are all urban, with 2 private and 1 public school (see Figure 4).
Figure 4. Disproportional stratified sampling (for five Flemish provinces)

In each final sample strata three random samples were selected and school campuses were contacted and invited to participate in the study. If the school in the first parallel sample would refuse participation the equivalent school in the second and third sample was contacted. In total 124 schools were contacted and 59 schools were sampled (47.6% of contacted schools enrolled in the programme). Schools that refused were often involved in other research projects or were not able to ensure participation of teachers and other school members. Only one school refuse was based on a lack of interest in the topic of the project.

**Dissertation project**

A purposeful heterogeneous subsample of six secondary schools was selected for participation in the quasi-experimental study. Schools were selected from the larger representative sample of 59 secondary school campuses. The six schools were sampled to vary in a range of conditions: public and private schools; urban and suburban areas; small, medium and large-sized schools; general, vocational and technical education schools. Two schools refused to participate; one because the school was already involved in three other large-scale research projects and in the other school the teacher advisory board – after a presentation of the research – decided not to participate due to the high workload that would be involved.

With support from the principal in each school five teachers were selected and asked to participate in the intervention study. Teachers had to be teaching
the first year of secondary education in general subjects and those teachers with most contact hours with the first-year students were addressed first for participation in the study. For each teacher, one of their first-year secondary classes was selected. Classes were selected in such a way that those classes with most contact hours with the teachers involved in the study were selected. In these classes videotaped observations were performed. Teachers were told that any lesson would do for the video-recording unless the entire hour would be devoted to an examination or test. In the one or two cases that the observed lessons involved something other than whole-class instruction, we would revisit the teacher at another time for filming an extra lesson (which resulted in observations of whole-class teaching for all teachers). In total, 30 teachers and their 500 pupils were video-recorded twice (pre-test and post-test at the beginning and end of the first year).

In the second year the initial goal was to observe the same classes. However, we soon learned that in only one third of the cases teachers were teaching the same class in the second year. This was quite a disappointment because we had also taken into account the chances that teachers would teach the same class the next year when sampling teachers and classes. We revised the sampling approach of classes for the second year and decided to observe new first-year classes in order to be able to use the data in a reproduction study for the first measurement. The plan to follow the central cohort of students with observations for two years was not achieved. Since the new cohort of first-year students was not being surveyed, no equally rich analyses could be performed on these data. We decided for this dissertation to focus all our attention on evaluating the effects of the quasi-experimental in the first year. Figure 5 illustrates which data in the end were used for the empirical studies in this dissertation.
0.4.3 Central variables and methods of analysis

In the larger Procrustes programme a wide range of variables is addressed at all levels of the theoretical framework.

In the dissertation project we focus on the dynamics of teacher-student interactions. We summarise the key variables in Table 4 and organise them at four levels: indicators of identity, perceptions, behaviour and achievement.
Table 4. Central variables in the dissertation research

<table>
<thead>
<tr>
<th>Teacher Identity</th>
<th>Student Identity</th>
<th>Perceptions</th>
<th>Teacher Identity</th>
<th>Student Identity</th>
<th>Perceptions</th>
<th>Behaviour</th>
<th>Behaviour</th>
<th>Achievement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender, years of teaching experience, subject taught, type of teaching qualification</td>
<td>Gender, socio-economic background, language background, migration background, initial ability maths, initial ability reading</td>
<td>Teacher reported measures of trust in students, autonomy support, structure provision, extra-role behaviour, experimentation and reflection, keeping up to date with professional literature, implicit and explicit gendered thoughts</td>
<td>Student reported sense of equity in the classroom</td>
<td>Giving students a turn, giving positive feedback, giving negative feedback, providing individual instruction</td>
<td>Calling out to the teacher, raising hands, responding to questions from teachers, unauthorised social interactions with other students (such as talking), being on-task, being off-task</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 6. Conceptual framework of dissertation research focus.

*Note.* PDP = Professional development programme.
Studying teacher-student interactions is especially interesting in the light of valorisation potential and recommendations towards teaching practice, teacher education and continued teacher professional development. Given the complexity of interactions we investigate both perceptions of interactions and their actual observations of interactions. Also we study interactions from the perspective of students and from the perspective of teachers. We not only describe but we also assess the effectiveness of interventions aimed at changing teacher-student interactions. We take into account multiple dimensions of social identity and analyse inter- as well as intra-gender differences and their intersects with other social identity dimensions. We use a wide range of methodologies to answer the different research questions. We aim to combine methodologies in such a way that the strengths of different methods are complementary. For example, to assess actual teacher bias in the provision of negative feedback the study of teacher perceptions alone will not suffice and a detailed analysis of observed student and teacher behaviours is needed.

In the following paragraphs and in Table 5 we provide an overview of the subsequent chapters that present the conducted studies in this dissertation. We recapitulate the limitations in previous research, the research questions that are addressed in the chapters and we summarise the methodologies that are used to answer the questions.

**Chapter 1. Reviewing European Research**

We observe that most previous research was performed in the USA and previous review studies have produced inconsistent findings with regards to the gendered character of positive teacher-student interactions. We wonder whether European studies confirm USA-based research findings or if indications of cultural and institutional differences influencing teacher-student classroom (positive) interactions might be found. Also we aim at detecting promising pathways to not only describe but also intervene in gendered classroom interactions. We opt for a systematic literature review of recent (>1980s) studies performed in Europe. We present an in depth narrative analysis of 18 selected studies.
Chapter 2. Increasing Teacher Awareness
Teacher gender bias is a controversial topic. Conflicting findings have been produced by previous research and depending on the theoretical tradition of the researchers similar results are sometimes interpreted differently (Beaman et al., 2006). Unclarity exists with regards to the extent to which boys are bringing increased levels of negative teacher feedback upon themselves, by being more misbehaved. When studying teacher expectation effects, we first have to assess whether teacher thoughts and beliefs are indeed gendered. Some previous research has measures teacher biased thinking by means of focus groups or semi-structured interview (e.g. Younger et al., 1999). However, Garrahay (2001) concludes that teacher gender bias is unconscious. Bearing this in mind, it is possible that focus groups and regular interviews will not succeed in grasping bias in teacher thinking. We opt for video-stimulated recall interviewing (Schmid, 2011; Tripp & Rich, 2012) in order to analyse differences between explicit and implicit gendered thoughts. Transcripts of the interviews are analysed in two ways: first a content analysis is conducted to compare associations between the noticing of positive and negative student characteristics and behaviour and student gender; second a thematic analysis is conducted to explore whether video-stimulated recall can serve both as tools for data collection and teacher professional development and becoming aware of gender bias.

Chapter 3. Gendered Classroom Interactions
A long tradition of over 40 years of research into gendered teacher-student interactions exists. One recurrent finding is that boys in general receive more negative feedback from their teachers than girls (e.g. Jones & Dindia, 2004; Bailey, 1993; Brophy, 1985b). Two remarks are usually added: not all boys and not only boys receive negative feedback. Despite the fact that this commentary has been formulated for over several decades already, still it is unclear which boys (and which girls) account for the increased levels of negative feedback. In order to assess which boys and which girls receive most negative feedback a student-focal coding is applied in which students are observed through video-taped classroom observations and students are surveyed and tested to measure their initial ability as well. Multiple dimensions of social identity can be related to differential interaction patterns and ability. We register the dyadic interactions between students and teachers
and are able to assess correlations between student (mis)behaviour and teacher feedback. We also investigate the relationship between students’ multidimensional identity and ability and student perception of equity in the classroom environment (which we measure in the survey), and student actual observed classroom interactions.

**Chapter 4. Feminised Teacher Professionalism**

The issue of teachers treating students differently according to their gender raises the question whether teacher gender influences teacher-student interactions as well. Previous research has shown that no teacher gender effects exist on teacher treatment of boys and girls (Howe, 1997) or student achievement (Helbig, 2012). The absence of gender differences, however, does not imply that teaching is not gendered. Research into student perceptions, for example, reports that teacher gender matters in the gender identity formation of students (Skelton et al., 2009). In this study we investigate teacher gendered perceptions of teacher-student relationships. Based on an extensive theoretical framework (drawn from feminist and educational literature) we hypothesise that teacher gender effects are mediated by emotional and rational elements of teacher professionalism. We use teacher survey data to assess the hypothesised model with structural equation modelling.

**Chapter 5. Effectiveness of Teacher Professionalisation**

A clear lack of intervention research was identified in the field of gendered teacher-student interactions. The great majority of previous research is of a descriptive nature (Beaman et al., 2006). Based on the work of Swinson and Harrop (2005) and White (2010) we hypothesise that an appreciative pedagogy can serve to reduce teacher negativity bias and deficiency-based approach towards boys (and girls). We suggest appreciative collaborative teacher inquiry to increase teacher awareness and investigation of good practices with regards to gender differentiated instruction which is information-based rather than stereotype-based. We assess the effectiveness of a one-year professional development programme aimed at increasing teachers’ levels of positive feedback to boys and girls. We adopt a quasi-experimental design with a pre-test and post-test and treatment and control groups. As suggested by van Veen et al. (2010) we assess the effectiveness of
the intervention at several levels: effects on teacher behaviour, student behaviour and student perceptions. Generalised Linear Mixed Modelling (GLMM) is used for the quantitative analysis since it can handle non-normal, nested and time dependent data.

Chapter 6. Predicting Grade Retention

The same variables and mechanisms may hold for explaining underachievement in boys and girls, with boys scoring higher than girls on characteristics that put them at risk. However, it is also possible that explanations for underperformance are different for boys and girls, with boys and girls being vulnerable to different risk factors (Laevers et al., 2011). This study is aimed at investigating the likelihood of students to be grade retained and the effects on this likelihood of student demographic background variables, ability measures, student perceptions of equity and student observed teacher-student interactions. Grade retention was chosen as a measure of early school underperformance since previous research has shown that is predicts later dropout decisions (Lamote et al., 2013). Binomial logistic regression are used to predict the odds that a student is grade retained (C-certificate) or not (A or B-certificate).

In Chapter 7 we finally summarise the results of the six studies and we discuss limitations of the dissertation research, suggestions for further research and recommendations and implications for practitioners and policy makers.
<table>
<thead>
<tr>
<th>Study</th>
<th>Topic</th>
<th>Method</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Reviewing European Research</td>
<td>Systematic review</td>
<td>Narrative analysis</td>
</tr>
<tr>
<td>2</td>
<td>Increasing Teacher Awareness</td>
<td>Video-stimulated recall interview</td>
<td>Content + thematic qualitative analysis</td>
</tr>
<tr>
<td>3</td>
<td>Gendered Classroom Interactions</td>
<td>Mixed methods</td>
<td>Intersectional quantitative analysis</td>
</tr>
<tr>
<td>4</td>
<td>Feminised Teacher Professionalism</td>
<td>Survey</td>
<td>Structural equation modelling</td>
</tr>
<tr>
<td>5</td>
<td>Effectiveness of Teacher Professionalisation</td>
<td>Quasi-experiment</td>
<td>Generalised linear mixed modelling</td>
</tr>
<tr>
<td>6</td>
<td>Predicting Student Achievement</td>
<td>Mixed methods</td>
<td>Binomial logistic regression analysis</td>
</tr>
</tbody>
</table>
Chapter 1
Reviewing European Research
Chapter adapted from:
Abstract

Gender differentiated teacher-student interaction has been the subject of extensive empirical research over the last decades. Through the years several review studies and meta-analyses have tried to recapitulate the existing evidence-base. An examination of eight existing literature reviews reveals that European studies account for only a small minority of the literature reviewed in the past, but are recently increasing in number. European studies coincide with three more general shifts in the research on gendered teacher-student interaction: (1) a shift from the quantity to the quality or type of interactions, (2) studying other variables than student gender alone, (3) a methodological move from predominantly quantitative methods to mixed methods designs. In order to gain insight in the empirical evidence that these movements have produced so far, and to increase our knowledge of European-based research, a systematic review of recent European research is performed. A total of 507 unique citations were found through a systematic electronic search and 18 articles were selected after careful scrutiny for inclusion in the review against pre-determined incorporation criteria. The reviewed studies remain inconclusive concerning the balance between negative and positive teacher-student interactions. Studies however do agree that boys receive more negative interactions. To decrease negative feedback towards boys, appreciative approaches to learning are argued to be a promising perspective in two ways: firstly as an appreciative pedagogical approach for teachers to respond to student behaviours and secondly as a strength-based professionalisation approach for teacher trainers to facilitate teachers' inquiry into their own teaching practices with the aim of increasing teachers’ awareness of gender bias.
1.1 Introduction

Gender differentiated teacher-student interaction has been the subject of extensive empirical research over the last decades. Through the years several review studies and meta-analyses have tried to recapitulate the existing evidence-base (see Table 6).

Table 6. Reviews on gendered teacher-student classroom interaction

<table>
<thead>
<tr>
<th>Year</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>1985</td>
<td>Brophy</td>
</tr>
<tr>
<td>1988</td>
<td>Kelly</td>
</tr>
<tr>
<td>1992</td>
<td>Olivares &amp; Rosenthal</td>
</tr>
<tr>
<td>1993</td>
<td>Bailey</td>
</tr>
<tr>
<td>1997</td>
<td>Howe</td>
</tr>
<tr>
<td>2004</td>
<td>Jones &amp; Dindia</td>
</tr>
<tr>
<td>2006</td>
<td>Beaman, Wheldall, &amp; Kemp</td>
</tr>
<tr>
<td>2013</td>
<td>Howe &amp; Abedin</td>
</tr>
</tbody>
</table>

An examination of existing reviews reveals that European studies account for only a small minority of the literature reviewed in the past, and European studies are best represented in the most recent literature. The increase of European studies coincides with three more general shifts in the research on gendered teacher-student interaction. First, the research scope has deepened by shifting the focus from the quantity to the quality or type of interactions. Second, the research scope has widened by expanding the focus to other variables than student gender alone. Third, this deepening and widening of the focus has translated into a methodological move from predominantly quantitative methods to mixed methods designs. We will elaborate on each of these shifts in the following paragraphs.
1.1.1 From quantity to quality

Research from the 1960s to 1990s reveals that teachers interact more with boys than with girls. Kelly (1988) describes that this overattention for boys is not always large but it is consistent, with none of the 81 studies involved in the meta-analysis showing girls to interact more with their teacher than boys. Bailey (1993) and Olivares and Rosenthal (1993) report similar findings of boys participating more in classroom interactions and teachers directing more overall attention to boys. Both reviews conclude that boys have the advantage in the interactions with teachers. Olivares points to the implicit lessons students are learning as a result of the imbalanced teacher attention: “Boys seem to be more valued in the class and girls and boys pick up that message. (...) The school environment seems to do a good job in perpetuating those [stereotypic gender] schemes.” (Olivares & Rosenthal, 1993, p. 13). Bailey (1993) mentions that teacher attention is not necessarily evenly distributed among all boys in the classroom and she refers to Sadker and Sadker (1985) who reported that a few “star” male students receive the majority of teacher attention and that one or two students often account for one fifth of teacher attention. Not all researchers, however, have agreed with such findings concerning boys’ advantages. Beaman et al. (2006) very successfully illustrate in their review the fierce debate between feminist researchers interpreting results from a “power and male domination” perspective and other researchers interpreting (sometimes similar) results from a “boys’ problem and underperformance” perspective. Galton, Hargreaves, Comber, Wall and Pell in 1999, for example, based on the large-scale 1976 UK ORACLE (Observation Research and Classroom Learning Evaluation) data, support that teachers do not favour one gender more than the other. Also, they heavily criticise other research by claiming that the adopted observation schedules are susceptible to bias of the researchers.

Research in the 1990s and onwards somewhat leave to rest the discussion about the quantity of teacher attention and continue research with a more fine-grained approach. The central focus of research is on the quality and type of interactions that are being addressed to boys and girls. One of the recurrent findings is that boys might receive more teacher attention in general, but it is mostly negative attention such as teacher criticism and reprimands (Howe, 1997; Beaman et al., 2006; Jones & Dindia, 2004) and thus maybe not such an ‘advantage’ after all. With regards to positive
interactions the reviews are contradictory. Howe in 1997 concluded that research on positive teacher feedback towards boys and girls is inconclusive. In 2004 Jones and Dindia find in their meta-analysis that boys and girls receive equal amounts of positive feedback than girls. They do note that the number of studies reporting positive teacher feedback was relatively small making it more difficult to achieve significant effects sizes. In 2013 Howe and Abedin conclude that boys occasionally receive more positive feedback than girls (Howe & Abedin, 2013).

1.1.2 Other variables than student gender alone

The move towards more fine-grained research from the 1990s onwards also includes a shift towards investigating the effects of other variables than student gender alone. A recurrent suggestion for further research in the reviews is to study which boys and which girls in specific are receiving more or less teacher attention. Howe and Abedin in 2013 conclude their review on classroom dialogue that the variables that have being studied in addition to gender are attainment, ethnicity and socio-economic status. The findings concerning the effects of these variables are, however, inclucive. Beaman et al. (2006) and Jones and Dindia (2004) also are unable to present conclusive findings concerning the effects of attainment, ethnicity and socio-economic status. They urge future research to continue investigating the interactions of gender with other student background characteristics.

1.1.3 Methodological shifts

From the 1960s to the 1990s several sophisticated coding systems and observation schedules have been developed to register and describe gender differentiated classroom interactions. The Brophy-Good Dyadic Child Interaction System (Brophy & Good, 1969), the INTERSECT system (Sadker, Sadker & Bauchner, 1984) and the Flanders Interaction Analysis Categories System (FIAC)(Flanders, 1970) have come to be well-known and widely used (in adapted form) by classroom interaction researchers in those years. From the 1990s on we see a drop in the use of these systems. Self-developed smaller and more focused observation schedules have been applied
to meet the need for more fine-grained research. Also, we see an increase in
the used of mixed methods research. A triangulation of methods is used to
gather a rich understanding of classroom interactions by means of not only
observations, but also interviews with students and teachers and the analysis
of curricular materials. This mixed methods approach meets Brophy’s
(1985b) framework for studying teacher expectancy effects which
emphasises the role of student perceptions and teacher beliefs as reinforcing
or opposing to negative self-fulfilling prophecies.

1.2 Positive interactions underexposed

Jones and Dindia (2004) mention that positive teacher feedback has
received relatively little attention in previous research. Other reviews too
illustrate the scarce attention of research for positive interactions in
comparison to the study of misbehaviour and teacher criticism. This is
unfortunate since some researchers point to the great importance of positive
classroom interactions for student learning (Swinson & Harrop, 2005). In
their methodological review of behavioural research Harrop and Swinson
(2007) state that research into positive feedback has yielded contradictory
findings due to differences in operationalisation. Some studies include neutral
and affirmative feedback such as ‘yes’ and ‘okay’ to the positive category
while other studies group these utterances in a separate neutral category. We
add to this explanation of Harrop and Swinson (2007) that a lack of attention
for positive interactions is also the consequence of the negativity bias in
much educational research, with scholars being more interested in whatever
is going wrong rather than investigating things that are going well. This is a
critique which has gained attention since the 2000s, with a movement of
scholars identifying themselves as studying the field of “positive psychology”
(Seligman et al., 2005; Hunter & Csikzentmihalyi, 2003). In 2000 the
American Psychologist devoted a millenium issue to the upcoming science
(cfr. Seligman & Csikzentmihalyi, 2000). Positive psychology is a term used
to address the study of positive emotions, positive personality traits and
positive institutions. Researchers use the same methods that have proven to
be effective in research of mental illnesses and mental health or wellbeing.
The research outcomes are intended to have an added value and do not aim to
replace the existing knowledge on human suffering, weaknesses and
disorders. The intent is to provide a fuller and more balanced scientific understanding of human experiences – the ups and the downs and everything in between. This positive approach (strength-based, appreciative approach) can be distinguished from deficiency-based and negative approaches that depart from problems. The appreciative approach studies the best of a phenomenon and states that change has more chance to succeed if attention is vested on the goals rather than on the problems. Appreciative Pedagogy (Yballe & O’Connor, 2000) was launched as a term to indicate appreciative approaches to learning (e.g. Yballe & O’Conner, 2000; Doveston & Keenaghan, 2006; Morsillo & Fisher, 2007; Giles & Alderson, 2008; Steyn, 2009; Kozik, Cooney, Vinciguerra, Gradel, & Black; 2009; Eow, Wan Zah, Rosnaini, & Roselan, 2010a, 2010b; Lehner & Hight, 2006; Seligman, Ernst, Gillham, Reivich, & Linkins, 2009). This self-labelling, however, gives the wrong impression of appreciative pedagogy being a new trend. Other theories have elaborated on the importance of positive and appreciative approach to foster learning as well: flow theory of Csikzentmihalyi (1990), the self-determination theory of Ryan and Deci (2000) or the U-theory of Senge, Scharmer, Jaworsky and Flowers (2004a, 2004b). A recently published book of Swinson and Harrop (2012) titled Positive Psychology for teachers too addresses the opportunities offered by positive psychology and appreciative approaches to learning for classroom interactions. We will review the literature with a special focus on positive interactions.

1.3 Research focus

Three major shifts have occurred in the field of gendered teacher-student classroom interaction research over the past decades. These shifts have taken place simultaneously to the rise of European interest into the topic of gendered interactions. These European studies, however, have not always received much attention in previous reviews since some of these reviews have only focused on reviewing research conducted in the USA (e.g. Jones & Dindia, 2004). The focus of this study will be to review recent European research holding the following research questions in mind: Which boys and which girls are receiving more or less positive and negative interactions? Which are the most interesting methodological suggestion for future research? Literature search
A systematic review approach was applied, which implies that the literature was searched in a systematic and transparent way (Petticrew & Roberts, 2006). The methodology of this systematic review includes the following procedure: (a) stating objectives and review questions, (b) identifying incorporation criteria, (c) literature search, (d) screening of studies, (e) analysis of data, and (f) synthesizing (Andrews & Harlen, 2006). In the literature search articles were found through an electronic search of the ERIC and Web of Knowledge databases. All relevant combinations of the following keywords were used for searching the databases: teacher-student interaction, student-teacher interaction, classroom interaction, or teacher feedback, in combination with gender or sex. In addition, the bibliographies of relevant articles were browsed to locate further articles of interest. The electronic search was undertaken in May 2012 and identified a total of 507 unique citations. All titles and abstracts were carefully screened against predetermined incorporation criteria. All articles had to: contain observational data for teacher-student interactions (either qualitative or quantitative); be published in a scholarly peer-reviewed journal after 1980; record demographic information of student gender; and include data collected in a European pedagogical setting in primary or secondary education. The reasons for exclusion were recorded and full texts were obtained for those articles that met (or could potentially meet) all of the incorporation criteria. After screening of the full texts 18 articles were left for review. It is striking that up to one third of all studies were excluded due to the fact that they concerned research conducted in the USA. Another considerable part of the studies did not qualify as the research did not include observation of teacher-student interactions but rather for example observation of peer-to-peer student interactions or questionnaires measuring teacher or student perceptions of interaction. The following information was extracted from the articles: country of data collection, school level (elementary, age 6-11; secondary, age 14-18), student sample size and gender distribution, teacher sample size and gender distribution, type of data analysis (quantitative, qualitative, mixed methods. Table 7 provides an overview of the 18 included studies and summaries of their methodologies and key results.
Table 7. Summary of studies included in the systematic literature review

<table>
<thead>
<tr>
<th>First author, year</th>
<th>Country</th>
<th>Population</th>
<th>Sample size</th>
<th>Method</th>
<th>Brief description of study</th>
<th>Brief description of results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cervoni 2011</td>
<td>UK &amp; USA</td>
<td>Primary</td>
<td>TF = 4, SM = 28, SF = 18</td>
<td>QL</td>
<td>Three consecutive year 3 (age 7-8) science lessons taught by four female teachers in four middle-class suburban (of which one in the USA) and working-class city coeducational schools were observed (twelve lessons in total). Verbal and non-verbal interaction were analysed to identify teachers’ choice of topic, use of artefacts and instructional discourse. Additionally, the four teachers and a sample of twelve students of each teacher were interviewed individually.</td>
<td>Boys in middle-class suburban schools are in some cases dominated by girls, with boys responding by becoming silent or expressing frustration. A pedagogy that provides students with high degrees of autonomy appear to open up possibilities for girls to engage in science and associate science with femininity.</td>
</tr>
<tr>
<td>Croll, 1985</td>
<td>UK</td>
<td>Secondary</td>
<td>TM=9, TF=25, SM=129, SF=97</td>
<td>QT</td>
<td>Thirty-four second year junior classrooms in twenty schools we observed. A total of two hours of interactions for each of one hundred twenty-nine control students were coded using a systematic observation schedule for individual interaction with the teacher.</td>
<td>Students that are seen by their teachers as having learning and behavioural problems receive significantly more individual attention than do other students in the class. In general boys receive a slightly higher average level of individual attention from their teachers. This higher average can be explained by a higher number of boys being identified as having learning and behavioural problems. Among pupils not having special educational needs there is still a higher average level of individual attention found for boys. This difference is accounted for by a small number of boys receiving very high levels of attention rather than a uniform tendency for most boys receiving more attention than most girls.</td>
</tr>
<tr>
<td>Drudy, 2002</td>
<td>Ireland</td>
<td>Secondary</td>
<td>T=136, SM=1715, SF=1502</td>
<td>MM</td>
<td>One hundred thirty-six modern languages, mathematics, science and humanities lessons in a variety of coeducational school types and locations were self-analysed by their teachers. Verbal interaction was coded using an adaptation of the FIAC observation schedule for the following categories: praise, acceptance, higher-order questions, lower-order questions, criticisms, lower-order answers, higher-order answers, and pupil initiations. Intra-rater reliability of minimum 0.90 was obtained. Additionally, qualitative responses of the teachers to the interaction patterns which they had observed in their own classrooms were obtained.</td>
<td>The predominant pattern was one of overall male domination of the various categories of interaction. Classroom gender composition was significantly related to the patterns of interaction. Both boys and girls participated in more interactions where their own sex was in the majority. The most equitable distributions were the classes that were either evenly balanced or where girls were somewhat in the majority. Student teachers attributed interaction patterns to the behaviour of male pupils rather than to aspects of their own behaviour or behaviour of female pupils. Many teachers expressed shock at the uneven patterns in their classrooms.</td>
</tr>
<tr>
<td>Einarsson, 2002</td>
<td>Sweden</td>
<td>Primary</td>
<td>T=40, S=n.r.</td>
<td>QT</td>
<td>Twenty intermetiate level (age 10-12) and twenty upper level (age 13-16) mathematics and mother tongue lessons were observed. Verbal interaction was coded for the following categories: teacher initiated, and student-initiated. Inter-rater reliability in a pilot study of 0.74 was obtained.</td>
<td>Boys in general have significantly more frequent interaction with their teachers than girls. This pattern is not valid for male teachers at the upper level. Male teachers increase the attention paid to girls, compared to boys, as they grow older. Female teachers always give more attention to boys. Boys initiate more interactions with female teachers than with male teachers at both levels, compared to girls.</td>
</tr>
</tbody>
</table>
French, 1984
Country: UK
Population: Primary
Sample size: T=1
SF=13
SM=16
Method: MM
One lesson was observed. Verbal interaction was coded for the following categories: turns taken by teacher, turns taken by pupils as chorus, turns taken by unidentified pupils, turns taken by boys, and turns taken by girls. Additionally, the interaction turns were broken down in detail and analysed qualitatively.

The research highlights the part played by students themselves in achieving gender-based imbalances in favor of boys. Boys generally monopolize the interaction in the lesson observed. The imbalance is due to a small subset of boys taking a disproportionately high number of turns. The data give the impression that students are actively seeking attention from their teachers for example by taking up unusual positions on issues of classroom discussion.

Harrop, 2011
Country: UK
Population: Primary
Sample size: T=38
S=n.r.
Method: QT
Thirty-eight classes in coeducational schools were observed. Teacher verbal interaction was coded for the following categories: questioning, instructions and redirections, approval for academic behaviour, disapproval for academic behaviour, approval for social behaviour, disapproval for social behaviour. Additionally, student on-task behaviour was coded. Inter-observer reliability of minimum 0.90 was obtained.

The primary school pattern of boys receiving both more approval and disapproval than the girls has not been repeated in the secondary school. In the secondary sample little difference was found between approval (to academic and social behaviour) to boys and girls. As for disapproval (academic and social) boys received more than girls. Little difference between the on-task behaviour of boys and girls was found in secondary schools whereas in primary schools the girls were considerably more on-task than the boys. Boys in both primary and secondary school however showed a significantly different pattern of off-task behaviour from that of girls. More girls were never off-task than were off-task at least once, whereas for boys the reverse is true.

Jungwirth, 1991
Country: Austria
Population: Secondary
Sample size: T=11
S=n.r.
Method: QL
Three to five successive mathematics lessons in eleven classes in coeducational schools were observed. Verbal and non-verbal interaction were analysed to identify gender-specific classroom management strategies.

Several types of gender-specific practices are identified that teachers use only in (certain) interactions with girls respectively with boys. Teacher gender does not matter. The more ambiguous a question the more students have to rely on trial and error, the less girls actively participate in the interaction. Teachers contribute to the concealment of boys’ failures as little mistakes or sudden lack of concentration. Within the interaction of girls however failure emerges and becomes bigger and bigger during the interaction. Teachers reject boys’ answers that are correct but unwelcome by negotiating towards the desired answer whilst with girls the teacher defends his/her answer as valid based on authority.

Merrett, 1992
Country: UK
Population: Primary
Sample size: T=70
S=n.r.
Method: QT
Thirty-two primary and thirty-eight secondary teachers in coeducational schools were each observed on three occasions. Verbal interaction was coded using a modified OPTIC observation schedule for the following categories: approval for academic behaviour, disapproval for academic behaviour, approval for social behaviour, disapproval for social behaviour. Additionally, students on-task behaviour was coded. Inter-observer reliability of minimum 0.90 was obtained.

For the primary sample there were virtually no differences between teachers’ responses to boys and girls. The mean on-task behaviour levels of boys and girls revealed no significant differences. Data from the sample of secondary teachers reveal that teachers gave significantly more positive responses to boys for their academic behaviour but not for social behaviour. Furthermore they gave significantly more negative responses to boys for both academic and social behaviours. Levels of on-task behaviour for boys and girls were shown to be very similar independent of teacher gender. Male teachers responded significantly
<table>
<thead>
<tr>
<th>Country</th>
<th>Population</th>
<th>Sample Size</th>
<th>Method</th>
<th>Sample</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>UK</td>
<td>Primary</td>
<td>T=36, S=144</td>
<td>MM</td>
<td>Thirty-six classes in total in years 1, 4, 5, 8, 9, and 10 in twelve primary, three middle and one high school (rural and urban, middle class and advantaged intake) were observed. Verbal and non-verbal interaction during whole class episodes were coded for the following categories: ask question, answer question after invitation, call out task related, call out task unrelated, joining in collective response, talk to neighbour, put hand up, off-task. Individual, pair or group work episodes were coded for the following categories: uncooperative response, request help, ask question, challenge a point, talk to self, watch someone else, talk to neighbour, off-task, work independently. Additionally, the achievement level of the students was noted and qualitative field notes were written during the observation. Furthermore, the teachers and a sample of four students (high achieving boy and girl, underachieving boy and girl) of each teacher were interviewed individually. The tests for statistical significance identify stronger statistical significance for achievement, than for gender. Underachievers, boys and girls alike, are the reluctant participators. High achievers dominate in positive learning interactions and underachievers dominate the more negative classroom interactions. Underachievers are more likely to be engaged in off-task interactions than high achievers. In early years off-task interactions are unrelated to gender. High achieving boys however tend to join the pattern of underachievers towards middle school. High achieving boys transform their pattern of interaction with increasing age and tend to become more like underachievers. High achieving girls remain consistently engaged in positive interactions and consistently on-task. Qualitative analyses show that high achieving boys not appearing too eager or enthusiastic in class can be explained by their construction of masculinity and peer group pressure.</td>
<td></td>
</tr>
<tr>
<td>France</td>
<td>Secondary</td>
<td>T=2, SM=42, SF=28</td>
<td>QT</td>
<td>Five coeducational physical education lessons (badminton as gender-neutral, circuit weight training as masculine) in each of two classes were observed. Direct verbal interaction between teacher and individual students was coded for 10 categories of teacher behaviour in two major dimensions: teacher reactive behaviours in response to students' performance, and spontaneous behaviours. Mean intra-coder reliability of 0.80 was obtained. Mean inter-coder reliability of 0.85 was obtained. Additionally, teachers' estimated levels of activity-specific skill of students during the observed lessons were noted.</td>
<td></td>
</tr>
<tr>
<td>Sweden</td>
<td>Secondary</td>
<td>T=15, SM=17, SF=34, SF=42</td>
<td>QL</td>
<td>Three grade 9 (age 15-16) classes were observed. Verbal and non-verbal interaction were analysed to identify strategies that girls use to enhance their influence and control in the classroom. Additionally, a sample of students and their teachers were interviewed individually. Observations showed that boys have a more prominent position in class. Interviews revealed girls however develop strategies (scripts) to increase their influence and control. Some scripts draw on apparent accommodation, i.e. to exaggerate a 'good pupil' script in order to get control. Girls as well as teachers consider girls more likely to get away with minor rule-breaking. Male teachers appear to be more susceptible to these scripts. Other scripts draw on confrontation. 'Personal' reactions of girls can make teachers feel uncomfortable. Teachers indicate conflicts with girls to be unpredictable and they can spread through girls' social networks. Working class girls are more likely to use confrontational strategies than middle-class girls.</td>
<td></td>
</tr>
</tbody>
</table>
Smith, 2007
Country: UK
Population: Primary
Sample size: T=12
SM=18
SF=n.r.
Method: QT

One hundred forty-four year 5 (age 9-10) literacy and numeracy lessons were observed live. Verbal interaction in all whole class sections of a lesson were coded using the computerized CIS observation schedule for: actor and receiver (teacher, whole class, boy, girl, other, whiteboard), type of discourse move (e.g. question, answer, evaluation), and additional information on discourse move (procedural, curricular, praise, criticism, acceptance, refocus, direct, probe). The CIS had intra-rater reliability of 0.78 and inter-rater reliability of 0.86.

Discourse moves are higher in frequency with boys than with girls. No significant differences between whiteboard lessons and non-whiteboard lessons were found for the first year of use. Open questions to boys, their answers and evaluation of their answers were significantly higher in the second year of interactive whiteboard use. Whiteboards seem to exacerbate the imbalance in classroom interaction. As the percentage of boys in class increases, so does the total number of discourse moves per hour with boys.

Swann, 1988
Country: UK
Population: Primary
Sample size: T=2
SM=7
SF=7
Method: MM

Two sequences of talk between sciences teachers and a small group of primary students (age 10-11 and age 9-10) in different schools were observed. Verbal interaction was coded for the following categories: total words spoken, total turns, total interchanges, silent turns, failed turns. Additionally, the interaction was analysed qualitatively.

Departing from the effectiveness of small group discussions and teacher directed discussion for science learning, the authors suggest that a systematic exclusion of girls from this interaction is denying them learning opportunities. Not all talk that was observed however is of value for learning. Nevertheless are girls excluded from valuable talk. Questions addressed to girls were less frequently the challenging and open variety typically addressed to the boys and more often rhetorical or yes/no questions. Secondly a hidden curriculum is at work as boys are acquiring and practising skills in competitive public speaking.

Swinson, 2009
Country: UK
Population: Primary
Sample size: T=18
SM=166
SF=178
Method: QT

Eighteen classes from five coeducational schools were observed. Teacher verbal interaction was coded for the following categories: questioning, instructions and redirection, approval for academic behaviour, disapproval for academic behaviour, approval for social behaviour, disapproval for social behaviour. Additionally, student on-task behaviour was coded. Mean inter-observer reliability of over 0.90 was obtained.

The most statistically significant difference was that boys received more instruction and redirection than girls. Teachers also directed significantly more approval for academic behaviour and more disapproval for social behaviour to the boys than to the girls. Boys were less on-task than the girls and their behaviour was more variable than girls'. The higher level of off-task behaviour of boys was not a consequence of the behaviour of a few boys. The ratio of approval to disapproval for boys was 85% of that of the girls. The boys' on-task behaviour was 92% of that of the girls.

Taino, 2011
Country: Finland
Population: Secondary
Sample size: T=28
S=n.r.
Method: QL

Twenty-eight lessons of different subjects were observed. Verbal and non-verbal interaction were analysed to identify the use of gendered address terms.

In classroom management teachers can reproach students without embarrassing the participants by using humour or by addressing the reproach to a group of students in stead of an individual student, such as students, guys, boys, girls. Using gendered address terms cannot be considered frequent. When used, the gendered address term boys was much more frequent in the data than girls. Students however resisted the use of this practice in various ways such as imitations to tease or sometimes even criticize the teachers' activities.
Tsouroufli, 2002
Country: Greece
Population: Secondary
Sample size: n.r.
Method: MM

Five teachers in different subjects in one small urban school and one class (age 14) which was taught by all the teachers was observed for forty-six hours. Verbal interaction was coded for the following categories: ask questions, assign tasks, address with first name, negative comment for behaviour, negative comment for work. Additionally, documents were collected where the teachers' behaviour was observed. Furthermore, documents were collected where the principal of the school and oral stories were collected to explore possible explanations for teachers' behaviour by looking at teachers' ideas about gender and their life experiences and gender constructions in that school.

Younger, 1999
Country: UK
Population: Secondary
Sample size: n.r.
Method: MM

Year 11 (age 15-16) classes across subjects and ability levels in two selective and two comprehensive schools were observed. Quantitative systematic observations and qualitative ethnographic note-taking focused on: type of teacher to student and student to teacher questions, requests for help, reprimands, praise, student work behaviour, and teaching approach. Additionally, the perspectives of 200 year 11 students in four selective and four comprehensive schools were gained through 48 focusgroups in single-sex and ability groupings of four. Furthermore, teachers were interviewed individually.

Younger, 2002
Country: UK
Population: Secondary
Sample size: n.r.
Method: MM

Two to six single-sex and mixed-sex year 9 and 10 (age 14-16) lessons of thirteen mathematics and geography teachers in one school were observed. Quantitative systematic observations and qualitative ethnographic note-taking focused on: type of teacher to student and student to teacher questions, requests for help, reprimands, praise, student work behaviour, and teaching approach. Additionally, achievement levels in the school at 16+ over the period 1988-99 were analysed.

In all single-sex classes teachers tended to ask more questions in boys' classes than in girls' classes. In mixed-sex geography classes girls asked twice as many questions as boys. In single-sex classes students' requests for help showed little variation according to gender. In mixed-sex geography classes girls made almost twice as many requests as boys. The majority of teachers did not explicitly adjust their teaching style to single-sex classrooms. Some did and in these cases a cooperative and collaborative ethos was more often present in girls' lessons. In the mixed classes boys were working steadily and quietly, without really participating.
1.4 Results

1.4.1 Descriptive results

A screening of the publication years of the included studies shows an increase of publications over time: three studies date back to the 1980s, four were published between 1990 and 2000 and eleven were published after 2000. With regard to the countries of data collection of the included studies an overrepresentation of UK studies can be noticed: eleven out of eighteen studies discuss research performed in the UK. As a consequence, overgeneralisation should be avoided. Three of the UK studies account for the four studies published before 1990, which indicates the pioneering role of UK research on the topic in Europe. Other countries of data collection are Ireland (1), Sweden (2), Austria (1), France (1), Finland (1), and Greece (1). It should be noted that this sample reflects research in Western European classrooms. As to the methodologies used an increase can be observed in the combination of qualitative and quantitative analyses in mixed methods towards the 2000s. In total seven studies were coded as using mixed methods, seven as using quantitative analyses and four as using qualitative analyses.

1.4.2 Negative teacher-student interaction

Of the eighteen reviewed studies, ten studies present evidence of positive and negative interactions between teachers and boys and girls (Drudy & Uí Chatháin, 2002; Harrop & Swinson, 2011; Merret & Wheldall, 1992; Myhill, 2002; Nicaise, Cogérino, Fairclough, Bois, & Davis, 2007; Öhrn, 1993; Smith, Hardman & Higgins, 2007; Swinson & Harrop, 2009; Tsouroufli, 2002; Younger et al., 1999). All ten support the findings of previous and non-European studies that boys are involved in more negative interactions than girls (e.g. criticism, disapproving comments, organization and misbehaviour feedback, reprimands, task-related and task unrelated calling out). The studies of Harrop and Swinson (2011) and Merret and Wheldall (1992) differentiate between disapproval for academic and disapproval for social behaviour and indicate that teachers in respectively primary and secondary education give significantly more negative responses to boys for both categories. A commonly cited hypothesis is that a small number of boys
account for this imbalance rather than a uniform tendency for most boys to receive more negative attention than most girls. The 2009 study of Swinson and Harrop explicitly checks for this possibility but does not provide any evidence for the hypothesis. Interviews with students and qualitative observations in the studies of Younger et al. (1999) and Öhrn (1993) reveal that boys and girls are well aware of the reported inequalities and suggest that girls avoid negative interactions and get away with minor rule-breaking by engaging in 'good pupil' behaviour and by flattering and placating the teacher. Myhill (2002) argues that underachieving boys, as well as high achieving boys once they reach puberty and adolescence, not only engage less in 'good pupil' behaviour but also actively cultivate a 'cool' anti-school attitude as a response to constructions of masculinity and peer group pressure.

1.4.3 Positive teacher-student interaction

The ten studies that present evidence of positive and negative interactions have produced contradictory results concerning the division of positive teacher-student interactions towards boys and girls. A number of studies conclude that boys are involved in more positive (praise, encouragement, reinforcement, approving comments) as well as negative interactions than girls (Drudy & Uí Chatáin, 2002; Harrop & Swinson, 2011; Merret & Wheldall, 1992; Smith et al., 2007; Swinson & Harrop, 2009; Younger et al., 1999). Other studies report the opposite: girls receiving more positive attention than boys (Myhill, 2002; Nicaise, 2007; Öhrn, 1993; Tsouroufli, 2002). Findings are also inconsistent with regard to differences in primary and secondary education or variations in approval for academic and social behaviour. A large part of this variation can probably be explained by methodological differences such as the definition of the observational categories mentioned by Harrop and Swinson in their methodological review (2007) or more general variations in methodology such as “how many times a single class is observed, how many classes are observed overall, whether coders are trained, whether interactions are coded from a videotape, whether there is volunteer bias reflecting the kinds of teachers who agree to be observed, and so forth” (Jones & Dindia, 2004, p. 457). Another aspect that is suggested to deserve further investigation is the correlation of interaction patterns and the subject matter taught. Tsouroufli (2002) and Nicaise (2007)
argue that boys received more praise and encouragement than girls when involved in male-traditional activities such as respectively circuit weight training and mathematics in contrast to reversed patterns of negative and positive interactions in gender-neutral or feminine-traditional activities.

1.4.4 Ratio of positive to negative teacher-student interaction

Despite conflicting results concerning positive interactions, most of the reviewed studies seem to agree on one thing: that the levels of praise are very low in comparison to the total interactions. Drudy and Uí Chathaíin (2002, p. 43) articulate their findings as follows “the praise and reinforcement category which helps to develop a positive classroom climate and to reinforce pupil success, was comparatively low at 8% of the total interactions”. Apart from the comparison of positive interactions to the total interactions another comparison that is put forward is the relationship between positive and negative interactions. In a study that is not included in the review (no results on student gender differences) Wheldall, Houghton and Merrett (1989) suggest to assess the overall effect of positive and negative teacher feedback in the form of a ratio of positive feedback divided by negative feedback. In all of the reviewed studies with quantitative data for positive and negative interactions the ratio for boys is lower than that for girls and the reported ratios vary from 1:>1 (more negative feedback than positive feedback) (Merrett & Wheldall, 1992) to 3:1 (Drudy & Uí Chatáin, 2002). In a systematic review of praise to reprimand ratio recommendations for classroom management White (2010) found very few peer-reviewed articles that specifically address the ratios as a key term. The recommended ratios White found range from 3:1 tot 10:1 for optimal promotion of appropriate student behaviours. In order to assess such relationships, research on teacher-student interaction has added measures of student behaviour to the observation schedules. Harrop and Swinson (2011), Merrett and Wheldall (1992) and Swinson and Harrop (2009) included a measure of pupil on-task behaviour. Myhill (2002) further developed the measurement of pupil off-task behaviour and distinguishes between more specific categories such as talking to neighbour or calling out task unrelated. Correlational analyses suggest that the higher the ratio of positive to negative interactions the more pupil on-task behaviour is displayed. White did not find any experimental
investigations where praise to reprimand ratios were directly manipulated to prove a causal relationship. Swinson and Harrop (2005), however, did publish a paper concerning the topic. They trained teachers to increase their amount of positive feedback to pupils. They found that the levels of on-task behaviour in students increased from 77 to 94 percent when positive feedback was three times more frequent than negative feedback.

1.4.5 Methodological shifts

An examination of the research designs of the reviewed studies reveals that beside quantitative analysis of observational data, interaction patterns have been studied using the following set of methods: qualitative analysis of observational data (Cervoni & Ivinson, 2011; Öhrn, 1993; Taino, 2011; Tsouroufli, 2002; Swann & Graddol, 1988; French & French, 1984), qualitative field notes on observations (Drudy & Úi Chatháin, 2002; Myhill, 2002), student and/or teacher interviews (Cervoni & Ivinson, 2011; Myhill, 2002; Öhrn, 1993; Tsouroufli, 2002; Younger et al., 1999), student achievement ratings (Myhill, 2002; Nicaise et al., 2007; Younger & Warrington, 2002), teacher questionnaires (Tsouroufli, 2002), and analysis of school policy documents (Tsouroufli, 2002).

This divergence of methods makes comparison and interpretation of results more complex but on the other hand allows for lifting the vagueness of too much generalisation. It is tempting to synthesise a study with general findings in terms of 'on the whole' or 'overall', but in doing this nuances are disregarded. Mixed methods research can deepen understanding and provide insights that would be missing when using only a single method (Johnson & Onwuegbuzie, 2004). This advantage is well illustrated in the study of Younger et al. (1999) that combines quantitative systematic observations, qualitative ethnographic note-taking, focusgroup interviews with students, and individual interviews with teachers. The study successfully both describes the quantity of different types of student-teacher interaction as well as gives significance to the identified interaction pattern by analysing teacher and student perceptions.

The reviewed studies do not only describe but also investigate correlations between interactions and variables. Correlations between student on-task
behaviour and interaction patterns have already been mentioned in the previous section of this chapter. Another relationship that has been examined is the (mediating effect) of student achievement level on gender differentiated interaction patterns (Myhill, 2002; Nicaise et al., 2007; Younger & Warrington, 2002). Myhill (2002) found achievement level to be of a stronger statistical significance for levels of interaction than gender. Younger and Warrington (2002) link their research on classroom interactions to the apparent problem of underachieving boys in Western countries that has yielded much concern amongst researchers and policy-makers over the last two decades.

Also, relationships between teaching practice and student behaviours have been studied. Two qualitative studies on this subject are worth mentioning because of the concrete recommendations for teaching practice that could be abstracted from them. The findings of Jungwirth (1991) reveal that teachers can stimulate classroom participation of girls by avoiding ambiguous questions to which the typical reaction of girls is to keep silent as they do not like to rely on trial and error to answer. Another teaching approach that has been found to stimulate girls participation (in science classrooms) is allowing students considerable autonomy (Cervoni & Ivinson, 2011).

Two of the reviewed studies have moved beyond descriptive and correlational research and have set up quasi-experimental research with the aim of manipulating unequal interaction patterns. The quantitative research of Smith et al. (2007) have investigated the effect of interactive whiteboards (IWB) for promoting interactivity and the mixed methods. The IWBs seemed to enlarge the imbalance in classroom interaction and with this the dominance of boys.

The study of Drudy and Úi Chatháin (2002) examined the influence of student teachers' systematic self-analysis on their awareness of and reflection on uneven interaction patterns. The study consisted of an action research project among student teachers, with the emphasis on teacher self-analysis and reflection. Student teachers were taught a systematic self-analysis method to observe video- and audiotape recordings of their own classroom teaching. Written reports of student teachers on the interaction patterns which they had observed showed that many teachers expressed shock at the uneven patterns. In addition they attributed the differences to the behaviour of male
students rather than to aspects of their own behaviour. They reported an increase in the awareness of imbalances and increased motivation to take remedial action. The need to further investigate the effectiveness of interventions aimed at raising teachers' awareness of and reflection on gender-related interaction is explicitly recommended by almost half of the reviewed studies (Cervoni & Ivinson, 2011; Drudy & Úi Chatháín, 2002; Jungwirth, 1991; Myhill, 2002; Smith et al., 2007; Tsouroufli, 2002; Younger et al., 1999; Younger & Warrington, 2002).

1.4.6 Breaking the taboo: student sexual maturity

The quantitative observational data of Younger et al. (1999) show that boys (age 15-16) across subjects and ability levels were reprimanded more often than girls. Teacher interviews suggest that this pattern can partially be explained by the fact that male teachers tend to feel uncomfortable when dealing with explicit challenges of their own authority by girls. Girls are believed to “use their feminine wiles on male teachers” (Younger et al., 1999, p. 332). Boys, and some girls too, share this perception and comment that girls are more willing to apologise or placate the teacher in order not to get punished. This flattering of male teachers by girls is also reported in the qualitative study of Öhrn (1993) with students of the same age. Even though Öhrn approaches the topic of student-teacher interaction from a different viewpoint than Younger and his colleagues (respectively girls' subordination versus boys' underachievement), she finds similar evidence. Again both teachers and students, boys as well as girls, express that teachers, male teachers in particular, are more indulgent towards girls. This lenience is in part ascribed to girls' exaggeration of 'good pupil' behaviour and 'flattering' of the teacher. Öhrn identifies this strategy as one of various scripts that girls use to gain influence and control in school. Einarsson and Granström (2002) raise the subject of romanticised pressure around the teacher and his/her pupils of the opposite sex more explicitly. They make it the focal point of their quantitative research as in previous research “no attention has been paid to 'sexual temptation' in the teaching profession” (Einarsson & Granström, 2002, p. 119). Swedish intermediate level (age 10-12), upper level (age 13-16) students and their teachers were observed and results of quantitative analyses indicate that male teachers pay relatively more attention to female
pupils as they grow older. Einarsson & Granström (2002) state that this change in attention has not been demonstrated in any previous research. In keeping with this statement, no alike patterns were found in any of the other reviewed studies. One mixed methods study in secondary education (Tsouroufli, 2002) even found evidence for the contrary and concludes that teachers tend to be more tolerant, friendly and lenient towards boys. Clearly, the question of sexuality in secondary education deserves more detailed research, even though it is often considered taboo.

1.5 Discussion

Research on gendered teacher-student interactions has been dominated by USA-based studies. We performed a systematic search and review of European studies and discussed the results of 18 selected empirical studies. USA-based research is confirmed with regards to the fact that boys receive more negative feedback from their teachers than girls. Reviewing European research has added three new perspectives to the field that we discuss in the following paragraph.

1.5.1 Three new perspectives

First our review reveals that inconsistencies in research concerning positive feedback might be related to cultural differences since UK-based and continental Europe studies show opposite findings. UK studies point to boys receiving more positive feedback than girls while the studies performed in continental Western-Europe show the opposite. We hypothesise that cultural differences between Anglo-Saxon and other western cultures might exist with regards to positive and appreciative approaches to learning (as opposed to deficit-based approaches). However, we also nuance that methodological differences between studies might lie at the base for inconsistent findings with regards to positive interactions. More research into positive interactions is clearly needed, and attention should be provided to possible cultural differences in the importance of appreciative versus deficiency-based approaches to learning.
Second we identified a new topic of research: sexual temptation in teacher-student relationships. Indications are found for romanticised pressure in the higher grades of secondary education in several of the European studies. Male teachers are found to pay more attention to girls as they get older and girls report using their ‘feminine wiles’ to influence their male teachers. Previous research has disregarded the issue of sexuality in secondary education. Contextual and cultural differences could explain this lack of attention to sexuality in previous research since it has mostly been USA based and in comparison to continental Western-European countries sexuality is more of a taboo there. In Europe sexuality in education is less censured. Governments support sex education and public health policies support widespread public education campaigns through internet, television, films, radio, billboards etc. with a focus on safe and pleasurable sexuality (Alford & Hauser, 2000). Further research into sexuality in teacher-student relationships is needed.

Third, we discuss the scarcity of experimental and intervention research aiming to change and not only describe gendered interactions in both USA and European research. Alex Harrop and Jeremy Swinson (2000, 2011) who extensively studied gendered teacher-student interactions were the only ones performing intervention research to our knowledge. The study, however, has not been published with a gender focus. Nevertheless, their suggestions are promising showing that an appreciative and positive pedagogy can be used by teachers to improve the ratios of positive to negative feedback to students (which are found to be lowest for boys). A higher ratio of positive to negative interactions is related to more on-task behaviour, but few empirical research has actually investigated the causal relationship between both. We conclude that more research is needed on the effects of ‘positive pedagogy’ as a strategy to counter teacher gender bias. Appreciative approaches to learning are also argued to be a promising perspective as a strength-based professionalisation approach for teacher trainers to facilitate teachers’ inquiry into their own biased teaching practices. We devote a separate section to the discussion of appreciative inquiry, since it provides interesting suggestions for future intervention research.
1.5.2 Appreciative Inquiry

It is often stated that the driving force for reflection is the recognition of uncertainty involving a problematic situation (Hoban & Hastings, 2006; Loughran, 2002). Consequently, many models for teacher inquiry depart from problem statements. Appreciative inquiry (Cooperrider & Srivasta, 1987), however, suggests a strength-based alternative, departing from the best of what is and the desired future state that could be. This ‘positive principle’ is assumed to increase positive emotions, which in turn leads to people being more flexible, integrative and open to information (Bushe, 2011). This appreciative approach to inquiry is promising, since teachers can be very resistant to critical and systematic inquiry into their own practice. Many teachers are unconcerned about theory and their teaching is often driven by intuition (Hoyle, 1980). Also, teachers are not familiar with seeing themselves as the producers of knowledge. Most teachers believe that the knowledge base for teaching is constructed by ‘experts’ and they merely have to apply the knowledge produced by others (Cochran-Smith & Lytle, 1992). The appreciative approach to inquiry is hypothesised to increase teachers’ motivation to step outside of their comfort-zone and increase their self-efficacy and optimism to engage in collaborative teacher inquiry (Verleysen, Lambrechts, & Van Acker, 2015).

1.6 Conclusion

This chapter has identified several promising perspectives for future research in the field of gendered classroom interaction research. Suggestions were based on an analysis of eight existing literature reviews and eighteen recent (Western) European studies. The reviewed studies confirm boys’ higher level of negative interactions with teachers and the studies remain inconclusive concerning positive interactions to boys and girls. To improve the ratio of positive to negative feedback for boys and girls, appreciative approaches to learning are argued to be a promising perspective in two ways: firstly as an appreciative pedagogical approach for teachers to respond to student behaviours and secondly as a strength-based professionalization approach for teacher trainers to facilitate teachers' inquiry into their own teaching practices. Exploring these questions will require a shift towards more
experimental research that tries to manipulate gender differentiated interaction patterns and as a consequence student behaviour responses and student achievement.
Chapter 2
Increasing Teacher Awareness
Chapter adapted from
Abstract

Teachers believe they do not interact any differently with boys than with girls. However, an examination of the evidence base on gendered student-teacher interactions, shows— at times contradicting—unequal interaction patterns for boys and girls (e.g., Beaman, Weldall, & Kemp, 2006; Jones & Dindia, 2004). In this study, the videotaped lessons of 13 secondary school teachers in three schools are analysed by both the teachers themselves and the researchers. Video-stimulated recall is used to bring to the surface the thoughts, images and emotions that evoke teachers' (re)actions in the classroom. Content analysis and thematic analysis of teachers' recall reveal three things: (1) when recalling specific situations, gender imbalances in the thoughts and images that accompany teachers' (re)actions come to the surface; (2) video-stimulated recall interviewing is a promising method to raise teacher awareness of these gendered thoughts and images; and (3) with a view to practitioners' professional learning process, it is important for researchers to allow practitioners to co-investigate their practice.
2.1 Introduction

Differential teacher attention to boys and girls has received increasing scrutiny as a predictor of boys' scholastic underachievement (Beaman et al., 2006). Boys in high school repeat a year more often, they are more likely to dropout of education, their overall grades are lower compared to girls', they deal more frequently with behavioural issues, are overrepresented in special education, and are less likely to attend higher education (Buchmann et al., 2008; Clark et al., 2006; Pollack, 1998). A growing amount of research suggests a relationship between more negative teacher feedback and lower pupils' interest, perceived competence, motivation, learning and social behaviour in the classroom (e.g. Brophy, 1985a; Eccles & Blumenfeld, 1985; Foote, 1999; Jones & Wheatly, 1990; Morgan, 2001). From the 1990s, research on gender differentiated teacher attention has focused predominantly on boys being disadvantaged (Galton et al., 1999; Merret & Wheldall, 1992). Many studies have shown boys to be involved in more negative interactions with their teachers than girls (e.g. Harrop & Swinson, 2011; Jones & Dindia, 2004; Myhill, 2002; Smith et al., 2007; Younger et al., 1999). Clearly, there is a dynamic interaction between boys' undesirable behaviours and teachers' reactions. Boys' undesirable behaviour in the classroom is, in general, of a more externalizing nature than girls' and therefore more visible (Francis, 2000). Boys' behaviours elicit conflict and the teachers' reactions give rise to even more aggressive behaviour. This interplay between actions and reactions makes studying gender bias in classroom interactions complex and results are prone to contradictory interpretations. For example, on the one hand a number of studies emphasise that boys receive more teacher attention in general: both negative and positive. Boys' higher levels of reprimands and criticism (e.g. Good, Cooper, & Blakey, 1980; Good & Findley, 1985; Irvine, 1986) are understood to be a reaction to their loud and obvious forms of inappropriate behaviour. More instruction and feedback are found to be a result of more calling out answers, volunteering for experiments, and teachers directing more questions to boys, possibly in attempts to retain their involvement and keep control (Francis, 2000; Warrington et al. 2000). On the other hand, several studies agree with the higher levels of negative attention but report lower levels of positive attention for boys (Myhill, 2002; Nicaise et al., 2007; Öhrn, 1993; Tsouroufli, 2002). Also, authors such as Younger et al.
Increasing Teacher Awareness

(1999) criticise that it is the behaviours of only a limited number of disobedient boys – and not all boys in general – that account for most of the reprimands. The 2009 study of Swinson and Harrop explicitly checks for this hypothesis but does not provide any evidence for it. They describe that most boys receive more negative attention than most girls. Forty-eight focusgroup interviews with students (Younger et al., 1999) reveal that girls are well aware of the reported inequalities and get away with minor rule-breaking by engaging in 'good pupil' behaviour and by flattering and placating the teacher. The study also shows that girls interact more inquisitively with the subject being taught, participate more in the inquiry process and ask most of the academic questions.

Studies carried out in the 1990s might not have cleared up all ambiguities, but they have learned that the quality of teacher attention is more important than the quantity, and that interaction patterns cannot be divided purely along gender lines (e.g. taking into account other variables such as ethnicity, student achievement, classroom composition and subject). Also, gender disparity in the interactions between pupils and teachers has been related to teachers' disparate beliefs and expectations. Research has shown, for example, that teachers have higher normative and academic expectations for girls than for boys (Van Houtte, 2007).

Despite the growing evidence base on teacher-student gender differentiated interaction there are strong indications that teachers are not aware of the fact that they take pupils' gender into account when teaching. Garrahy (2001) suggested that teachers work from a gender-blind position. This gender blindness of teachers is also argued by Younger and Warrington (2005) who found that teachers perceive little difference in the way they treat boys and girls, while pupils – especially boys – feel that boys and girls are treated differently in terms of classroom management, attention and support, questioning and teachers' attitudes. Raider-Roth et al. (2008) confirm this lack of awareness, stating there are powerful but often unconscious forces of relational disconnection in teachers' attitudes towards boys. They investigated how teachers' conceptions of their relationships with boys and their capacity to connect with the boys they teach are shaped. Their findings suggest that the way teachers enter into relationships with boys shape, and are shaped by, teachers' personal and professional identities, boys' responses to school and classroom culture, and the forces of school culture on teachers and boys.
Awareness of these forces would allow teachers to investigate how they develop thoughts, images and feelings that evoke gender specific practice, and examine and eventually change their practice seeking support from colleagues. Raider-Roth and her colleagues conclude that there is need for professional development programmes that support teachers' investigation of, and reflection on, their own practice in relation to issues of gender, identity, teaching and learning. Several other researchers working on gender differentiated student-teacher interaction (e.g. Myhill, 2002) emphasise this need to further investigate the effectiveness of interventions aimed at raising teachers' awareness and reflection on gender related interaction.

Teachers' recognition that there are, or might be, problematic issues in their relationship with the boys they teach is a first prerequisite for eliciting professional learning on this topic. It is the recognition of uncertainty involving a problematic situation that initiates and drives the reflective process of professional learning (Hoban & Hastings, 2006; Loughran, 2002). However, considering the gender blindness referred to in the research literature, it is likely that most teachers do not perceive problems embedded in their own practice and they are not likely to spontaneously question their attitudes towards boys and girls or seek to develop gender-sensitive teaching strategies. Therefore, attention has to be paid to specific strategies for raising teachers' awareness. Dewey (1933) identifies this phase of the teacher becoming perplexed about a situation as the pre-reflection phase. In this phase it is often highlighted that teachers need to see their practice through a different perspective, for example by reading educational literature or by observation by a colleague or teacher educator (Loughran, 2002; Hoban and Hastings, 2006). Hoban and Hastings (2006) focus on viewing one’s own practice through the lens of one’s own students. They discuss four strategies for teachers to question their assumptions by gathering data from their own students: conducting interviews with students, asking students to write about their class experiences in learning logs, asking students to complete a classroom observation schedule and getting students to fill in a survey.

Recent technological evolutions and videography in particular have added powerful – and increasingly popular - possibilities to switch perspectives about one’s teaching (Seidel & Strümer, 2014; Kersting, Givvin, Sotelo, & Stigler, 2010). The richness of video allows displaying the complexity of a real classroom setting by capturing voices, body language, interactions, and
providing a realistic picture of the learning environment (Koc, Peker, & Osmanoglu, 2009). One videography strategy that we examine in this study is the use of video-stimulated recall interviews (Schmid, 2011; Tripp & Rich, 2012) in which teachers watch video’s of their own teaching and discuss these with others. This enables teachers to focus their analysis and gain a new perspective on their practice (Tripp & Rich, 2012). Sherin (2004) states that such video recordings provide: “the opportunity to develop a different kind of knowledge for teaching – knowledge not of 'what to do next', but rather, knowledge of how to interpret and reflect on classroom practices' (p. 14). Video-stimulated recall can be assumed to bring to the surface ‘how teachers think when they deliberate about what to do in the classroom’ or the ‘practical arguments’ of teachers as Pendlebury describes (1990, p. 171). Earlier work on practical arguments and practical reasoning from educational philosophers such as Green (1976) and Fenstermacher (1986) emphasise the more structural rational thinking of teachers. Noel (1999) builds on that by investigating how practical reasoning is not rigid and restricting thoughts but can incorporate tacit knowledge, moments of insight, allowing for emotion and imagination in teacher thinking. All teachers’ knowledge is therefore different, constructed out of a combination between teachers’ personal background and the particular situation they are in. It is wrong to assume that theories and guidelines about (gender-sensitive) teaching that are developed by educational researchers could automatically translate into classroom teaching practice and, thereby, in improved student learning (Connelly & Clandinin, 1997). Clandinin stresses the need to change teachers’ personal practical knowledge. A type of knowledge which does not refer to an objective truth that is generated by those who specialize in research on teaching, but to a mix of cognitive, affective and intuitive processes (Fenstermacher, 1994). The aim is to construct and deconstruct teachers’ personal practical knowledge, through a dialectic process that links theory and research, on the one hand, and educational practice, on the other hand.
2.2 Research questions

The research questions that will be addressed in this study are the following:

(1) to what extent are teachers' thoughts, images and emotions that evoke (re)actions to students gendered?

(2) can video-stimulated recall raise teachers’ awareness of their possible gender bias?

2.3 Method

2.3.1 Participants

This study is based on the first phase of a larger research programme involving a representative sample of 59 secondary schools in Flanders (Dutch speaking part of Belgium). The study reported in this paper is based on the research and development activities with a sub-sample of three experimental schools. The three schools involved were public schools in urban areas: one small, one medium-sized and one large-sized school. The schools also differed in the types of secondary education they organize - only general; vocational and technical; and both general, vocational and technical – thus supplying a variation from quiet studious classes to more wilful, challenging ones. With support from the principal, a team of five teachers who volunteered to participate was composed in each school. They had to be teachers in the first two years of secondary education who taught general subjects, partly in the same classes, or at least they needed to know the students in the others classes. All teachers had between 3 and 23 years of experience as a teacher (M=12). In each school both male and female teachers participated with a total of 5 male and 10 female teachers. Teachers were aged between 27 and 49 years of age (M=36).

All participants (principals, teachers, students, parents) were informed about the research goals and research activities. Participants could decide not to participate and refuse to be video-recorded. No refusals occurred in the schools involved in this study. Due to practical reasons (illness of teacher, loss of recordings) the video-stimulated recall interviews of only 13 teachers could be analysed.
2.3.2 Data collection

Video-taped lessons of 13 secondary school teachers in three schools are used in video-stimulated recall interviews (Schmid, 2011; Tripp & Rich, 2012) to help teachers become aware of their patterns of interaction with their students, and of the thoughts, images and emotions that evoke their (re)actions. The researcher would typically introduce the interview as follows:

'Together we will take a look at some extracts of the recorded lesson and I will ask you to comment on what we see. As a teacher you're constantly observing what's happening in your classroom and you're constantly taking decisions: whether to react or not, whether to make a joke or to get angry, whether to let a student speak or not, and so on. We will try to understand how you make these decisions. Your perspective is what counts. I selected the extracts in which you seem to take such decisions. As soon as you start talking, we can stop the video, rewind, or go to whichever part of the lesson you want to look at or show me. I will first let you watch for a little while just to get used to it; most people find watching themselves on video a little weird.'

Afterwards, the researcher helped the teachers talk about their teaching by applying appreciative and active listening techniques such as nodding, humming, showing empathy, repeating, summarizing, paraphrasing, etc. The episodes that were presented to the interviewees were selected in such a way that the same amount of disobedience by boys and girls occurred. Also, episodes of the pupils behaving in an orderly manner were shown. Teachers, however, did have the freedom to discuss certain episodes, more or less, and were able to select new episodes to discuss. This way, the attention of teachers was neither biased by the content of the episodes nor overly steered by the researcher. It should be noted that the interviews were not designed as a video-stimulated reflection interview, but as a recall interview. The central focus was on exposing teachers’ practical reasoning and a possible gender bias in it. In a reflection interview the interviewer would actively stimulate teacher reflection by pointing to alternative perspectives and conflicting statements of teachers. In this study the interviewer stayed as neutral as
possible in order to capture the effect of merely confrontation with the videotaped perspective of one’s own teaching and a facilitators support in recalling. The interviewer did not explicitly question nor stimulate teachers’ increased awareness.

Most interviews lasted between 60 and 90 minutes and all interviews were recorded and transcribed.

2.3.3 Data analysis

2.3.3.1 Content analysis: gendered thoughts

A content analysis was performed to investigate the degree of gender-differentiation of thoughts, images and emotions that evoke teachers' interactions in the classroom. Content analysis was chosen as it is a method for systematic and rule-guided classification and the description of text material with the aim of comparing categories in a, somewhat, quantitative way (Berg, 2007). So that other researchers can obtain comparable results, selection criteria were consistently applied to systematically determine if a message content could be assigned to a certain category. Trying to achieve some kind of reliability of the coding was essential, especially since researchers may introduce bias due to a heightened attention for gender differences as gender researchers. Establishing intercoder reliability is an attempt to reduce the bias generated when individuals (unconsciously) make errors when processing the large amounts of textual data generated by qualitative research. To achieve acceptable levels of reliability, the following process was followed: segmentation of text, codebook creation, coding, assessment of reliability by two independent coders, codebook modification, and final coding – with coding, assessment of reliability and codebook modification conducted several times in iteration (Hruschka, Schwartz, Cobb St. John, Picone-Decaro, Jenkins, & Carey, 2004).

A number of statistical indexes can assess to what degree a set of texts were consistently coded by different coders. Despite all the efforts that methodologists have devoted to developing and testing indices of intercoder reliability, there is no consensus on a single, 'best' index (Lombard, Snyder-Duch & Bracken, 2002). The commonly used 'coefficient of agreement', which measures the proportion of decisions where coders agree, can
dramatically overestimate the true degree of intercoder reliability by not taking into account the agreement by chance. Therefore, we relied on Krippendorff's alpha (Krippendorff, 2013), which prevents the inflation of reliability scores by correcting for chance agreement, although other statistics also satisfy these criteria (e.g. Cohen's Kappa, Scott's pi). In addition to the choice of the appropriate index of intercoder reliability, another difficulty was determining what constitutes an acceptable level of reliability. Again, there are no established standards. As Krippendorff's alpha is known to be a more conservative index, a more liberal criterion of .70 was used to determine if the coding of a variable was considered reliable. Table 8 gives an overview of the kalpha's for the key codes in the codebook.

<table>
<thead>
<tr>
<th></th>
<th>Cod1: 1</th>
<th>Cod1: 0</th>
<th>Cod2: 1</th>
<th>Cod2: 0</th>
<th>kalpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boys explicit</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>75</td>
<td>0.8831</td>
</tr>
<tr>
<td>Boys implicit</td>
<td>15</td>
<td>3</td>
<td>0</td>
<td>63</td>
<td>0.8452</td>
</tr>
<tr>
<td>Girls explicit</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>73</td>
<td>0.7506</td>
</tr>
<tr>
<td>Girls implicit</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>74</td>
<td>0.9169</td>
</tr>
</tbody>
</table>

*Note. C1: 1 = coder 1 has assigned the code, C2: 0 = coder 2 has not assigned the code.*

### 2.3.3.2 Thematic analysis: raised awareness

In order to answer the question whether stimulated recall raises teachers' awareness of their gendered interaction, and of the need for gender-sensitive attitudes and teaching strategies, the transcripts of the interviews were analysed thematically. Thematic analysis differs from other analytic methods in the sense that it is not directed towards theory development. It is aimed at organizing and describing data in rich detail, not bound to one specific theoretical framework. Braun and Clarke (2006, p. 79) state that: 'it can be seen as a very poorly “branded” method, in that it does not appear to exist as a “name” analysis in the same way that other methods do (e.g. narrative analysis, grounded theory).’ These authors argue that thematic analysis is often used but not explicitly identified as any particular method at all, or
claimed as something else. According to the guidelines of Braun and Clarke (2006), the process of carrying out a thematic analysis breaks down into six stages:

- data familiarization,
- initial coding generation,
- searching for themes based on the initial coding,
- review of the themes,
- theme definition and labelling,
- report writing.

The thematic analysis was performed by a different researcher to the one performing the video-stimulated recall interviews. A risk for euphemising the effects of the interview was avoided by not allowing the same researcher to both perform the interviews and analyse the effect of these interviews on raising teachers' awareness.

2.4 Results

2.4.1 Content analysis: gendered thoughts

2.4.1.1 Perceived misconduct and negative attributions

An analysis of co-occurring codes shows a clear difference between comments of teachers when talking in general and comments made when recalling and describing specific students and situations on the videotaped lesson. When discussing off-task behaviour and misconduct teachers ascribe behaviours such as not listening, cheating, dreaming, lying, rolling eyes or forgetting material at home, roughly as often to boys as they do to girls (Figure 7). When recalling their thoughts on specific situations, however, we see that almost three quarters of the thoughts and images on off-task behaviour and misconduct are made in relation to boys (Figure 7). The following quotes of geography teacher Edward illustrate the difference.
Edward (talking about misbehaviour in general):

'Oh if you put four boys together, they talk. If you put four girls together, you get exactly the same talking.'

Edward (talking about misbehaviour in recalled specific situations):

'Jordan and Kenny are talking about all kinds of things here.'

'Kyle is someone who likes rap and beatboxing and things like that. He is sometimes making noises. You go like: please, stop that.'

'I don't know if you can see it here when Sean is coming in. (...) He gets a remark about it, you can't see it on the tape but I show him the red card. He knows ... when he was entering he was making way too much noise.'

These results say nothing about the actual misbehaviour of boys and girls. They only show that it is the misbehaviour of boys that receives most attention of teachers when teaching. The results also show clearly that when consciously reflecting on their classroom experiences in general, teachers do not pick-up on this imbalance; they genuinely believe that the differences are negligible.

We see the same awareness gap when looking at negative attributions such as students being untidy, impolite, lazy, quiet, deceiving or similar. When talking in general teachers identify these kind of negative traits equally in girls and boys (Figure 7). When commenting on their thoughts about certain episodes in the videotaped lessons we again see three quarters of the negative attributions are related to boys (Figure 7). The comments of English teacher Rebecca on the motivation of students serve as an example:
Rebecca (talking about negative attributions in general):

'You have very motivated boys and very motivated girls. And you have both boys and girls that you have to struggle to get them engaged.'

Rebecca (talking about negative attributions in recalled specific situations):

'During the first lessons I had the feeling not to give Cody too much attention because he was trying to catch my attention in a negative way. (...) I think he was trying to ... looking for a way to measure off if he was amongst the strong students for English or not. Asking himself the question if he had to profile himself or not. But underneath I think he is rather unsure about himself. That's my impression.'

'David is often not in my lesson because he is an English native, and he does French during my lessons. (...) At the beginning of the year David was lying in my lessons with an attitude ... like ... pff I know everything anyways.'
Figure 7. Teachers’ explicit and implicit comments on perceived misconduct and negative attributions for boys and girls.
2.4.1.2 Perceived good conduct and positive attributions

Appropriate and on-task behaviour such as paying attention, listening, helping the teacher, cooperating in group assignments, being quiet and working are rarely discussed in general by the teachers. When analysing the co-occurrence of appropriate behaviour with boys or girls in student specific situations only, we see that the majority of the comments on good behaviour were made concerning boys (35/14). Positive student characteristics ascribed to specific students show a slightly more balanced pattern. Boys are described as having positive traits a little more often than girls (23/19). More interesting than these quantitative ratios are the qualitative differences in the positive comments made about boys and girls. The array of positive characteristics ascribed to boys is much richer than those linked to girls. Girls are said to be easy going, motivated, hard workers and wanting to please, whereas boys are ascribed a set of additional non-academic positive traits such as having a sense of humour, being mature, being extravert, authentic, able to build good relationships with other students, honest etc. Comments about the positive traits of girls seem to have an academic character, whereas boys are also noticed for their social traits.

2.4.1.3 Gender blindness

Most teachers' articulations corroborate the aforementioned gender blindness. The explicit thoughts of teachers display a gender neutral pattern while the implicit thoughts of teachers are clearly gender differentiated. Teachers were not aware of the fact that their comments about individual students and classes implicitly expressed gender-differential expectations and approaches.
### 2.4.2 Thematic analysis: raised awareness

#### 2.4.2.1 Nature of teachers' comments

Table 9. Overview of nature of themes of teachers’ comments (thematic analysis)

<table>
<thead>
<tr>
<th>Nature of comments</th>
<th>Theme</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practical knowledge and beliefs</td>
<td>About students</td>
<td>Knowledge and beliefs about students in general as well as about classes and individual students. The thoughts may express students' needs, students' behaviour, teachers' expectations in regard to students, labelling (groups of) student – possibly boys vs. girls.</td>
</tr>
<tr>
<td></td>
<td>About instructional strategies, teaching and learning</td>
<td>Knowledge and beliefs about rules, principles, instructional approaches that work for them; personal 'theories' and vision</td>
</tr>
<tr>
<td></td>
<td>About student-teacher interaction and relationships</td>
<td>Knowledge and beliefs about characteristics of the interaction and relationship with students in general or with particular classes and students – possibly boys vs. girls.</td>
</tr>
<tr>
<td></td>
<td>About classroom management</td>
<td>Knowledge and beliefs about rules and principles for classroom management; personal 'theories' and vision</td>
</tr>
<tr>
<td></td>
<td>About facilities</td>
<td>Information and beliefs about how facilities (e.g. Size of classroom, teaching aids) determine their decisions and how they would do things differently if they had the opportunity</td>
</tr>
<tr>
<td></td>
<td>About self as teacher</td>
<td>Knowledge and beliefs about their strengths and weaknesses, about the kind of teacher they ware or want to be</td>
</tr>
<tr>
<td></td>
<td>About subject matter and curriculum</td>
<td>Knowledge and beliefs about subject-specific aspects of teaching and learning and the opportunities and limitations of the curriculum</td>
</tr>
<tr>
<td>Becoming aware</td>
<td>Of student behaviour</td>
<td>Utterances of surprise about students' behaviour they observer on video and reinterpretations of student behaviour</td>
</tr>
<tr>
<td></td>
<td>Of own behaviour</td>
<td>Utterances of surprise about own behaviour they observe on video and reinterpretations of their behaviour</td>
</tr>
<tr>
<td>Emotions</td>
<td>miscellaneous</td>
<td>Reliving emotions that occured in action</td>
</tr>
</tbody>
</table>
Critical reflection

| Critical reflection | About own behaviour and beliefs | Critical analysis of own behaviour or beliefs during the interview; considering alternative approaches for the future or intentions to change. |

The overview in Table 9 presents the nature and themes of the teachers' comments in the stimulated recall interviews. The greater part of the interview time goes to teachers propounding 'practical knowledge and beliefs'. The category was labelled so because the sub-themes in this category bear a close similarity to Meijer, Verloop and Beijaard’s (1999) list of categories of teachers' practical knowledge and beliefs, which also resulted (partly) from stimulated recall interviews with experienced teachers. The comments used in the content analysis described previously are illustrations of comments that fall under this category. A second category is 'becoming aware' that gathers utterances of teachers when becoming aware of aspects of their own behaviour or students' behaviour that they had not seen before or interpreted in different ways. The categories 'emotions' and 'reflection' were least represented. The label 'emotions' was used during the interviews when teachers relive an emotion that occurred in-action. 'Reflection' was attributed when teachers critically analysed their behaviour or beliefs during the interview and/or voiced their intention to change their approach in the future.

### 2.4.2.2 Articulating tacit knowledge

In-action teachers continually try to read situations and make intuitive or deliberate decisions (cfr. practical reasoning). During the stimulated recall interviews teachers are asked to articulate the knowledge underpinning their interpretation of situations and the rules underpinning their decision-making. This underpinning of practical knowledge might be explicit or tacit. We refer here to the categories of knowledge mentioned by Polanyi (1966) who distinguished knowledge that is transmittable in formal, systematic language (explicit knowledge) from knowledge that is rooted in a person’s action and involvement in a particular context and is therefore hard to formalize (tacit knowledge). Clandinin (1985) states that personal practical knowledge is mainly made up of tacit knowledge which has arisen from experience and is embodied in the actions of the teacher. As mentioned earlier, teachers
Increasing Teacher Awareness

communicate a lot of practical knowledge and beliefs to explain their understanding of situations and the thoughts underpinning their decisions. For the first time, during the interview, several teachers point out that they externalise some of these thoughts.

Stephen:

'I don’t expect them to be focused all the time. If it relaxes them to talk about something else for a minute – because now [watching the video] they’re working again – I let them. Then afterwards, they can concentrate again. That is how I work. I never thought of it before.'

Apparently, watching their concrete actions and communicating about them, helps teachers to externalise their tacit knowledge into explicit knowledge.

2.4.2.3 Raised awareness and reflection

A theoretical assumption was that video recordings of teachers’ own classroom practice would be an appropriate tool to stimulate reflection and raise teachers’ awareness. We would like to remind the reader of the fact that the interviewer was merely intended to support teachers in recalling thoughts and not in stimulating in depth reflection by confronting teachers with contradicting thoughts or alternative perspectives. Yet, some teachers critically analysed their behaviour or cognitions-in-action spontaneously during the interview and thought about alternative approaches for the future. About half of the teachers expressed how looking at, and talking about, the video of their own lessons made them notice aspects of their own and their students’ behaviour they were not previously aware of. These episodes of increased self-awareness during the stimulated recall, while they do occur, are so sparse that a more systematic analysis of this sub-category of data was not possible. We do present some examples of becoming aware. Some of the teachers became aware of aspects in their own behaviour.
Stephen:

'In fact I should have reacted more strictly here but it was so close to noon. In fact, yes, I should have reacted more severely because I did when Tony disturbed the lesson and not when Bjorn did. [...] It’s not right, I think. I try to make sure that I treat them all equally. There are always student that you prefer. But there’s no need for them to know. All equal before the law.'

Ann:

'Whoa, I’m... I’m overreacting there. That was so very direct. [...] Students say that I’m harsh but I think I’m not. I think I tolerate a lot. Well, but with this class there’s no need to be too severe. Still I’m so... well so... I could have said “Mehmed, don’t forget to jot down the distance”. But no: “Mehmed, distance!”’. That makes a difference. I’m shocked.'

Teachers also discover that the way boys and girls behave in their classroom is different from what they thought or expected. The become aware of aspects in students’ behaviour.

Mary:

'I think I'm ignoring the laughter there [on the video]. In the end it was not funny anymore. But I'm surprised to see how many of the girls are laughing too. [...] These are things you don't notice when teaching. At least not so clearly...'

Stephen:

'Here ... see ... they are all at work. I'm surprised. You wouldn’t think ‘oh, so much noise: they are all playing’, but no they’re not. I didn’t realise that.'
2.5 Discussion

Content analysis of teachers' recall makes clear the gender-differentiated character of teachers' thoughts, images and emotions that accompany classroom interactions. Misconduct and negative character traits are more often commented on in relation to boys than to girls. We see a clear awareness gap concerning these gendered thoughts in the sense that gender imbalances that are present in the teachers' implicit recall are not present in their explicit reflection. Methodologically, this finding raises questions on research that tries to capture gendered thoughts of teachers – of which teachers are often unaware of - by means of regular interviews or questionnaires. The findings of our video-stimulated recall, which measures thoughts that accompany classroom interactions, are in-line with the research that investigates actual behaviour and actual teacher-student interactions, showing that boys are involved in more negative interactions than girls (e.g. Beaman et al., 2006; Brophy, 1985a; Jones & Dindia, 2004). With regard to the noticing of positive student traits and desirable behaviours, no quantitative but qualitative differences were found between boys and girls. Comments about girls seemed to be much more based on good academic behaviour and attitude than comments about boys who are also noticed for social traits. Our results are not in-line with the results of some studies showing that girls receive more positive attention in general than boys (Myhill, 2002; Nicaise et al., 2007). Our findings would rather support the idea that boys receive more attention in general, both positive and negative (Harrop & Swinson, 2011; Merret & Wheldall, 1992), with the remark that qualitative differences exist. The question remains however what effects these gender differentiated thoughts of teachers have on students’ perception and academic achievement. In this particular study no measure of student perception or achievement was included, but previous research does point to positive correlations between teacher thoughts and expectations and student perceptions and outcome. Good and Brophy (1978) found that teachers criticize students more frequently for incorrect public responses if they hold lower expectations for them. Recent work of Van Houtte (2007) shows that teachers have lower expectations for boys and we could hypothesise that this translates in an unfairly high amount of criticism in comparison to girls. Still, a direct
relationship between teachers’ behavioural bias towards boys and boys’ underachievement is not self-evident. Brophy’s extensive work on the self-fulfilling prophecy and teacher expectations (1982) shows that students will differ in their interpretation of and response to teacher behaviour, meaning that similar behaviour of teachers may produce different student outcomes. Students play an active role in whether teachers’ bias and negative expectations are internalised or whether they are actively resisted. The large-scale mixed methods work of Younger et al. (1999) showing that both teachers and students feel that girls get away with bad behaviour more easily, and the similarly rigorous work of Myhill (2002) showing that high achieving boys tend to become underachieving boys towards middle school, however seem to support the hypothesis that boys do internalise teachers’ (and not only teachers’) negative expectations. More research investigating the relationship between teacher practical knowledge, teacher behaviour and student outcomes however is needed to make such statements.

Most strategies for realizing change in educational practice have emphasised the transfer to practice of research findings, models and theories conceptualized by researchers. We observed that, even if such knowledge produced by scholars was significant for improvement in practice, this rarely led to well-understood integration of this type of knowledge in teachers’ practical working models of the educational practice in a specific context. In the last decennium a lance was broken for practitioners’ research and the long standing but limited epistemological perspective that theoretical knowledge is to be developed by researchers, and afterwards implemented or applied by practitioners, is being discarded (Fenstermacher, 1994). However, involving teachers in research and not merely seeing them as consumers of knowledge but also as the creators of knowledge, while maintaining the standard approaches to educational research, is not an answer to the problem of disconnection between research findings and implementable knowledge. We searched for a model in which the different stakeholders could contribute, starting from their divergent areas of professionalism, in which they were respected for the specific knowledge they could provide, and in which they preserved ownership with regard to their core tasks (Nonaka & van Krogh, 2009). The combination of teachers’ practical knowledge – involving the externalization of tacit knowledge – combined with the knowledge generated by the researcher through systematic analysis, was central in the design.
As to the question whether video-stimulated recall can increase teacher awareness of unconscious gender bias, it is obvious that in depth reflection and inquiry will not happen spontaneously and data does not talk to everyone (in the same way). The changing of perspective and investigating student perspectives is not enough for teachers to elicit pre-reflection and the increased awareness of problems. Hoban and Hastings (2006) in their discussion also state that teachers’ readiness to consider personal feedback on their teaching needs to be taken into account. We could hypothesise that teachers are not open to self-question themselves and therefore fail to consider that their practice is problematic, without some extra support or help from others in taking different perspectives. Video stimulating recall interviewing can be a promising method to increase teacher awareness if efforts are made to provide teachers with the necessary support and safety conditions. Despite the fact that the teachers involved in this study hardly knew the researcher, were not familiar with the practice of stimulated recall interviews, and most of them felt uncomfortable watching themselves on video, they communicated rather openly about their practice. In meta-reflection later on (in a follow-up professional development programme using collaborative teacher inquiry), teachers attributed this to the non-evaluative nature of recalling on actions with an external researcher. This finding led us to the conclusion that it seems to be crucial to involve a non-threatening facilitators when engaging with teachers in video-based reflection. We believe principals, school district evaluators and inspectorate members are not the best positioned actors to facilitate in video-based reflection since it is difficult for them to guarantee a sufficiently safe environment for teachers to questions their practice since they combine a role of high-stakes assessment and guidance and support towards the teacher. We also want teachers to develop reflective intelligence (MacGillchrist, Myers & Reed, 2004) while at the same time respecting teachers’ autonomy and ownership of the change process. We could criticize that an external researcher might guarantee neutrality, but threatens the achievement of sustained support to teachers by establishing dependence on researchers (Butler, Novak Lauscher, Jarvis-Selinger & Beckingham, 2004). Butler et al. (2004) propose a combination of an experienced external facilitator that teachers have access to while at the same time fostering the development of teachers’ competences to support one another in video-based reflection. Future research could further investigate
the effect of different profiles and roles for facilitators in video-based recall and reflection. Also, the content focus of the recall could be on a different topics chosen by the participants to increase the chance of teachers to be willing to self-question themselves and become aware of potential problems in their behaviour.

In summary, plenty of opportunities for video-stimulated recall and reflection are revealed as a means for stimulating pre-reflection and inquiry-based professional learning and it’s potential for successful co-development of scholarship and practice.
Chapter 3
Gendered Classroom Interactions
Chapter adapted from
Abstract

Gendered teacher-student classroom interactions have been the subject of extensive empirical research over the last four decades. However, several gaps and inconsistencies in the research-base still exist. With regard to positive teacher feedback, past research is contradictory, and with regard to negative teacher feedback, it is unclear which boys in particular are responsible for the increased levels of criticism. An incongruency exists between students’ perception of negative teacher bias towards boys and observational studies claiming that boys bring criticism upon themselves due to increased levels of disruptive behaviour. In this mixed methods study, the relationship between student demographic background, teacher-student classroom interactions and students’ perception of equity in the classroom environment is investigated. Questionnaires and standardised tests were administered to 6380 first-year secondary students. For a sub-sample of 180 students, fifteen-minute fragments of video-recorded lessons were coded. An intersectionality-informed approach is used to analyse additive and multiplicative effects of demographic background variables and their intersecting categories. Increased levels of negative feedback towards boys are confirmed, but some degree of unfair teacher differential treatment is supported. Boys report a lower sense of equity in the classroom environment than girls, and girls from a foreign language background appear to be an extremely vulnerable group, not participating in classroom interactions and frequently being off-task. We suggest that teachers be careful with tips and tricks on ‘how to handle boys and girls’. We see more merit in increasing teachers’ awareness and monitoring skills to control incorrect and negative expectancy effects.
3.1 Introduction

Gendered teacher-student classroom interactions have been the subject of extensive empirical research over the last four decades in many countries with a strong tradition in English-speaking countries such as USA, UK and Australia (Howe & Abedin, 2013). Beaman et al. (2006) kick off their widely-cited literature review on the topic with the statement that “the issue of sexual inequality in the classroom has been of concern for over 30 years, frequently generating more heat than light, in both academic and public debate” (p. 339). One could state that the amount of research information on the topic has exploded, generating an “information overload” (Petticrew & Roberts, 2006) for scholars, teacher educators and policy makers. Fortunately, through the years several authors have tried to synthesize the growing – and at times contradictory – evidence-base. We identified eight review studies on gendered teacher-student classroom interactions by searching the ERIC and Web of Knowledge databases and by browsing the bibliographies of articles of interest. In the references section of this paper the reviews are marked with an asterix. Each decade since the 1980s is represented by one to three reviews, and a careful examination of these works brings to the surface the richness of the study field and the congruences, inconsistencies and gaps in past research. The empirical study presented in this paper will be introduced by making as much use of these existing reviews as possible. When reviews report a cumulation of evidence on a certain topic, we limit ourselves to discussing the state of the art by means of the reviews. When reviews are inconclusive we proceed to discussing primary studies as well.

A first inspection of the reviews shows three major groups that can be distinguished in the body of research (often used to structure the reviews): studies investigating student variables, studies investigating teacher variables, and a third group of studies investigating an amalgam of other variables such as stereotyped curricular materials and biased evaluation instruments. The vast majority of research has focused on the study of student variables and in particular the effect of student gender. Since our interest lies in teacher-student interactions no discussion of the third group of research concerning materials and instruments will be provided.
3.1.1 Student gender and quantity of interactions

One central finding of Kelly’s (1988) extensive meta-analysis of 81 empirical studies that has been reproduced by all following reviews is that teachers tend to interact more with boys than with girls, in preschool and throughout high-school. Kelly concludes that teacher-initiated behaviours at least in part lie at the base for this increased interaction with boys, by asking boys more questions and providing them with more response opportunities than girls. Five years later, Bailey (1993) adds to that finding by noting that teacher attention to boys is not evenly distributed among all boys. It is “certain subgroups” of boys, she states, that are causing the increased level of interactions for boys in general. Due to the limited amount of research that examined differences in interactions in terms of race, ethnicity and socio-economic status, Bailey however could not draw any conclusions about the specific profile of these subgroups receiving the bulk of teacher attention. All following reviews concur with the finding that the disproportionately high share of boys’ contributions stems from the extreme volume of contributions of some boys. In 2006, Beaman, Wheldall, and Kemp explicitly critique in their suggestions for future research that “classroom interaction patterns can not be divided along purely gender lines, this is too simplistic and drives by rhetoric rather than by evidence.” (p. 361). Curious about which student characteristics are known to influence classroom interactions we broadened our literature search and hit upon Howe and Abedin’s 2013 review of four decades of classroom dialogue research. In the 92 studies they review that investigate student variables, 45 studies included gender, 17 included attainment and 16 included ethnicity. Gender has always been, they conclude, at the center of research on classroom dialogue, with other variables – usually attainment and ethnicity – only gaining serious attention in the 1980s.

With regard to attainment the pattern that has been recurring is that low achieving boys are least likely to receive teacher feedback and opportunities to respond (contrary to boys in general). The early work of Brophy and Good (1974) is one of the first to describe this pattern and – Howe and Abedin (2013) state – the pattern is confirmed up until today. The authors call attention to the fact that this increased attention is probably – at least partly – the consequence of high attainers initiating interactions with their teachers more actively (such as raising hands).
With regard to overall participation of ethnic minorities, Howe and Abedin (2013) conclude that the findings of past research are inconclusive. Nevertheless, they conclude that the effects of gender, attainment and ethnicity as predictor variables of classroom dialogue have been studied plentifully and should no longer be prioritized in research. An occasional check – so they state – should suffice.

3.1.2 Student gender and positive and negative interactions

Much research has tried to study this quality of interactions by focusing on teacher positive and teacher negative feedback towards students. Olivares and Rosenthal’s (1992) review concludes that boys receive both more negative and more positive teacher feedback than girls. This first part about boys receiving more negative feedback than girls is confirmed by all subsequent reviews (except for the review of Plumm (2008) which doesn’t discuss negative and positive interactions). With regard to the higher levels of more positive teacher feedback directed at boys than at girls, the reviews are indecisive. Howe (1997) and Consuegra and Engels (2014) conclude their reviews by explicitly calling the existing research concerning the topic contradictory. The meta-analysis of Jones & Dindia (2004) suggests that boys and girls receive equal amounts of positive teachers interactions, though it should be noted that the power to detect an effect size significantly different from zero was low. In their 2013 systematic review of classroom dialogue, Howe and Abedin are noticeably unsure but they write that boys “occasionally” receive more positive feedback than girls. In our search for understanding the reasons behind these inconclusive findings, we found Harrop and Swinson’s (2007) critique on the varied operationalizations of ‘praise’ in much classroom interaction research to be essential. Harrop and Swinson’s methodological review shows that some studies include non-verbal praise such as nodding while other studies don’t. Some take into account neutral comments and encouragement such as ‘yes’ and ‘continue’ while others only count very explicit positive remarks such as ‘excellent’ and ‘well done’. Another explanation for the contradictory findings might lie in the gendered contexts in which teacher-student interactions are taking place. The review of Consuegra and Engels (2014) suggests that subject matter of the observed classes influences patterns of positive feedback. The authors
refer to studies of Tsouroufli (2002) and Nicaise et al. (2007) arguing that boys receive more praise and encouragement than girls when involved in male-traditional activities such as circuit weight training and mathematics in contrast to reversed patterns of negative and positive interactions in gender-neutral or feminine-traditional activities. Despite conflicting results concerning positive interactions, most of the review studies do agree that the relative frequency of positive and negative remarks is more negative for boys than for girls. Howe in 1997 even suggests studying this proportion rather than the separate frequencies of positive and negative feedback, since it may be more significant than the absolute amounts in terms of impact on pupils. In her doctoral dissertation on praise to reprimand ratio recommendations, White (2010) reports a range from 3:1 to 10:1 for optimal promotion of appropriate student behaviours and classroom management. Correlational analyses suggest that the higher the ratio of positive to negative interactions the more pupil on-task behaviour is displayed. White did not find any experimental studies where praise to reprimand ratios were directly manipulated to prove a causal relationship.

3.1.3 Student behaviour and teacher response

The implications of teacher feedback on students have been a recurrent concern in the field of gendered classroom interactions research. Nevertheless, few empirical studies have actually investigated effects of teacher differential treatment towards boys and girls. Howe (1997) comments that virtually all of the classroom interaction research is limited to descriptions of interactions that take place. Few studies measure effects on student perceptions, academic performance or social attitudes. This information is however key to formulating suggestions for improving teaching practices in order to reduce or eliminate gender bias. In his 1985b review Brophy suggests future research should focus not on the existence of self-fulfilling prophecy effects, but on how exactly these complex expectancy effects are mediated through differential teacher treatment. He discusses 17 possible mediational mechanisms of teacher expectations and he successfully disentangles the following concluding elements to take into account when studying teacher differential treatment and its impact on students:
1. teachers differ in whether or not they project inaccurate expectations to students
2. students differ in their susceptibility to being conditioned by teachers’ expectations
3. many differential interaction patterns may represent student effects on teachers rather than teacher bias, at least in part
4. some forms of differential treatment are appropriate. (pp. 312-313)

Brophy illustrates very well the complexity of teacher expectancy and teacher bias effects. In the 1990s, however, serious doubt emerges about the degree to which differences in interactions are really a matter of teacher gender bias (Beaman et al., 2006). It is hypothesized that boys bring criticism upon themselves, as a consequence of their higher levels of misbehaviour. Howe (1997) concludes her review somewhat along these lines, claiming that boys receive more overall interaction due to their being more visible. Boys would be, so she states, eight times more likely than girls to call out in class, raise their hands more quickly and are more restless (hence attracting more teacher attention). This would result in teachers gazing roughly twice as often at boys than girls because they are monitoring where they anticipate misbehaviour. Increased levels of negative feedback towards boys are thus suggested to be justified, based on boys’ higher levels of disruptive behaviour. Howe concludes that differences in whole-class teacher instructions are unlikely to disadvantage boys and influence their school performance.

It is a pity that the reviews of Consuegra and Engels (2014), Plumm (2008) and Jones and Dindia (2004) do not report a systematic discussion of results concerning the reciprocity of student behaviours and teacher response. Jones and Dindia (2004) do criticize that few studies have succeeded in grasping this powerful dynamic and they conclude that research has failed to take into account the active participatory role of the student. Howe’s (1997) review provides some examples of what this active role of students could be. She refers to some studies indicating that girls develop ‘compensatory strategies’ for their lower levels of participation in whole-class settings, such as demanding more individual instruction time from teachers. This would result in boys being more prominent in whole-class settings and girls interacting with their teachers on a more individual basis, without one or the other being at a disadvantage per se. We continued our search for empirical evidence on
the active role of students by reviewing primary studies that investigate
teacher-student interactions from a student-perspective. A qualitative study of
Öhrn (1993) identifies two compensatory strategies that girls use to gain
influence and control on teachers in classrooms which are dominated by
boys. The study discusses confrontational strategies in which social networks
of girls intimidate teachers to make them feel uncomfortable with going into
conflict with one of the members of the network. Another strategy is girls
exaggerating ‘good girl’ behaviour and flattering the teacher to get away with
rule-breaking behaviour. Öhrn found boys as well as girls expressing that
teachers are more lenient towards girls. The study of Younger et al. (1999)
partly confirms this perception of students that teachers treat boys and girls
differently. Boys report feeling unfairly treated, discriminated against and not
deserving their bad reputation. One group of girls that was interviewed
agreed with these opinions. But the other groups of girls accepted that they
received more support and less negative feedback, but mostly they felt that
this was justified. The study also shows the strong mismatch between boys’
(and some girls’) perception of teachers’ unfair treatment towards boys and
teacher themselves being convinced of unbiased approaches.

3.1.4 Teacher characteristics

We have already mentioned that teacher variable effects on teacher-student
classroom interactions have received far from the same amount of attention
as student variables. Also, hardly any general systematic approach can be
detected in this group of research. Considering this, it should not surprise that
many of the reviews only discuss the topic in parenthesis; for example, to
help explain inconsistent findings on student variable effects and without the
ambition to synthesize results. Nevertheless, authors do acknowledge that
teacher effects might be in play. Brophy (1985b) emphasises that teachers’
personal characteristics and beliefs about teaching and learning will
determine teachers’ susceptibility to negative self-fulfilling prophecy effects
(which he calls Golem effects). He discusses a typology of teachers as
proactive, reactive or overreactive to characterize their proneness to rigid and
stereotypical expectations and biased student treatment. We are simplifying
his argument, but after his discussion of a set of – at times questionable –
assumptions, he concludes that teachers with a stronger mastery-approach will be less susceptible to Golem effects than teachers with a performance-approach. We are not familiar with any studies confirming this hypothesis. In 1997 Howe makes the effort to examine the existing evidence on teacher variable effects, and concludes that there are few indications that teacher variables of any kind (including teacher gender) exert influence on interactions. In 2004 Jones and Dindia aim to analyze teacher gender effects on teacher-initiated interactions with boys and girls, but due to too few studies reporting sufficient data on teacher variables no such analysis was possible. The review of Consuegra and Engels (2014) concludes that while it might be true in general that teacher gender does not affect patterns of teacher-student interactions, effects can be found if specific settings are taken into special consideration. Consuegra and Engels (2014) refer to a study of Einarsson and Granström (2002), which found a teacher gender effect in higher secondary education where male teachers seem to pay more attention to female pupils (in contrast to the general pattern) and no similar effect was found for female teachers (and male students). The authors relate this to romantic and libidinous aspects in teacher-student interactions in the higher grades of secondary education, which are often considered taboo. Nevertheless, Einarsson and Granström (2002) criticize previous studies for not having “paid attention to sexual temptations in the teaching profession” (p. 119). Not only grade level and age of students but also other aspects could help in detecting possible conditional teacher gender effects. Brophy in 1985(b) sums up several elements that researchers have to take into account as influencing classroom interactions such as group size, time of year and nature of the subject content taught. Again, little systematic research has been conducted focusing on these elements in specific so it is difficult to draw any general conclusions. In 1997 Howe puts forward one central limitation, namely that the existing evidence base in secondary schools has predominantly focused on mathematics and science. In 2006 Beaman et al. still identify this need for research in high school classes to be carried out over a broader range of discipline areas.
3.2 Research focus

The empirical study presented in this paper is aimed at closing some of the gaps in the existing evidence-base.

First – and in line with the closing recommendation of Beaman et al. (2006) in the previous section – the study is performed in secondary education over a broad range of disciplines (mathematics, sciences, mother tongue and second language courses). The first year of secondary education is studied (11 to 12 years old) so no major sexual temptations are expected.

Second, the design is set up to focus on student variable effects and not on teacher variable effects since previous research has provided too few indications of significant effects in this regard to formulate clear hypotheses. In 2013 Howe and Abedin stated that the effects of student variables such as gender, attainment and ethnicity as predictors of classroom dialogue have been studied sufficiently and should no longer be a focus of research. We however do not share this opinion. Recent studies have clearly shown the complexity and conditional character of gendered interactions (e.g. “some” boys receiving more negative interactions). In order to better grasp these complexities (e.g. which boys in specific) we suggest furthering research on student variables, granted that a proper intersectionality informed approach be used (Collins, 1998; Hankivsky, 2014; Rouhani, 2014). In this intersectionality approach, background variables such as gender, ethnicity or socio-economic status are not studied as distinctive social categories of which the effects can merely be added. Social categories are investigated as mutually constructing one another and the intersects of axes of diversity are studied in depth. This approach will enable moving forward in a more fine-grained study of student variable effects in teacher-student classroom interactions. We will focus on the following student background variables: gender, reading ability, math ability, socio-economic status, migration background and foreign language background. This last variable is included to meet the need for new variables to be investigated in regard to gendered classroom interactions. Foreign language background has struck us as not being included in previous research, notwithstanding its self-evident relationship with students’ abilities to interact with their teachers.

Third, the presented study will not only investigate teacher-initiated behaviours towards students (including a focus on positive teacher feedback,
which has produced inconclusive findings in the past). The study will also include student-initiated behaviours and teacher reactions to student behaviour. This will provide us with the opportunity to examine – to some extent – the dynamic relationship between student behaviour and teacher response, as discussed in the previous sections. Also, student perceptions of equity in the classroom environment will be investigated. This triangulation of perspectives and mixed methods (Johnson, Onwuegbuzie, & Turner, 2007) allow us to verify the congruency between actual equity of interaction patterns and perceived equity of interactions. Figure 8 illustrates the hypothesized relationships under scrutiny in this study.

Figure 8. Hypothesised model of student background variables predicting sense of equity and teacher-student classroom interactions.

3.3 Method

All data collection was carried out in the period October – November 2012. First, questionnaires were used to gather information concerning pupils' socio-demographic background and their perception of the classroom environment. Second, standardised tests measuring mathematics and reading ability were administered. Third, video-recorded classroom observations were performed to collect data on teacher-pupil classroom interactions.
3.3.1 Sample

Survey and standardised tests sample. A representative sample of 59 schools in Flanders (Dutch speaking part of Belgium) was selected to participate in the study. Schools were selected using disproportionate stratified sampling with three strata: (1) geographical location in one of five provinces and the capital city, (2) rural and urban area, and (3) public and private educational network, the latter principally organized by groups affiliated with the Catholic Church. All first-year secondary students of these 59 schools were selected to participate in the survey. In total, 6380 pupils (response rate 97%) completed the questionnaire. The mean age of all pupils when taking the survey and the tests was twelve years old (SD=.51). Table 1 provides an overview of the number and percentage of pupils for four axes of diversity: gender, socio-economic background, foreign language and migrant background and its intersecting categories. In total, 6139 pupils participated in a mathematics test (response rate of 93%) and 5897 pupils participated in a reading ability test (response rate 89%).

Classroom observation sample. A purposeful sample of six schools from among the 59 schools was selected for additional data collection through the video recording of lessons. The six schools involved were public and private schools in urban and suburban areas: two small, two medium-sized and two large-sized schools. The schools also differed in the types of secondary education they organize: only general; vocational and technical; and both general and vocational and technical. With support from the principal, a team of five teachers who volunteered to participate was composed in each school. Teachers had to be teaching the first year of secondary education in general subjects. For each teacher, one of their first-year secondary classes was selected and one lesson was video-recorded with two cameras (positioned opposite one another in two corners of the classroom) by one female researcher. In total, 30 teachers and their 500 pupils were video-recorded. Half of the 500 pupils was randomly selected for student-centered coding in which a single subject is observed continuously. Due to practical reasons (out of sight, quality of recordings) the teacher-student interactions of 180 pupils or 36% of all 500 pupils was coded. The
mean age of the observed students was equal to the mean age of the total survey sample, twelve years old. Table 10 presents the number and percentage of observed pupils along the four axes of diversity and its intersects.
### Table 10. Survey and observation sample, by gender, SES, foreign language background and migration background

<table>
<thead>
<tr>
<th></th>
<th>Gender</th>
<th>Socio-economic background</th>
<th>Foreign language background</th>
<th>Migration background</th>
<th>Observation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Boy</td>
<td>Girl</td>
<td>Low</td>
<td>Mid/High</td>
</tr>
<tr>
<td>Gender</td>
<td>Boy</td>
<td>783</td>
<td>2541</td>
<td>12.3</td>
<td>39.8</td>
</tr>
<tr>
<td></td>
<td>Girl</td>
<td>580</td>
<td>2290</td>
<td>9.1</td>
<td>35.9</td>
</tr>
<tr>
<td>Socio-economic</td>
<td>Low</td>
<td>17</td>
<td>19</td>
<td>10.0</td>
<td>11.1</td>
</tr>
<tr>
<td>background</td>
<td>Mid/High</td>
<td>66</td>
<td>72</td>
<td>33.7</td>
<td>40.0</td>
</tr>
<tr>
<td>Foreign language</td>
<td>No</td>
<td>68</td>
<td>74</td>
<td>37.8</td>
<td>41.1</td>
</tr>
<tr>
<td>background</td>
<td>Yes</td>
<td>18</td>
<td>18</td>
<td>10.0</td>
<td>10.1</td>
</tr>
<tr>
<td>Migration</td>
<td>No</td>
<td>57</td>
<td>68</td>
<td>31.7</td>
<td>37.8</td>
</tr>
<tr>
<td>background</td>
<td>Yes</td>
<td>25</td>
<td>20</td>
<td>13.9</td>
<td>11.1</td>
</tr>
<tr>
<td>Survey</td>
<td>Total</td>
<td>86</td>
<td>94</td>
<td>47.8</td>
<td>52.2</td>
</tr>
<tr>
<td></td>
<td>Missing</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
<td>3.3</td>
</tr>
</tbody>
</table>

*Note.* The upper half of the table gives frequencies for the survey sample. $N_{\text{survey}} = 6380$. The bottom half of the table gives frequencies for the observation sample. $N_{\text{observation}} = 180$. For each cell the upper figure is the absolute frequency, the bottom figure is the relative frequency by the total N.
3.3.2 Instruments

Survey

Questionnaires were administered by the research team during class. Pupils were asked about their socio-demographic background and were asked to rate the extent to which students are treated equally by the teachers. To measure socio-economic status, pupils were asked for the last occupation of their mother and father and the answers were converted into 8 categories of occupational prestige based on the Erikson, Goldthorpe and Portocarero (1979) classification. The highest of both parents was used as an indicator of the SES of pupils. For the analyses, SES was dichotomised into the following categories: (a) unemployed or manual workers and (b) routine non-manual employees up to professionals, entrepreneurs and large proprietors. Migrant background was operationalised by segregating pupils whose grandmothers from mothers' side originated from a non-Western-European country. Foreign language background was measured by asking pupils which language they usually speak with their parents. If no Dutch was spoken with either of the parents, students were categorised as being from a foreign language background. To measure sense of equity the 6-item equity subscale of the What Is Happening In This Class? (WIHIC) survey was used (Dorman, 2003). The WIHIC is a recently-developed and widely validated instrument that assesses classroom psychosocial environment in high schools. The scale is developed in such a way that it does not pre-define the basis for the discrimination that pupils' experience. All items and their Dutch translation can be found in Appendix 1. Sample item: “I get the same opportunity to contribute to class discussion as other students”. All items are scored 0, 1, 2, 3, and 4, respectively, for the responses of almost never, seldom, sometimes, often and almost always. A principal component analysis was performed to verify the underlying structure of the construct and the reliability of the scale was checked (\(\alpha = .865\)).

Standardised tests

Math ability was measured using the Dudal-orientation test Middle 6th Grade (Dudal, 2003). The test comprises 50 exercises and measures the knowledge
of numbers, mental calculation, arithmetic, mathematical problems, and measures. Reading ability was measured using the CITO-test Group 8 (CITO, 2010). The test presents students with text fragments, followed by multiple-choice questions. The test is comprised of 25 questions that required students to retrieve and summarize information from the text, identify the goal of the text, describe relationships and define concepts. Both tests are aimed at assessing student ability at the end of primary education, and thus the initial ability of students when entering the first year of secondary school. The standardised tests were provided to the 59 schools along with short manuals with clear instructions on how to correctly administer the tests. Mathematics and Dutch teachers were asked to administer the tests during regular classes. Some teachers were not able, or refused, to participate.

**Classroom observation schedule**

Two educational researchers acted as the observers, using continuous focal coding for 15-minute video-recorded fragments. The middle 15 minutes of all video-recorded lessons were selected for coding teacher-student interactions since these minutes were guaranteed to be instruction time. The first and last fifteen minutes of a 50-minute lesson are traditionally used to enter and leave the classroom, fill in the school diary, hand out corrected assignments, and take quizzes. A content analysis of pupils' classroom interactions was performed using ten codes and a code for missing/out of sight (see Table 11). The codebook was based on the recurring codes in three recent studies (cf. Harrop & Swinson, 2011; Myhill, 2002; Younger & Warrington, 2002). Teacher-initiated behaviours as well as student-initiated behaviours were coded. Detailed selection criteria were consistently applied to determine if a behaviour could be assigned to a certain category. Several iterative cycles of coding, assessment of reliability and codebook modification were conducted by the two coders until reliability for all codes was high (Hruschka et al., 2004). To ensure that high levels of observer agreement would not be mere reflections of high chance agreement, a Krippendorff's alpha (Krippendorff, 2013) was calculated. As Krippendorff's alpha is known to be a more conservative index (Lombard et al., 2002), a more liberal criterion of .60 was used to determine if the coding of a variable was considered reliable. Training continued until kalpha's were above .60 and from that point on the
two coders scored the remaining tapes, being aware that five lessons, taken at random, would be scored by the other observer at the end. Table 11 gives an overview of the mean kalpa's for each code for both reliability checks combined. In line with recommendations of Harrop and Swinson (2007) we present a contingency table so that agreements and disagreements on occurrences and non-occurrences of behaviour recorded by the two observers can be inspected. The C, H, T, R, N and P categories were analysed as events, generating the frequency or occurrence of each behaviour for each student. The O, F, I and S categories were analysed as a state, generating the total duration of each behaviour as a proportion of time in sight for each student. The h code was also coded as a mean duration state (mean duration of hand up). A P-N code was calculated by subtracting the frequency of negative feedback from the positive feedback. Coding was performed using JWatcher (Blumstein, Daniel, & Evans, 2012). More information concerning the observation schedule and protocol can be found in Appendix 2.

Table 11. Observation codebook and intercoder reliability (Krippendorff’s alpha, kalpha) for each code

<table>
<thead>
<tr>
<th>Code</th>
<th>Short definition</th>
<th>C1 (0)</th>
<th>C1 (0)</th>
<th>C1 (1)</th>
<th>C1 (1)</th>
<th>kalpha’s</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>Call out to teacher</td>
<td>1089</td>
<td>20</td>
<td>18</td>
<td>73</td>
<td>.78</td>
</tr>
<tr>
<td>H</td>
<td>Hand up</td>
<td>1078</td>
<td>6</td>
<td>3</td>
<td>113</td>
<td>.96</td>
</tr>
<tr>
<td>T</td>
<td>Turn from teacher</td>
<td>1147</td>
<td>9</td>
<td>8</td>
<td>36</td>
<td>.80</td>
</tr>
<tr>
<td>R</td>
<td>Response to question or comment from teacher</td>
<td>1137</td>
<td>17</td>
<td>11</td>
<td>35</td>
<td>.70</td>
</tr>
<tr>
<td>N</td>
<td>Negative feedback from teacher</td>
<td>1169</td>
<td>5</td>
<td>9</td>
<td>17</td>
<td>.70</td>
</tr>
<tr>
<td>P</td>
<td>Positive and neutral feedback from teacher</td>
<td>1170</td>
<td>1</td>
<td>11</td>
<td>18</td>
<td>.75</td>
</tr>
<tr>
<td>I</td>
<td>Individual instruction from teacher</td>
<td>1172</td>
<td>3</td>
<td>11</td>
<td>14</td>
<td>.66</td>
</tr>
<tr>
<td>S</td>
<td>Social behaviour without permission</td>
<td>971</td>
<td>53</td>
<td>28</td>
<td>148</td>
<td>.75</td>
</tr>
<tr>
<td>O</td>
<td>On-task behaviour</td>
<td>145</td>
<td>35</td>
<td>78</td>
<td>942</td>
<td>.66</td>
</tr>
<tr>
<td>F</td>
<td>Off-task behaviour</td>
<td>955</td>
<td>94</td>
<td>34</td>
<td>117</td>
<td>.58</td>
</tr>
</tbody>
</table>

Note. C1 (0) = coder 1 recorded non-occurrence. C2 (1) = coder 2 recorded occurrence.

3.3.3 Analysis

Non-normal observational data

Normality of the data was assessed by looking at the histograms and QQ-plots, and by running a Kolmogorov-Smirnov and Shapiro-Wilk test. The observational data show a strongly skewed distribution. Many behaviours have values close to zero (e.g. many students never call out to their teacher) or a have a natural limit (e.g. number of times students raise their hand during
a fifteen minute fragment). In an attempt to reduce the skewness, both a log transformation and a square root transformation of the data were performed. None of the procedures was successful at making the data normal enough for parametric statistics, so non-parametric tests were used to analyse the observational data.

Three steps of two-stages

The effect of the socio-demographic background variables and math and reading ability on the two dependent variables is assessed in separate analyses. First, the effects of the independent variables on sense of equity will be investigated. Second, differences in actual classroom interactions will be investigated for the set of independent variables. In a final and third step, the relationship between perceived equity of interactions and actual interactions will be assessed. In each step of analyses, a two-stage analytical strategy will be used (Rouhani, 2014) to allow for a systematic analysis of intersectionality. First, traditional additive approaches are used to examine the individual effects of each of the demographic background variables on the dependent variable. In the further stage of analysis, a multiplicative approach is used to account for the conditional effects of intersecting categories. For the survey data a multiple regression analysis is performed adding two-way interaction terms of the demographic categories to the equation predicting sense of equity. Since the observational data are non-normal, no regression analysis with interaction terms could be performed on these data. As an alternative, six new variables were constructed for each intersect of two dimensions of diversity. The intersect of gender and SES, for example, resulted in one variable with four categories: low SES boy, middle/high SES boy, low SES girl, middle/high SES girl. Differences for the behaviours between these intersecting categories were assessed. When investigating the relationship between sense of equity and actual interactions, split case correlation analyses are performed to investigate differences in correlational patterns for several (intersecting) groups.
3.4 Results

3.4.1 Perception of interactions: sense of equity in the classroom

*Additive analysis.* Independent samples t-tests show that significant differences exist for all axes of demographic diversity in sense of equity. Boys experience less equity than girls ($t(5771.120)=-9.39$, $p=.000$, $d=-.25$, $M_b=3.15$, $SD_b=.78$, $M_g=3.33$, $SD_g=.69$) and so do low SES pupils ($t(1666.567)=-8.86$, $p=.000$, $d=-.43$, $M_l=3.06$, $SD_l=.93$, $M_{m/h}=3.30$, $SD_{m/h}=.71$), pupils from a foreign language background ($t(537.962)=10.90$, $p=.000$, $d=.90$, $M_n=3.28$, $SD_n=.72$, $M_y=2.84$, $SD_y=.87$) and pupils from a migrant background ($t(1414.617)=11.82$, $p=.000$, $d=.63$, $M_n=3.31$, $SD_n=.70$, $M_y=2.99$, $SD_y=.82$).

Before running the regression models, all assumptions are checked. Firstly, the sample size is deemed adequate given the number of independent variables to be included. The collinearity statistics are well within accepted limits, so the assumption of multicollinearity is met. Residual and scatter plots indicate the assumptions of normality, linearity and homoscedasticity are all satisfied.

A multiple regression analysis predicting sense of equity assuming that all demographic categories are mutually exclusive is performed. Math and reading ability are entered in the model in a separate step, in order to control them for the four socio-demographic background variables. The additive Model 1 is statistically significant ($F(6,4890)=83.312$, $p<.000$), and accounts for approximately 9 percent of the variance in equity. Equity is primarily predicted by reading ability ($\beta=.164$, $p=.000$) and to a lesser extent by gender ($\beta=.112$, $p=.000$), math ability ($\beta=.089$, $p=.000$), language background ($\beta=-.070$, $p=.000$) and migration background ($\beta=-.064$, $p=.000$). Regression statistics can be found in Table 12. After inclusion of ability, the effect of SES and ethnicity decreases. This can point to a partial mediation of SES and ethnicity through ability.
<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE B</th>
<th>β</th>
<th>B</th>
<th>SE B</th>
<th>β</th>
<th>B</th>
<th>SE B</th>
<th>β</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Step 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender (1)</td>
<td>.183</td>
<td>.020</td>
<td>.125**</td>
<td>.163</td>
<td>.020</td>
<td>.112**</td>
<td>.192</td>
<td>.049</td>
<td>.132**</td>
</tr>
<tr>
<td>SES (2)</td>
<td>.028</td>
<td>.006</td>
<td>.071***</td>
<td>.000</td>
<td>.006</td>
<td>.001</td>
<td>.005</td>
<td>.007</td>
<td>.014</td>
</tr>
<tr>
<td>Language (3)</td>
<td>-.228</td>
<td>.046</td>
<td>-.080***</td>
<td>-.201</td>
<td>.045</td>
<td>-.070***</td>
<td>-.776</td>
<td>.131</td>
<td>-.272***</td>
</tr>
<tr>
<td>Migration (4)</td>
<td>-.181</td>
<td>.033</td>
<td>-.092***</td>
<td>-.126</td>
<td>.032</td>
<td>-.064***</td>
<td>-.155</td>
<td>.059</td>
<td>-.078**</td>
</tr>
<tr>
<td><strong>Step 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math</td>
<td>.006</td>
<td>.001</td>
<td>.089***</td>
<td>.006</td>
<td>.001</td>
<td>.089***</td>
<td>.023</td>
<td>.002</td>
<td>.161***</td>
</tr>
<tr>
<td>Reading</td>
<td>.023</td>
<td>.002</td>
<td>.164***</td>
<td>.023</td>
<td>.002</td>
<td>.164***</td>
<td>.114</td>
<td>.011</td>
<td>.200***</td>
</tr>
<tr>
<td><strong>Step 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 x 2</td>
<td>-0.045</td>
<td>.049</td>
<td>-0.030</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 x 3</td>
<td>.096</td>
<td>.091</td>
<td>.023</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 x 4</td>
<td>.017</td>
<td>.064</td>
<td>.006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 x 3</td>
<td>.092</td>
<td>.095</td>
<td>.023</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 x 4</td>
<td>-.062</td>
<td>.065</td>
<td>-.023</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 x 4</td>
<td>.617</td>
<td>.114</td>
<td>.200***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R²</td>
<td>.05</td>
<td></td>
<td></td>
<td>.09</td>
<td></td>
<td></td>
<td>.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>64.965***</td>
<td></td>
<td>83.312***</td>
<td></td>
<td>44.725***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ΔR²</td>
<td>.05</td>
<td></td>
<td>.01</td>
<td>.04</td>
<td></td>
<td>.04</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note.** Gender is coded 0 = boy and 1 = girl. SES is coded 0 = low and 1 = middle/high. Foreign language background is coded 0 = no and 1 = yes. Migration background is coded 0 = no and 1 = yes.
**Multiplicative analysis.** In the multiplicative approach a model is tested including two-way interaction terms of the demographic categories in order to account for conditional effects. Results (see Table 12) show that the interaction terms in Model 2 add some (but little) explanatory power to the model \( F(12,4884)=44.725, \ p<.000, \ \Delta R^2=.01 \). The only interaction term that helps to explain variance in sense of equity is the interaction between migration and language background \((\beta=.200, \ p=.000)\). Equity is primarily predicted by language background \((\beta=-.272, \ p=.000)\), followed by the interaction term of language and migration background \((\beta=.200, \ p=.000)\), reading ability \((\beta=.161, \ p=.000)\) and gender \((\beta=.132, \ p=.000)\). None of the interaction terms with gender are significant predictors of sense of equity.

### 3.4.2 Actual interactions: teacher-student classroom interaction

**Additive analysis.** Mann-Whitney U Tests were performed to compare the scores of the different demographic variables for the different behaviours. Significant differences were found for five types of behaviours (see Table 13). Boys tend to call out, be off-task and receive negative feedback more often than girls. Pupils from a migrant and foreign language background raise their hand less often than non-migrant and Dutch speaking pupils. Also low SES, migrant and non-native speakers raise their hands for significantly shorter periods than respectively high SES, non-migrant and native speakers. Pupils from a migrant background are also less on task. All effect sizes \( r \) values are small to moderate.
Table 13. Results Mann-Whitney U test group comparisons for teacher-student interactions

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>SD</th>
<th>U</th>
<th>p</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Call out</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>2.47</td>
<td>5.47</td>
<td>3408.000</td>
<td>.044</td>
<td>.14</td>
</tr>
<tr>
<td>Female</td>
<td>2.75</td>
<td>2.75</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hand up (event)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Migration (no)</td>
<td>5.24</td>
<td>5.41</td>
<td>1414.617</td>
<td>.000</td>
<td>.20</td>
</tr>
<tr>
<td>Migration (yes)</td>
<td>3.33</td>
<td>4.07</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Foreign language (no)</td>
<td>5.10</td>
<td>5.27</td>
<td>1852.000</td>
<td>.010</td>
<td>.13</td>
</tr>
<tr>
<td>Foreign language (yes)</td>
<td>3.61</td>
<td>5.75</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hand up (mean duration)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low SES</td>
<td>2603.47</td>
<td>2727.27</td>
<td>1857.000</td>
<td>.029</td>
<td>-.22</td>
</tr>
<tr>
<td>Middle/High SES</td>
<td>4150.51</td>
<td>3891.42</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Migration (no)</td>
<td>4512.79</td>
<td>3782.65</td>
<td>2120.000</td>
<td>.011</td>
<td>.36</td>
</tr>
<tr>
<td>Migration (yes)</td>
<td>2029.71</td>
<td>2501.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Foreign language (no)</td>
<td>4441.65</td>
<td>3821.000</td>
<td>1185.000</td>
<td>.000</td>
<td>.45</td>
</tr>
<tr>
<td>Foreign language (yes)</td>
<td>1434.43</td>
<td>1675.61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>On-task</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Migration (no)</td>
<td>.83</td>
<td>.12</td>
<td>2192.500</td>
<td>.046</td>
<td>-.07</td>
</tr>
<tr>
<td>Migration (yes)</td>
<td>.85</td>
<td>.17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Off-task</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>.07</td>
<td>.09</td>
<td>3118.000</td>
<td>.019</td>
<td>.17</td>
</tr>
<tr>
<td>Female</td>
<td>.04</td>
<td>.08</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative feedback</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1.01</td>
<td>1.80</td>
<td>3392.000</td>
<td>.034</td>
<td>.35</td>
</tr>
<tr>
<td>Female</td>
<td>.50</td>
<td>.83</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Multiplicative approach. Kruskal-Wallis H Tests were performed to check if there are statistically significant differences for the behaviours between two or more intersecting groups of each pair of crossing axes of diversity. Post hoc Mann-Whitney tests on each pair of groups were performed to assess which specific groups are statistically significantly different from each other. Only the differences for the different gender subgroups will be addressed. Girls with a foreign language background participate least actively in the classroom interactions. They score significantly lower than the other groups for raising hands ($\chi^2(3)=13.989$, $p=.003$), mean duration of raising hands ($\chi^2(3)=27.157$, $p=.000$), getting a turn ($\chi^2(3)=9.189$, $p=.027$) and responding ($\chi^2(3)=9.891$, $p=.020$). Also, whereas girls in general are less off task than boys, this is not the case for girls with a foreign language background ($\chi^2(3)=9.362$, $p=.025$) who are significantly more off task than all other girls. Boys who are significantly more off task than other groups are boys with a
middle to high SES ($\chi^2(3)=9.993, p=.019$) and autochthonous boys ($\chi^2(3)=12.065, p=.007$). In the groups of students with low SES no significant gender difference was found for off-task behaviour.

### 3.4.3 Relationship between perceived and actual interactions

*Additive approach.* A spearman rho correlation analysis was performed for testing the relationship between all behaviours, math and reading ability and sense of equity. Students who report higher levels of equality in the classroom environment are those students who score higher in reading ($r_s(157)=.290, p=.000$) and math initial ability ($r_s(164)=.204, p=.009$) and who raise their hand more often ($r_s(174)=.234, p=.002$), raise their hand for a longer period ($r_s(171)=.236, p=.002$), who get a turn more often ($r_s(174)=.176, p=.002$) and respond more frequently ($r_s(174)=.189, p=.012$).

Table 14 presents the results of two split case correlation analyses exploring differences in correlational patterns in boys and girls. The above described correlations with sense of equity appear only to remain standing for girls. Boys’ sense of equity is not significantly correlated with any of the behaviours nor with the initial ability measures.
Table 14. Spearman rho correlation matrix for boys and girls

<table>
<thead>
<tr>
<th></th>
<th>1.</th>
<th>2.</th>
<th>3.</th>
<th>4.</th>
<th>5.</th>
<th>6.</th>
<th>7.</th>
<th>8.</th>
<th>9.</th>
<th>10.</th>
<th>11.</th>
<th>12.</th>
<th>13.</th>
<th>14.</th>
<th>15.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. call out</td>
<td>--</td>
<td>.060</td>
<td>.024</td>
<td>.180</td>
<td>.176</td>
<td>.326**</td>
<td>.300**</td>
<td>-.012</td>
<td>.224*</td>
<td>.188</td>
<td>-.316</td>
<td>.102</td>
<td>-.298**</td>
<td>-.133</td>
<td>-.045</td>
</tr>
<tr>
<td>2. hand up</td>
<td>.182</td>
<td>--</td>
<td>.689**</td>
<td>.512**</td>
<td>.533**</td>
<td>.047</td>
<td>.265**</td>
<td>.147</td>
<td>-.209*</td>
<td>-.157</td>
<td>-.117</td>
<td>.019</td>
<td>.260*</td>
<td>.429**</td>
<td>.359**</td>
</tr>
<tr>
<td>3. hand up dur</td>
<td>.181</td>
<td>.542**</td>
<td>--</td>
<td>.308**</td>
<td>.371**</td>
<td>.071</td>
<td>.201</td>
<td>.080</td>
<td>-.136</td>
<td>-.168</td>
<td>-.205</td>
<td>.065</td>
<td>.134</td>
<td>.315**</td>
<td>.309**</td>
</tr>
<tr>
<td>4. turn</td>
<td>.326**</td>
<td>.519**</td>
<td>.128</td>
<td>--</td>
<td>.966**</td>
<td>.325**</td>
<td>.578**</td>
<td>.229*</td>
<td>-.022</td>
<td>-.046</td>
<td>-.169</td>
<td>-.173</td>
<td>.029</td>
<td>.191</td>
<td>.188</td>
</tr>
<tr>
<td>5. response</td>
<td>.374**</td>
<td>.528**</td>
<td>.182</td>
<td>.925**</td>
<td>--</td>
<td>.360**</td>
<td>.624**</td>
<td>.265**</td>
<td>-.007</td>
<td>-.012</td>
<td>-.208*</td>
<td>-.184</td>
<td>.007</td>
<td>.167</td>
<td>.255*</td>
</tr>
<tr>
<td>6. negative</td>
<td>.554**</td>
<td>.097</td>
<td>.040</td>
<td>.229*</td>
<td>.335**</td>
<td>--</td>
<td>.253*</td>
<td>-.499**</td>
<td>.278**</td>
<td>.063</td>
<td>-.290**</td>
<td>-.008</td>
<td>-.113</td>
<td>.204</td>
<td>.073</td>
</tr>
<tr>
<td>7. positive</td>
<td>.349**</td>
<td>.177</td>
<td>.061</td>
<td>.519**</td>
<td>.562**</td>
<td>.113</td>
<td>--</td>
<td>.653**</td>
<td>.148</td>
<td>.059</td>
<td>-.204</td>
<td>-.184</td>
<td>-.090</td>
<td>-.003</td>
<td>.116</td>
</tr>
<tr>
<td>8. pos-neg</td>
<td>-.180</td>
<td>.022</td>
<td>-.021</td>
<td>.125</td>
<td>.077</td>
<td>-.715**</td>
<td>.532**</td>
<td>--</td>
<td>.021</td>
<td>.045</td>
<td>.034</td>
<td>-.229*</td>
<td>.009</td>
<td>-.131</td>
<td>.028</td>
</tr>
<tr>
<td>9. individual</td>
<td>.316**</td>
<td>-.171</td>
<td>-.099</td>
<td>.078</td>
<td>.091</td>
<td>.224*</td>
<td>.175</td>
<td>-.004</td>
<td>--</td>
<td>.238*</td>
<td>-.232*</td>
<td>-.132</td>
<td>-.315**</td>
<td>-.042</td>
<td>-.061</td>
</tr>
<tr>
<td>10. social</td>
<td>.532**</td>
<td>-.120</td>
<td>.024</td>
<td>.024</td>
<td>.020</td>
<td>.374**</td>
<td>.037</td>
<td>-.227*</td>
<td>.305**</td>
<td>--</td>
<td>-.520**</td>
<td>.150</td>
<td>-.134</td>
<td>-.109</td>
<td>.059</td>
</tr>
<tr>
<td>11. on task</td>
<td>-.536**</td>
<td>-.096</td>
<td>-.174</td>
<td>-.102</td>
<td>-.129</td>
<td>-.448**</td>
<td>-.035</td>
<td>.321**</td>
<td>-.419**</td>
<td>-.627**</td>
<td>--</td>
<td>-.315</td>
<td>.202</td>
<td>0</td>
<td>-.267</td>
</tr>
<tr>
<td>12. off task</td>
<td>.266*</td>
<td>-.103</td>
<td>.058</td>
<td>-.065</td>
<td>-.072</td>
<td>.300**</td>
<td>-.187</td>
<td>-.370**</td>
<td>.241*</td>
<td>.288**</td>
<td>-.759**</td>
<td>--</td>
<td>-.042</td>
<td>-.146</td>
<td>-.088</td>
</tr>
<tr>
<td>13. math</td>
<td>-.027</td>
<td>.079</td>
<td>.068</td>
<td>-.095</td>
<td>-.026</td>
<td>-.039</td>
<td>-.116</td>
<td>-.009</td>
<td>-.190</td>
<td>.128</td>
<td>.097</td>
<td>-.126</td>
<td>--</td>
<td>.549**</td>
<td>.223*</td>
</tr>
<tr>
<td>14. reading</td>
<td>-.141</td>
<td>.005</td>
<td>.184</td>
<td>-.101</td>
<td>-.046</td>
<td>-.192</td>
<td>-.015</td>
<td>.098</td>
<td>-.080</td>
<td>-.003</td>
<td>.071</td>
<td>-.124</td>
<td>.560**</td>
<td>--</td>
<td>.324**</td>
</tr>
<tr>
<td>15. equity</td>
<td>.042</td>
<td>.067</td>
<td>.137</td>
<td>.182</td>
<td>.117</td>
<td>-.102</td>
<td>.150</td>
<td>.150</td>
<td>.128</td>
<td>-.107</td>
<td>.115</td>
<td>-.141</td>
<td>.193</td>
<td>.216</td>
<td>--</td>
</tr>
</tbody>
</table>

Note. The upper half of the table is for girls. $N_{\text{girls}} = 94$. The bottom half of the table is for boys. $N_{\text{boys}} = 86$.

Coefficients in the top right corner are for the subsample of boys. Coefficients in the bottom lefts corner are for the subsample of girls.

* correlation is significant at the .01 level (2-tailed)

** correlation is significant at the .05 level (2-tailed)
Multiplicative approach. We conducted several split case correlation analyses for intersectional groups. We report only the results of subgroups of gender that differ considerably from the general pattern. For girls with a foreign language background, sense of equity was only significantly correlated with on task (r_s(18)=-.482, p=.043) and off task behaviour (r_s(18)=.473, p=.048). For middle/high SES boys sense of equity is correlated with individual instruction (r_s(66)=.272, p=.028) and social interaction with other students without permission (r_s(66)=.259, p=.037). No other significant correlations were found for this group of middle/high SES boys.

When exploring the behaviours with which negative teacher feedback is correlated, again different patterns emerge. Part of the general correlations are found for both girls and boys, such as a significant correlation with calling out (r_s(180)=.464, p=.000), getting a turn (r_s(180)=.280, p=.000), responding to teachers (r_s(180)=.326, p=.000), lower levels of on-task behaviour (r_s(177)=-.383, p=.000), higher levels of off task behaviour (r_s(177)=.189, p=.012) and individual instruction (r_s(177)=.249, p=.001). For boys in general however an extra correlation exists with unauthorised social behaviour with classmates (r_s(86)=.374, p=.000). For girls in general, this relationship does not exist. Girls do however talk and interact with their classmates as often as boys. For girls from a foreign language background and for boys from a middle/high SES the correlations with negative feedback follow the general pattern.

### 3.5 Discussion

#### 3.5.1 Which boys and which girls?

Since the 1990s researchers concur that not all boys as such but some boys in particular are responsible for boys’ increased levels of classroom participation and (negative) teacher feedback. The little research that has been performed on other variables than student gender alone suggests that high achieving students participate more actively in classroom interactions. With regards to ethnic minorities findings are inconclusive (Howe & Abedin, 2013). With regard to attainment, our findings for girls are in line with the general pattern of high achievers participating more (raising hands,
responding to teachers). For boys however this pattern is not reproduced. No correlation exists with any of the behaviours observed. Myhill’s (2002) work can shed some light on this phenomenon. Myhill found that towards middle school high achieving boys tend to join the pattern of interaction of underachievers. Based on qualitative analyses she relates this to boys’ construction of masculinity and peer group pressure not to appear too eager or enthusiastic in class. This hypothesis deserves more attention in future research since our findings appear to support it.

With regard to girls our intersectionality analyses paid off by bringing to the surface a group of girls that appears to be exceptionally vulnerable, namely girls from a foreign language background (not speaking the instruction language with any of their parents). We observed that these girls (unlike girls in general) are off-task significantly more often and they raise their hand less often and for less long, they get a turn from teachers less often and they respond less to questions. Language background did not only interact with gender, it was also in itself the strongest predictor of sense of equity with a very high practical relevance (d=.90) for differences in sense of equity. We suggest continuing research on the effects of language background and classroom interactions. Multidisciplinary cooperations with language researchers can increase our understanding of these effects.

3.5.2 Teacher bias or disruptive boys?

In the introduction of this chapter we discussed an incongruence between student perceptions of teacher gender bias (Younger, Warrinton, & Williams, 1999) and the review of Howe (1997) claiming that boys’ higher levels of (negative) feedback are justified due to their higher levels of disruptive behaviour. Our findings confirm previous research that boys feel discriminated against more often than girls and they receive significantly more negative feedback than their female peers. We do not find any significant differences between boys and girls with regard to positive and neutral teacher feedback nor in the ratio of positive to negative feedback. Our study confirms that boys’ increased negative feedback appears to be – at least in part – justified by their calling out and being off-task more often. However, we also observe some significant differences in the behaviours with which negative feedback correlates. Boys and girls do not differ in their
levels of talking with classmates without permission from the teacher, but for girls this disruptive behaviour does not relate to higher levels of negative teacher feedback while for boys it does. This could indicate some degree of unfair differential teacher treatment as a result of teachers’ differential attention for boys and girls (Consuegra, Engels, & Willeghems, in press). Boys might be aware of this teacher bias, which could explain their lower levels over sense of equity in the classroom environment. However, the fact that we do not see any significant correlations between boys’ sense of equity and their classroom interactions could indicate a sense of discrimination that is not the direct result of teacher differential treatment. All boys and not only boys attracting unfair negative attention might be impacted by an accumulation of disproportionate negative feedback to boys over a longer period of time. We agree with Beaman, Wheldall and Kemp (2006) that the issue of (unfair) negative teacher feedback to boys is not one to be dismissed readily.

3.5.3 Teaching implications?

Howe (1997) critiques that virtually all of the classroom interaction research is limited to descriptions of interactions that take place, rather than measuring effects of these differential interactions on student perceptions, academic performance or social attitudes. This cross-sectional study is descriptive and correlational as well so we need to be careful with formulating suggestions for teaching. It is not unthinkable that recommendations end up being counterproductive due to the use of a model that is too simplistic and fails to meet the complexity of teacher expectancy, teacher treatment, student perceptions and student behaviour effects. In 1993 Bailey suggested – based on mainly descriptive studies – techniques such as deliberate turn-taking and increasing wait-time before calling on students (which would increase the number of female volunteers) in order to achieve equal classroom interactions. Brophy in 1985(b) however already proposed that it is not always desirable to treat all students as equal. Optimal instruction, he argues, implies some degree of individualization of instruction. Our critique on Bailey’s suggestions would be that it fails to acknowledge the multidimensional identity of students. Our findings show that students from a foreign language background (and girls in specific) are hesitant to raise their
hand and they tend to drop their hand quickly after raising it. Introducing a longer waiting time might be interesting for girls in general, but this specific group of girls and other students from a foreign language background would most certainly not be served with it. Hence it is important to move beyond gender as an isolated category of analysis and to take into account simultaneous interactions between multiple dimensions of diversity. Brophy (1985b) suggests – and we follow his suggestions – that teachers should counter negative teacher expectancy effects by keeping expectations for individual students current and up to date by monitoring student progress closely and stressing present performance over past history. A suggestion that Bailey (1993) formulates, which fits very well with this idea of increased monitoring that Brophy describes, is sensitizing teachers about the mechanisms of gender bias and getting feedback on interaction patterns in their own classrooms. Olivares and Rosenthal (1992) propose similar interventions of teachers developing skills to become aware of their biased perceptions and elements in the learning environment which inhibit students’ development. The above recommendations advocate for an increased professionalism in the form of teacher awareness and monitoring skills rather than teacher-proof tips and tricks on how to handle boys and girls. We believe this is indeed the way to proceed given the gaps in evidence, since implementing changes without proof of effectiveness is a risky business and could result in unsought effects. The review by Plumm (2008) on technology in education and gender-biased interactions very well illustrates these risks. They explain how it had been assumed that technology for the most part was gender-neutral and that its introduction would undo gender-bias. However, as the review shows, biases have simply been reshaped and have in some cases even increased gender-bias in student perceptions, teacher perceptions and curricular materials. Plumm (2008) concludes that educators and researchers need to be very cautious when attempting to reduce or eliminate gender-bias from classroom interactions with clear-cut remedies, since the mechanisms are so complex that without solid empirical data it is hard to predict effectiveness (or counter-productivity) of interventions.
3.6 Conclusion

The findings presented in this chapter point to significant interaction effects between student gender, socio-economic background, migration background and student foreign language background for participation in teacher-student interactions during whole-class teaching in a wide range of subjects in lower secondary education. Girls who do not speak the instruction language at home with one of their parents appear to be interacting least with their teachers (raising hand, getting a turn, responding to teacher). Boys from a middle to high SES and autochthonous boys are found to be most off-task. Boys in general receive more negative feedback from their teachers than girls but no significant gender differences were found for positive/neutral feedback and for the ratio between positive/neutral and negative feedback. Girls interact with their classmates without permission as much as boys. For boys but not for girls this behaviour correlates significantly with negative teacher feedback. This differential teacher treatment could in part explain boys’ significantly lower levels of sense of equity in the classroom environment. Given the descriptive and correlational nature of this study and given the demonstrated complexity of interactions between multiple dimensions of student diversity, we are reluctant to formulate clear-cut teaching strategies to handle boys and girls. We follow the suggestions of authors such as Brophy (1985b) and Olivares and Rosenthal (1992) who conclude that teachers need to control incorrect and negative expectation effect for students by increasing their awareness of bias mechanisms, by monitoring their differential treatment of students and by stressing present student achievement over previous student achievement.
Chapter 4
Feminised Teacher Professionalism
Chapter adapted from:
Abstract

Feminist theories argue that an increased rationality in teaching will suppress emotionality and educational theories argue that rationality is needed to tackle the supremacy of intuition in teaching. Rationality is thus claimed to threaten as well as foster teacher professionalism. This study aims to investigate the gendered character of ‘emotionality’ and ‘rationality’ in teacher professionalism and it’s effects on teacher-student relationships. The data of a representative sample of 1225 first-year secondary school teachers are analysed using structural equation modelling. Our results show that emotionality and rationality do not conflict and that both mediate – at least in part – teacher gender effects on the provision of structure, autonomy support and trust in students. Male and female teachers appear to rely on different strategies to achieve qualitative relationships, with men relying more on ‘rationality’ and women on ‘emotionality’. Implications for teacher education programs and teacher professional development are discussed.
4.1 Introduction

In 1990 Joan Acker concludes her widely cited paper on ‘gendered organizations’ by stating that “feminists wanting to theorise about organisations face a difficult task because of the deeply embedded gendering of both organizational processes and theory” (p. 154). She critiques a trend in theory on gender and work organisations that advocates for the creation of gender-neutral, asexual and disembodied organisations and states that gender is not merely an addition to ongoing - assumed to be gender-neutral - processes. She suggests not to enter gender in studies merely as a stereotypical analogy between femininity, oppression, weakness and passivity but rather to investigate gender as an integral part of processes of control, structures and relations in the workplace. Today, the critique of Acker about a lack of sophisticated gender theory substantiating empirical research in the field of work and organizations, remains relevant. Disconnections between feminist fields of study and adjacent critical but non-feminist disciplines of research exist. This multidisciplinary study aims at bridging one of these gaps by combining feminist theory and educational theory in a study investigating the gendered character of teaching and teachers’ perception of teacher-student relationships.

4.1.1 Gendered professions

When discussing gendered professions we find the work of Britton (2000) to be very helpful. Britton (2000) reviewed an extensive amount of literature and concludes that the labelling of a profession as gendered is still theoretically and empirically unclear. She discusses three of the most commonly employed interpretations of gendered work organisations. First, organisations and professions can be described to be gendered based on the extent to which they are numerically dominated by men or women. Teaching is generally known to be numerically feminised. The OECD (2012) average percentage of female teachers in kindergarten, primary, lower secondary, higher secondary and post-secondary education is respectively 97 percent, 82 percent, 68 percent, 56 percent and 54 percent. Second, an organisation can be perceived to be inherently gendered, implying that the organisation and
the work itself has been defined, conceptualised and structured in terms of a distinction between femininity and masculinity. In this perspective, male or female dominated professions can be gendered in ways that do not match their sex composition. In the case of teaching, all teachers (men and women) are ‘feminised’ since the job itself is already gendered and identified with ‘soft’, caring and emotional skills (Basten, 1997; Bolton & Muzio, 2008). Third, organisations and professions can be gendered because the discourse draws on *hegemonically defined* masculinities and femininities. Hegemonic masculinities and femininities refer to the at that point most celebrated and highly regarded ways of being a man or a woman (Connell & Messerschmidt, 2005). They do not necessarily embody the most prevalent masculinity or femininity but rather they represent the norm with which all others tend to compare themselves. Pulsford (2014) critiques the recent plea for a shift of the norm in teacher professionalism towards the hegemonic masculine traits of management, rationality, control and strategic knowledge. This norm reproduces gender inequality since female teachers have difficulties moving away from their gendered identities and women are associated with traits that are perceived to be unprofessional. As already mentioned, Britton (2000) identifies all approaches to gendered organisations as inadequate. She critiques that they immediately assume that gendered organisations produce and reproduce inequality and hierarchy. She refers to studies showing that increased bureaucracy in some cases appears to reduce gender inequality. Similar optimistic reflections and openings towards ‘positive’ gendered organisations as advocated by Britton can be found in the work of Connell and Messerschmidt (2005). Connell and Messerschmidt (2005) state that we have to acknowledge the possibility that a positive hegemonic masculinity could one day prevail, abolishing hierarchy and making equality of women and men the norm. Hegemonic masculinities and the gendered character of professions are theorised to be open to historical change leaving a window for less oppressive ways of doing gender, resulting in a decrease or even an abolition of gender hierarchies. In the following paragraphs we discuss teacher professionalism in the perspective of two utopias: first, we discuss the abolishment of the gap between emotionality and rationality (studied extensively in feminist literature); second, we discuss the abolishment of the divide between theory and practice (which has regained attention in educational sciences recently).
Finally we explore possible implications of (gendered) teacher professionalism with regard to the quality of teacher-student relationships.

4.1.2 Emotionality and rationality

Bolton and Muzio (2008) criticise the prevailing paradigm that claims that professional work needs to take place in an objective way, free of emotion and that teaching will never reach full professional status due to the image of teachers as a caring, nurturing, maternal figures providing emotional support. This critique has echoed in the field for quite a long time (e.g. Etzioni, 1969). In 1986 Apple theorised that the increasing rationalization of the work of educators will result in teachers being more dependent of prespecified curricula and tests and suffering a chronic sense of work overload. Self-direction and leisure will be lost and the risk of isolation will increase, leading to teachers having insufficient time to keep up with the developments in the field and thus de-professionalising. Forrester's 2005 analysis of UK primary education policy illustrates that qualities such as nurturing and caring are overshadowed by a culture of management, accountability, performance and monitoring and the friction between both results in increased stress and burnout in teachers. Research into the emotionality of teachers is growing, highlighting the impact of emotions on student learning, student-teacher relationships and the quality of education in general (Schutz & Zembyias, 2009). Recent studies show that female teachers anticipate on emotions more often and regulate their emotions in a better way than male teachers (Akin, Aydin, Erdogan, & Demirkasimogly, 2014). They will be more successful at, for example, ventilating sad or angry emotions in order to genuinely feel and act cheerfully when needed. Also, when faced with disruptive and disengaged students, they typically employ more emotion tactics to re-engage students (Demetriou, Wilson, & Winterbottom, 2009). Gender role conflict theory (O'Neil, 2008) provides further theoretical insight in the gendered character of emotionality. The theory states that men are taught to avoid emotions because these are considered not masculine. Rather than viewing emotions as human, emotions are equated with femininity and as a result men may repress their feelings by adopting a more instrumental and stoic approach. However, as men enter into teaching, they are expected to be emotionally expressive
and to better regulate and manage their feelings. Men may experience a conflict between the demands of their socialisation and the demands of their job situation (O’Neil & Lújan, 2009; Wester, Vogel, O’Neil, & Danforth, 2012). This ‘restricted emotionality’ is proposed as one of several factors underlying gender role conflicts. Empirical studies indicate that men who experience a more restricted emotionality report more violence and anger in relationships, decreased relationship intimacy, lower levels of social connectedness and more negative attitudes towards help-seeking behaviour (Blazina & Watkins, 1996; O’Neil, 2008). Zamarripa, Wampold and Gregory (2003) showed that gender role conflict is also present in women and the detrimental effects of restricted emotionality operate similarly in men and women. However, men did show to have higher levels of restricted emotionality. Pulsford (2014) recently suggested that further research on the study of emotions in (male) teachers is needed.

4.1.3 Theory and practice

In the previous section we related the critique that external control over the work of teachers should not be increased (Pulsford, 2014) to the tension between emotionality and rationality. This critique can also concern the long standing gap between theory and practice. Cochran-Smith and Lytle (1992) show that most teachers believe that the knowledge base for teaching is constructed by ‘experts’ such as university researchers. Teachers see themselves as the consumers of knowledge rather than as it’s creators. Many authors strongly criticise this practitioner-researcher and practice-theory gap (e.g. Broekkamp & Hout-Wolters, 2007; Grossman, 1989; Vanderlinde & van Braak, 2010; Korthagen & Kessels, 1999; Zeichner, 2010). Inspired by the principles of situated and constructivist learning, teachers are suggested to regain true autonomy and professionalism by self-regulating their learning. The concept of “teacher researcher” and teachers systematically inquiring into their own practice was proposed decades ago (see Dewey, 1933), but the amount of studies on the topic is peaking recently (e.g. Cochran-Smith, 2005; Cochran-Smith & Lytle, 1999, 2009; Gravani, 2008; Leitch & Day, 2000). In 2008 Mary Christianakis in a theoretical paper describes teacher research as a feminist act, upsetting hierarchy in educational knowledge production, much
like feminism upsets patriarchal hegemony. Practical and theoretical knowledge are integrated and co-exist on an equal plane. School-university and practitioner-researcher collaborations are seen as the only way to sustainably increase teacher professionalism and teaching quality (Kirschner, Kickinson, & Blosser, 1996; Noffke, Clark, Palmeri-Santiago, Sadler, & Shuja, 1996; Sim, 2010; Burbank & Kauchak, 2003; Butler et al., 2004), which is known to be one of the – if not the first – predictor of student learning (Hattie, 2009).

In order to realize teacher-researcher collaborations, a shift is needed in teacher thinking. Teachers' professional orientation (Beijaard, Meijer, & Verloop, 2004) thus needs to change: teachers' views on their roles, what they should know and how they should act. Teachers engaging in practitioner research, move from a “restricted” to an “extended” professional orientation. In the 1970s Hoyle first described this continuum (Hoyle, 1980). A restricted professional derives his skills from experience and intuition solely, has a narrow classroom-based orientation, values classroom autonomy and is unconcerned about theory. An extended professional derives his skills from rational activity, has a broad orientation including the context of education, values professional collaboration and seeks interaction between practical and theoretical knowledge. Recent empirical research supports this restricted/extended continuum (Evans, 2008; Lamote & Engels, 2010). No empirical studies are, however, known to the authors that discuss teacher gender differences in extended professionalism.

4.1.4 Gendered teacher-student relationship

One recurrent concern since the mid-1990s has been that a lack of male teachers may have negative consequences for the achievement and behaviour of boys in particular (Beaman et al., 2006). Relatively little is in fact known about the effects of ‘gender matching’ of teachers and pupils (Skelton et al., 2009). The few studies that have been performed indicate no major teacher gender effects. Helbig (2012), Driessen (2007), and Siongers (2002) all confirm that no teacher gender effects whatsoever exist on student achievement in primary and secondary education. Howe’s 1997 review on verbal and non-verbal classroom interactions concludes that there are few
indications that the gender of teachers makes a difference in classroom interaction patterns. Despite these results, some researchers do find teacher gender effects in very particular situations. Einarsson and Granström (2002), for example, found that male teachers’ attention for girls significantly increases in higher grades of secondary education and no such effect exists for female teachers and male students. The authors relate this to romantic and libidinous aspects of teacher-student relationships. Students themselves, also, report that teacher gender matters in terms of the construction of their own gender identities and the definition of ‘proper masculinity’ and ‘proper femininity’ (Skelton et al., 2009). Some students believe female teachers to be less strict and rigorous than male teachers. The effects of teacher gender might thus be subtle and of a qualitative nature. Male and female teachers might, for example, direct the same amount of questions to students but the type of questions could be different. Subtle differences in teacher-student relationships might exist which can not easily be grasped with quantitative observation schedules. Two well-established and fine-grained theories of teacher-student relationships are proposed to try and grasp any subtle teacher gender effects that might exist. The first framework is that of teacher need supportive teaching which is widely used as it is related to Deci and Ryan’s (1985) self-determination theory. Central in this theory is the assumption that three universal psychological needs are fundamental, namely the need for competence, autonomy and relatedness. With regard to teaching this is related to three dimensions of ‘need supportive’ teaching: teachers’ support of students’ autonomy and the provision of students with choice and the incorporation of their interests and preferences; teachers’ provision of structure by means of clarity, guidance, encouragement and feedback; and teachers’ involvement by showing affection, expressing attunement, dedicating time and being available for support. The extensive review of Stroet, Opdenakker and Minnaert (2013) on students’ and teachers’ perception of these dimensions of need supportive teaching shows a significant effect on early adolescents’ motivation and engagement for school. However, no teacher gender differences in need supportive teaching are mentioned. Deci, Spiegel, Ryan, Koestner, and Kauffman (1982) did find teacher gender effects in – again – very particular settings. In conditions in which the work of teachers is much controlled and structured with few autonomy for teachers, no teacher gender differences were found in the level
of supportiveness towards students. In noncontrolling conditions, where teachers have great autonomy and serve as a guide to students, the female teachers were judged much worse than their male colleagues. The second framework to investigate teacher-student relationships is teacher trust in students (Van Mæle, Forsyth, & Van Houtte, 2014). Trust is defined as “an individual’s or group’s willingness to be vulnerable to another party based on the confidence that the latter party is benevolent, reliable, competent, honest and open” (Hoy & Tschannen-Moran, 2003, p. 7). Higher levels of trust shown by teachers in students is known to promote the effectiveness of schools by promoting a culture in which students can succeed (Goddard, Tschannen-Moran, & Hoy, 2001). Again, no systematic teacher gender effects are reported but in some specific settings differences can be observed. Van Houtte (2007) found that in vocational education male as well as female teachers report more trust in girls than boys, but for male teachers the difference between teaching girls and boys is bigger than it is for female teachers. In general education this difference is not present.

4.1.5 Research focus

This study is aimed at investigating the effects of teacher gender on teacher-reported teacher-student relationship. Based on feminist and educational theory two mediators are hypothesised to carry the effect of gender: restricted emotionality and extended professionalism. It is unclear from the theoretical framework what the effects of the mediators will be on teacher-student relationship quality. On the one hand the increasing importance of rationalism, accountability and theory (present in teacher extended professionalism) is criticised to neglect and oppress emotionality in teaching with this conflict resulting in a loss of teaching quality (Forrester, 2005; Pulsford, 2014). On the other hand, the role of rationality, systematic inquiry and objectivity in teacher professionalism is heavily stressed in an attempt to diminish teachers’ ‘primacy of practice’ (Tillema, 2000) and the belief that teachers are merely consuming knowledge and not producing knowledge which impedes them to systematically improve their teaching quality (Cochran-Smith & Lytle, 1992). Arguments, thus, are provided for teacher extended professionalism and restricted emotionality to both promote and
hinder teaching quality. This study will investigate the gendered character of both mediators and their effects on need supportive teaching and trust in students.

4.2 Method

4.2.1 Mediation Analysis with Structural Equation Modelling

Since this study builds on an extensive amount of theory a quantitative design was chosen to verify and check the hypothesised relationships. This quantitative design will better allow for untangling both the principal relationships that are being studied and the confounding influences of variables which are relevant but not under central scrutiny in this study (Johnson & Onwuegbuzie, 2004). Also, we take the critique of Britton (2000) into account emphasizing the need for research on gendered organizations to bear in mind possible policy and professional uses of research findings. In this regard, a quantitative study and its conclusions might have higher credibility and impact on many people in power (e.g. administrators, politicians, funding agencies)(Johnson & Onwuegbuzie, 2004). We do understand that the focus on theory testing will result in missing out on opportunities for theory generation. In order to test the effect of multiple mediators and multiple outcomes simultaneously we use structural equation modelling (SEM). The advantage of estimating the entire model simultaneously is that we can learn if the mediation is independent of the effect of the other mediators (Kenny, 2014a).

4.2.2 Participants

A large and representative sample was selected for participation in this study. Data were collected in the school year 2012-2013 by means of written teacher questionnaires across a representative sample of 59 secondary schools in Flanders (the Dutch speaking part of Belgium). Schools were selected using disproportionate stratified sampling with three strata: (1) geographical location in one of five provinces and the capital city; (2) rural and urban area;
and (3) public and private educational network. In total, 1793 teacher questionnaires were distributed among the first-grade teachers (teaching students of approximately 13 and 14 years old) and 1247 completed questionnaires were returned – a response rate of 69.5 percent. Of the teachers responding to the survey, 71 percent were female and 28 percent were male. This is a slightly higher percentage of female teachers than the Flemish average in lower secondary education which is 68.1 percent (OECD, 2012). The average age of the respondents is 40 (SD=10.31), with an average of 14 years of teaching experience (SD=9.87). The majority of teachers have a bachelor degree in education, obtained after a three-year teacher education programme at a college of higher education that prepares for teaching in the grades 1 to 4 of secondary education. Eight percent of the respondents have a master's degree and were prepared for the teaching profession in a one-year postgraduate teacher education program. Teachers with these profiles are qualified to teach in grades 5 to 6 of secondary education. Seven percent of all male teachers report they have a master's degree compared to 9 percent of the female teachers. Equal percentages of men and women (28 percent) are mathematics and/or science teachers.

4.2.3 Instruments

This study adopts existing and previously validated scales to measure the central concepts of teacher extended professionalism, teacher restricted emotionality and teacher perception of teacher-student relationships. Each of the concepts will be discussed separately. An overview of all instruments and their items can be found in Appendix 1.

4.2.3.1 Teacher extended professionalism

In Hoyle's (1980) conceptualisation of extended professionalism we can identify the following dimensions: (a) a broad orientation towards the context of education that reaches beyond the classroom walls, (b) valuing professional collaboration and seeking interaction with theoretical knowledge (c) deriving skills not only from intuition but being critical towards one’s own habitual practices. For each of these three dimensions one instrument was selected.
Teachers’ broad orientation was measured using the *extra-role behaviour towards the students (EXT)* scale. This instrument was developed by Somech and Drach-Zahavy (2000) and measures behaviours that go beyond specific role requirements, that are voluntary and not formally rewarded and directed towards benefiting the organisation. Originally the scale comprised of 8 items but two items were left out because of a mismatch with the Flemish context. “Participating in private celebrations of my students (e.g., birthdays)” and “inviting students to visit me at my home” are very uncommon practices in Flanders and could even be perceived to be inappropriate. The abbreviated scale is made up of 6 items that are assessed by means of a 7-point Likert scale (0 = (almost) never, 6 = (almost) always). Sample item: “I stay late in school to help students with their lessons”.

Teachers’ tendency to seek interaction with colleagues and theoretical knowledge was measured using the *keeping up to date (KUD)* scale. This subscale from the Dutch School Improvement Questionnaire (see Geijsel, Sleegers, Stoel, & Krüger, 2009) measures the extent to which teachers collect new knowledge and information and keep up with developments in the field of education by reading professional literature and participating in non-mandatory trainings. The instrument consists of 4 items that are assessed by means of a 4-point Likert scale (0 = (almost) never, 3 = (almost) always). Sample item: “I read professional literature”.

Teachers’ ability to be critical towards one’s own habitual practices and intuition is measured using the *experimentation and reflective practice (EXP)* scale. This subscale from the Dutch School Improvement Questionnaire (see Geijsel et al., 2009) measures the extent to which teachers try out new things and/or undertake actions explicitly meant to change and improve their practices and/or enable reflection on their practice. The instrument consists of 5 items that are scored on a 4-point Likert scale (0 = (almost) never, 3 = (almost) always). Sample item: “I try out new knowledge and skills in my classes”.

### 4.2.3.2 Teacher restricted emotionality

Teacher restricted emotionality is measured using the *restricted emotionality (REM)* scale of Blazina, Pisecco and O’Neil (2005) which is an adapted version of the 37-item Gender Role Conflict Instrument developed in the 1980s (O’Neil, Helms, Gable, David, & Wrightsman, 1986). The adapted
measure presents respondents with 9 items related to a general restriction of affections that are assessed by means of a 6-point Likert scale (0 = strongly disagree, 5 = strongly agree). Sample item: “Intense emotions are hard for me to understand”.

4.2.3.3 Teacher perception of teacher-student relationship
Three dimensions of teacher-student relationship were selected based on the literature review. Provision of structure, autonomy support and trust in students were chosen considering previous research indicating a clear relationship with student motivation and student outcomes.

Provision of structure (STR). This 15-item subscale is part of the Teacher As Social Context Questionnaire (TASCQ) developed by Wellborn, Connel, Skinner, and Pierson (1992). The items tap into teachers’ reports of clarity of expectations, consistency, instrumental help and support and adjustment of teaching strategies to each child in the classroom. Statements are assessed on a 4-point Likert scale (0 = strongly disagree, 3 = strongly agree). Sample item: “I talk with students about my expectations for them”.

Autonomy support (AUT). This 12-item subscale of the TASCQ covers issues such as teacher controlling behaviour, respect and acknowledgement of students’ opinions, feelings and agendas, encouraging students to follow their own interests and providing them with options and providing a rationale for learning activities. Items are scored on a 4-point Likert scale (0 = strongly disagree, 3 = strongly agree). Sample item: “I try to give students a lot of choices about classroom assignments”.

Faculty Trust in Students (TRU). This 5-item instrument is a subscale of the Faculty Trust Scale of Hoy and Tschannen-Moran (2003) and measures teachers’ willingness to be vulnerable to students based on the confidence that they are benevolent, reliable, competent, honest and open. Items are scored on a 5-point Likert scale (0 = strongly disagree, 4 = strongly agree). Sample item: “I can count on students to do their work”.

4.2.4 Covariates
Using SEM to support complex models could result in misleading estimates if important variables are omitted (Myers, Well, & Lorch, 2010) or if causal
ordering of the variables is incorrect (Fairchild & MacKinnon 2009). With regard to the causal ordering of the variables we refer to the theoretical framework in which the hypothesised model is grounded. With regard to the omitted variables three important variables are included as covariates in the model: years of teaching experience, qualification (1 = master’s degree; 0 = bachelor’s degree), and subject taught (1 = mathematics and/or science, 0 = other). Gender, our central independent variable, is coded 0 = man and 1 = woman.

4.2.5 Preliminary analyses

Prior to testing the model, the relevant assumptions are tested. The sample size is deemed adequate for one sample analysis with an acceptable ratio of more than 10 participants per estimated parameter (Schreiber, Nora, Stage, Barlow, & King, 2006). Twenty-two extreme univariate outliers were observed in initial data screening and were excluded, leaving a final sample size of 1225 (1247 minus 22). The missing patterns are handled by using full-information maximum likelihood (FIML).

Assumptions of multivariate normality were tested using the Henze-Zirkler test and the Royston test. Both tests indicated that the assumption of multivariate non-normality had to be rejected (Henze-Zirkler test: 1.54, \( p > .001 \); Royston test: 361,32, \( p > .001 \)). We therefore conducted a follow-up analysis of univariate skewness and kurtosis. Results demonstrated that, although some items exhibited a skewness-values higher than 1 (rem2, rem6, kud1), none exceeded 2. For kurtosis, some items exhibited a value above 1 (str2, str6, str14, str15, aut5, aut10, aut11, rem2, rem6). In addition, two items exhibited a value above 2 (str13 and tru5). The latter two items were removed from the subsequent analyses. To accommodate for the remaining mild problems with skewness and kurtosis we opted for a maximum likelihood estimator with robust standard errors as it is better suited to handle this type of non-normality (Kline, 2005).
4.3 Results

4.3.1 Measurement model

In order to determine the adequacy of the measurement model portion of the structural equation model we performed confirmatory factor analyses (CFA) with the Lavaan 0.5-17 R Package (Rosseel, 2012). The seven latent constructs were tested separately. An overview of the comparative fit index (CFI), the Tucker-Lewis fit index (TLI), the RMSEA and the SRMR is provided for each construct in Table 15. A CFI and TLI value of .95 or higher was used to indicate a good model, while a value over .90 was used as an indication of acceptability. Similarly RMSEA and SRMR cut-off values of .08 and .05 were used to indicate a good or an acceptable model respectively (Hu & Bentler, 1999).

For REM, EXT and TRU the values indicated an acceptable to good fit between the model and the observed data. No post-hoc modifications were performed on these variables because of the good-fit indexes and the residual analysis did not indicate any problems. The scales measuring KUD and EXP exhibited a bad to mediocre fit to the data. Modification indices demonstrated that allowing for some covariances between the items would improve the models. Table 16 provides an overview of the covariates that were allowed in the models. The improved models received the name suffix _COV and the fit indices show acceptable to good fits. The scales measuring STR and AUT demonstrated a bad fit to the observed data. For STR the modification indices suggested to allow for some covariances which resulted in an acceptable fit (STR_COV). For AUT the factor loadings of three items (aut10, aut11, aut12) were very low (<.200) so we removed them from the model and based on modification indices we allowed for some covariances. The downsized model with covariances (AUT_XS_COV) shows an acceptable to good fit to the observed data. Cronbach coefficient alpha scores for all the factors were calculated and together with the means and standard deviations of the final scales they are presented in Table 15. Standardised parameter estimates and unstandardised estimates for the final models are provided in Table 16.
Table 15. Goodness-of-fit indicators for the latent variables

<table>
<thead>
<tr>
<th>Latent Variable</th>
<th>df</th>
<th>χ²</th>
<th>CFI</th>
<th>TLI</th>
<th>RMSEA</th>
<th>SRMR</th>
<th>N items</th>
<th>Scale range</th>
<th>α</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>REM</td>
<td>27</td>
<td>220.225</td>
<td>***</td>
<td>.965</td>
<td>.953</td>
<td>.079</td>
<td>9</td>
<td>0-5</td>
<td>.929</td>
<td>1.65</td>
<td>.92</td>
</tr>
<tr>
<td>KUD</td>
<td>2</td>
<td>104.698</td>
<td>***</td>
<td>.909</td>
<td>.728</td>
<td>.209</td>
<td>4</td>
<td>0-3</td>
<td>.766</td>
<td>1.70</td>
<td>.57</td>
</tr>
<tr>
<td>KUD_COV</td>
<td>1</td>
<td>3.118</td>
<td>.998</td>
<td>.989</td>
<td>.042</td>
<td>.007</td>
<td>4</td>
<td>0-3</td>
<td>.766</td>
<td>1.70</td>
<td>.57</td>
</tr>
<tr>
<td>EXP</td>
<td>5</td>
<td>36.494</td>
<td>***</td>
<td>.953</td>
<td>.905</td>
<td>.074</td>
<td>5</td>
<td>0-3</td>
<td>.644</td>
<td>1.57</td>
<td>.46</td>
</tr>
<tr>
<td>EXP_COV</td>
<td>4</td>
<td>15.233</td>
<td>**</td>
<td>.983</td>
<td>.958</td>
<td>.049</td>
<td>5</td>
<td>0-3</td>
<td>.644</td>
<td>1.57</td>
<td>.46</td>
</tr>
<tr>
<td>EXT</td>
<td>9</td>
<td>45.678</td>
<td>***</td>
<td>.961</td>
<td>.934</td>
<td>.059</td>
<td>6</td>
<td>0-6</td>
<td>.714</td>
<td>3.75</td>
<td>.88</td>
</tr>
<tr>
<td>STR</td>
<td>77</td>
<td>1295.211</td>
<td>***</td>
<td>.641</td>
<td>.576</td>
<td>.121</td>
<td>14</td>
<td>0-3</td>
<td>.770</td>
<td>2.07</td>
<td>.28</td>
</tr>
<tr>
<td>STR_COV</td>
<td>72</td>
<td>358.327</td>
<td>***</td>
<td>.916</td>
<td>.893</td>
<td>.060</td>
<td>14</td>
<td>0-3</td>
<td>.770</td>
<td>2.07</td>
<td>.28</td>
</tr>
<tr>
<td>AUT</td>
<td>54</td>
<td>1015.976</td>
<td>***</td>
<td>.480</td>
<td>.365</td>
<td>.133</td>
<td>12</td>
<td>0-3</td>
<td>.709</td>
<td>1.57</td>
<td>.29</td>
</tr>
<tr>
<td>AUT_XS</td>
<td>27</td>
<td>647.538</td>
<td>***</td>
<td>.581</td>
<td>.441</td>
<td>.151</td>
<td>9</td>
<td>0-3</td>
<td>.746</td>
<td>1.41</td>
<td>.34</td>
</tr>
<tr>
<td>AUT_XS_COV</td>
<td>22</td>
<td>93.334</td>
<td>***</td>
<td>.952</td>
<td>.921</td>
<td>.057</td>
<td>9</td>
<td>0-3</td>
<td>.746</td>
<td>1.41</td>
<td>.34</td>
</tr>
<tr>
<td>TRU</td>
<td>2</td>
<td>37.481</td>
<td>***</td>
<td>.951</td>
<td>.854</td>
<td>.123</td>
<td>4</td>
<td>0-6</td>
<td>.755</td>
<td>2.70</td>
<td>.46</td>
</tr>
</tbody>
</table>

Note. *** p ≤ .001, ** p ≤ .01, * p ≤ .05
Table 16. Standardised and unstandardised coefficients for confirmatory factor analyses

<table>
<thead>
<tr>
<th>Observed variable</th>
<th>Latent construct</th>
<th>$\beta$</th>
<th>B</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>rem1</td>
<td>REM</td>
<td>.700</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>rem2</td>
<td>REM</td>
<td>.636</td>
<td>.783</td>
<td>***</td>
</tr>
<tr>
<td>rem3</td>
<td>REM</td>
<td>.638</td>
<td>1.002</td>
<td>***</td>
</tr>
<tr>
<td>rem4</td>
<td>REM</td>
<td>.894</td>
<td>1.362</td>
<td>***</td>
</tr>
<tr>
<td>rem5</td>
<td>REM</td>
<td>.899</td>
<td>1.341</td>
<td>***</td>
</tr>
<tr>
<td>rem6</td>
<td>REM</td>
<td>.677</td>
<td>.880</td>
<td>***</td>
</tr>
<tr>
<td>rem7</td>
<td>REM</td>
<td>.739</td>
<td>1.046</td>
<td>***</td>
</tr>
<tr>
<td>rem8</td>
<td>REM</td>
<td>.838</td>
<td>1.256</td>
<td>***</td>
</tr>
<tr>
<td>rem9</td>
<td>REM</td>
<td>.883</td>
<td>1.342</td>
<td>***</td>
</tr>
<tr>
<td>ext1</td>
<td>EXT</td>
<td>.543</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>ext2</td>
<td>EXT</td>
<td>.479</td>
<td>.821</td>
<td>***</td>
</tr>
<tr>
<td>ext3</td>
<td>EXT</td>
<td>.489</td>
<td>.608</td>
<td>***</td>
</tr>
<tr>
<td>ext4</td>
<td>EXT</td>
<td>.664</td>
<td>.963</td>
<td>***</td>
</tr>
<tr>
<td>ext5</td>
<td>EXT</td>
<td>.623</td>
<td>1.218</td>
<td>***</td>
</tr>
<tr>
<td>ext6</td>
<td>EXT</td>
<td>.492</td>
<td>.759</td>
<td>***</td>
</tr>
<tr>
<td>kud1</td>
<td>KUD_COV</td>
<td>.822</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>kud2</td>
<td>KUD_COV</td>
<td>.756</td>
<td>1.061</td>
<td>***</td>
</tr>
<tr>
<td>kud3</td>
<td>KUD_COV</td>
<td>.549</td>
<td>.745</td>
<td>***</td>
</tr>
<tr>
<td>kud4</td>
<td>KUD_COV</td>
<td>.452</td>
<td>.600</td>
<td>***</td>
</tr>
<tr>
<td>kud3 ~~ kud4</td>
<td></td>
<td>.361</td>
<td>.152</td>
<td>***</td>
</tr>
<tr>
<td>exp1</td>
<td>EXP_COV</td>
<td>.354</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>exp2</td>
<td>EXP_COV</td>
<td>.547</td>
<td>1.493</td>
<td>***</td>
</tr>
<tr>
<td>exp3</td>
<td>EXP_COV</td>
<td>.530</td>
<td>1.872</td>
<td>***</td>
</tr>
<tr>
<td>exp4</td>
<td>EXP_COV</td>
<td>.720</td>
<td>2.115</td>
<td>***</td>
</tr>
<tr>
<td>exp5</td>
<td>EXP_COV</td>
<td>.513</td>
<td>1.599</td>
<td>***</td>
</tr>
<tr>
<td>exp1 ~~ exp4</td>
<td></td>
<td>-.207</td>
<td>-.061</td>
<td>**</td>
</tr>
<tr>
<td>str1</td>
<td>STR_COV</td>
<td>.408</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>str2</td>
<td>STR_COV</td>
<td>.441</td>
<td>1.087</td>
<td>***</td>
</tr>
<tr>
<td>str3</td>
<td>STR_COV</td>
<td>.496</td>
<td>1.537</td>
<td>***</td>
</tr>
<tr>
<td>str4</td>
<td>STR_COV</td>
<td>.323</td>
<td>1.063</td>
<td>***</td>
</tr>
<tr>
<td>str5</td>
<td>STR_COV</td>
<td>.423</td>
<td>.891</td>
<td>***</td>
</tr>
<tr>
<td>str6</td>
<td>STR_COV</td>
<td>.495</td>
<td>1.128</td>
<td>***</td>
</tr>
<tr>
<td>str7</td>
<td>STR_COV</td>
<td>-.208</td>
<td>-.593</td>
<td>***</td>
</tr>
<tr>
<td>str8</td>
<td>STR_COV</td>
<td>.532</td>
<td>1.428</td>
<td>***</td>
</tr>
<tr>
<td>str9</td>
<td>STR_COV</td>
<td>.436</td>
<td>1.023</td>
<td>***</td>
</tr>
<tr>
<td>str10</td>
<td>STR_COV</td>
<td>.469</td>
<td>1.120</td>
<td>***</td>
</tr>
<tr>
<td>str11</td>
<td>STR_COV</td>
<td>.597</td>
<td>1.453</td>
<td>***</td>
</tr>
<tr>
<td>str12</td>
<td>STR_COV</td>
<td>.545</td>
<td>1.341</td>
<td>***</td>
</tr>
<tr>
<td>str14</td>
<td>STR_COV</td>
<td>.584</td>
<td>1.296</td>
<td>***</td>
</tr>
<tr>
<td>str15</td>
<td>STR_COV</td>
<td>.618</td>
<td>1.563</td>
<td>***</td>
</tr>
<tr>
<td>str9 ~~ str10</td>
<td></td>
<td>.754</td>
<td>.163</td>
<td>***</td>
</tr>
<tr>
<td>str5 ~~ str6</td>
<td></td>
<td>.472</td>
<td>.086</td>
<td>***</td>
</tr>
<tr>
<td>str2 ~~ str3</td>
<td></td>
<td>.228</td>
<td>.066</td>
<td>***</td>
</tr>
<tr>
<td>str11 ~~ str12</td>
<td></td>
<td>.234</td>
<td>.046</td>
<td>***</td>
</tr>
<tr>
<td>str1 ~~ str6</td>
<td></td>
<td>.174</td>
<td>.037</td>
<td>***</td>
</tr>
<tr>
<td>aut1</td>
<td>AUT_XS_COV</td>
<td>.538</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>aut2</td>
<td>AUT_XS_COV</td>
<td>.625</td>
<td>1.147</td>
<td>***</td>
</tr>
<tr>
<td>aut3</td>
<td>AUT_XS_COV</td>
<td>.689</td>
<td>1.333</td>
<td>***</td>
</tr>
<tr>
<td>aut4</td>
<td>AUT_XS_COV</td>
<td>.195</td>
<td>.366</td>
<td>***</td>
</tr>
<tr>
<td>aut5</td>
<td>AUT_XS_COV</td>
<td>.248</td>
<td>.442</td>
<td>***</td>
</tr>
</tbody>
</table>
4.3.2 Multiple mediation model

The examination of correlations revealed that no independent variables are highly correlated (see Table 17). Also, the collinearity statistics are well within accepted limits with the variance inflator factor (VIF) ranging from 1.01 for REM to 1.39 for KUD_COV (thus <10) and tolerance levels ranging from .992 for REM and .718 for KUD_COV (thus >.2)(Hair, Anderson, Tatham, & Black, 1995). The SEM analysis was performed with the Lavaan 0.5-17 R Package (Rosseel, 2012) and again a maximum likelihood estimation with robust standard errors was used. Based on the results of bivariate analyses we hypothesised the SEM described graphically in Figure 9. Circles represent latent variables and rectangles represent measured variables. Figure 9 does not show the measurement component which is described in the previous section and in Tables 15 and 16. We note that the model is slightly overidentified because independent samples T-tests showed that the relationships between teacher gender and autonomy and trust in students are not significant, nevertheless, we identified them in the model in order to be able to calculate all the relevant indirect, direct and total paths of gender effects to allow for analysis of mediation through restricted emotionality and experimentation. No significant gender differences were found for extra-role behaviour and keeping up to date, so they are entered in the model as covariates and not as mediators (Little, Card, Bovaird, Preacher, & Crandall, 2007).
Table 17. Correlation table for the independent variables

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>REM (1)</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KUD_COV (2)</td>
<td>-0.088**</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EXP_COV (3)</td>
<td>-0.085**</td>
<td>-0.400***</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>EXT (4)</td>
<td>-0.042</td>
<td>-0.329***</td>
<td>-0.459***</td>
<td>1</td>
</tr>
</tbody>
</table>

*Note.*** p ≤ .001, ** p ≤ .01, * p ≤ .05*

Figure 9. Hypothesised model of teacher background variables predicting teacher-reported teacher-student relationship.

The fit indices provide ambiguous results. The RMSEA and SRMS indicate a good fit, respectively .04 and .05. The CFI and TLI, however, indicate a bad fit with respectively .88 and .87. The TLI and CFI values below .90 might be a penalty for the complexity of our model (χ²/df ratio is 2.32). Kenny (2014b) suggest that when a measurement model has a RMSEA smaller than .158 (which is the case for the majority of our measurement model), an incremental measure of fit such as TLI and CFI may not be that informative. Considering that our main focus of research is not on the items themselves but on testing the direct and indirect effects between the latent variables in
our mediation model and since the factors in our models are no “obscure invalid assumptions” we double checked the structural part of our model in a less complex model by replacing the latent variables by item parcels created by taking the mean of all items within a factor (Little, Cunningham, & Shahar, 2002, p. 169). This simplified model shows excellent to good fit, with an RMSEA of .05, an SRMR of .03, a CFI of .96 and a TLI of .89. Considering Kenny’s (2014b) suggestion that incremental measures of fit are not so informative in our case and considering a good to excellent fit for the simplified structural part of our model, we opt to interpret the model as having an acceptable fit and not to perform post-hoc modifications to the model. Figure 10 provides a graphical representation of the results and Table 18 shows the results for all direct, indirect and total effects in the multiple mediation model.
Figure 10. Results structural equation model.

Note. Only statistically significant relationships at the p < .05 level are shown. RMSEA = .04; SRMS = .05; CFI = .88; TLI = .87. $R^2_{REM} = .05$; $R^2_{EXP-COV} = .06$; $R^2_{STR-COV} = .27$; $R^2_{AUT_XS-COV} = .20$; $R^2_{TRU} = .09$. 
Table 18. Results from structural equation model: direct, indirect and total effects

<table>
<thead>
<tr>
<th></th>
<th>STR_COV</th>
<th></th>
<th></th>
<th>AUT_XS_COV</th>
<th></th>
<th></th>
<th>TRU</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \beta )</td>
<td>( B )</td>
<td>SE</td>
<td>( \beta )</td>
<td>( B )</td>
<td>SE</td>
<td>( \beta )</td>
<td>( B )</td>
<td>SE</td>
</tr>
<tr>
<td>Direct</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gender</td>
<td>.12</td>
<td>.06</td>
<td>***</td>
<td>.02</td>
<td>-0.00</td>
<td>-0.00</td>
<td>.03</td>
<td>-0.01</td>
<td>-0.00</td>
</tr>
<tr>
<td>REM</td>
<td>-0.28</td>
<td>-0.08</td>
<td>***</td>
<td>.01</td>
<td>-0.13</td>
<td>-0.05</td>
<td>**</td>
<td>-0.10</td>
<td>-0.04</td>
</tr>
<tr>
<td>EXP_COV</td>
<td>.05</td>
<td>.05</td>
<td>.07</td>
<td>.30</td>
<td>.40</td>
<td>***</td>
<td>.11</td>
<td>.06</td>
<td>.08</td>
</tr>
<tr>
<td>Indirect</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender (mediator REM)</td>
<td>.05</td>
<td>.03</td>
<td>***</td>
<td>.01</td>
<td>.02</td>
<td>*</td>
<td>.01</td>
<td>.02</td>
<td>*</td>
</tr>
<tr>
<td>Gender (mediator EXP_COV)</td>
<td>-0.01</td>
<td>-0.00</td>
<td>.00</td>
<td>-0.03</td>
<td>-0.02</td>
<td>*</td>
<td>.01</td>
<td>-0.00</td>
<td>-0.00</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>.17</td>
<td>.09</td>
<td>***</td>
<td>.02</td>
<td>-0.01</td>
<td>-0.01</td>
<td>.03</td>
<td>.00</td>
<td>.00</td>
</tr>
</tbody>
</table>

Note. *** \( p \leq .001 \), ** \( p \leq .01 \), * \( p \leq .05 \). \( R^2_{REM} = .05; R^2_{EXP_COV} = .06; R^2_{STR_COV} = .27; R^2_{AUT_XS_COV} = .20; R^2_{TRU} = .09. \)
4.3.2.1 Direct effects
Teacher gender is positively and significantly related to the provision of structure (women scoring higher, $\beta=.121$, $p=.001$). Autonomy support and trust in students are not significantly predicted by teacher gender. Restricted emotionality is significantly related to structure ($\beta=-.281$, $p=.000$), autonomy ($\beta=-.132$, $p=.005$), and trust ($\beta=-.100$, $p=.011$). Experimentation and reflection only significantly predicts autonomy support ($\beta=.303$, $p=.000$).

4.3.2.2 Indirect effects
We hypothesised that teacher gender effects would be mediated through restricted emotionality and experimentation and reflection. Four indirect effects through restricted emotionality and experimentation and reflection are significant: the effect of teacher gender through restricted emotionality on structure ($\beta=.052$, $p=.000$), on autonomy ($\beta=.024$, $p=.013$), and on trust in students ($\beta=.018$, $p=.020$); and the indirect effect of teacher gender through experimentation and reflection on autonomy ($\beta=-.029$, $p=.020$). Male teachers report higher levels of restricted emotionality which is negatively correlated with the provision of structure, autonomy support and trust in students. Male teachers report higher levels of experimentation and reflection, which translates into higher levels of autonomy support.

4.3.2.3 Total effects
The total effect of the mediation model predicting structure is statistically significant ($\beta=.168$, $p=.000$). A partial mediation through restricted emotionality is assumed since the direct effect of teacher gender remains significant (Little et al., 2007).

The total effect of the multiple mediation model predicting autonomy is not significant ($\beta=-.008$, $p=.847$). An examination of the model shows that the direct effect of teacher gender is not significant but the two indirect effects through restricted emotionality and experimentation are. The indirect effects are of approximately equal size but with an opposite sign (standardised coefficients of respectively .024 and -.029). The sum of the two indirect effects is not significant ($\beta=.005$, $p=.774$) and close to zero resulting in the appearance of “no effect” being mediated. Kenny (2014a) calls this an
opposing mediation. Men report higher levels of restricted emotionality which has a negative effect on autonomy support, and at the same time, men report higher levels of experimentation which has a positive effect on autonomy support. Both effects cancel each other out.

The total effect of the mediation model predicting trust in students is not significant ($\beta=0.001, p=0.970$). The significant indirect effect of teacher gender through restricted emotionality on trust without a significant direct effect of teacher gender could point to a full mediation. We have to be careful, however, with such interpretations since a suppression effect could just as well be affecting the model (Little et al., 2007; Rucker, Preacher, Tormala, & Petty, 2011). As is the case in predicting autonomy, another source of variance with an opposed sign could be mediating the teacher gender effect.

4.4 Discussion

4.4.1 Conclusion

This study was aimed at investigating the effects of teacher gender on (teacher-reported) teacher-student relationship. Based on feminist and educational theory two mediators were hypothesised to carry the indirect effect of teacher gender: restricted emotionality and extended professionalism. This last was broken up in three dimensions: keeping up to date with professional literature, experimentation and reflection, and extra-role behaviour. Teacher-student relationship was operationalised by means of teacher-reported measures of provision of structure, autonomy support and trust in students.

All three dimensions of extended professionalism indicate a broadening of teachers’ role-orientation towards the adoption of more theory and rational activity and point to less isolation of teachers and less supremacy of intuition of practice. Based on previous empirical research and theoretical work restricted emotionality and extended professionalism are hypothesised to be gendered, with women leaning more towards the stereotypically feminine emotionality and men being more susceptible to the stereotypically male rationality and theory-driven approaches that characterise extended professionalism.
Our results show that emotionality and some dimensions of teacher extended professionalism are indeed gendered along these traditional lines. Male teachers are found to be more restricted in their emotionality and they report more experimentation and reflection. No significant differences were found for teachers’ level of keeping up to date with professional literature and extra-role behaviour.

With regard to the effects of teacher gender on teacher-student relationship our analyses suggest a marginal to no total effect. In predicting autonomy support (taking into account students’ opinions, interests etc.) and in predicting trust in students, no significant effect was found. Provision of structure is slightly gendered with women reporting higher levels of structure than men. However, gender is the weakest significant predictor in the model. Extra-role behaviour and keeping up to date combined - which do not differ for men and women – have a coefficient four times larger than gender. These findings are important in itself, since the feminisation of teaching can not be problematised based on the claim that male or female teachers establish less qualitative relationships with their students.

The near absence of teacher gender effects on teacher-student relationship, however, does not mean that teaching is not gendered. A detailed analysis of the structural model clearly shows a complex interplay of gender effects. In predicting autonomy this is most apparent. Two opposing mediations are at play, cancelling each other out and giving the impression that gender has no effect whatsoever: male teachers’ higher level of restricted emotionality results in lower levels of autonomy support, but at the same time, men’s higher level of experimentation and reflection results in higher levels of autonomy support. Since both opposing effects are of similar size but with opposite signs, in sum, no significant gender effect exists. In predicting trust (willingness to be vulnerable towards students based on the confidence that they are reliable etc.) a similar mechanism is possible at play. A clear indirect effect of teacher gender on trust in students that is carried by restricted emotionality is observed (female teachers reporting less restriction and more trust), but the direct effect and total effect of teacher gender on trust in students is not significant. A total mediation or – similar to the model predicting autonomy – an unknown suppressor variable is influencing the teacher gender effect. Further research is needed to investigate if teacher gender effects on trust in students are indeed fully mediated by restricted
emotionality or if a more complex mechanism is taking place. In predicting structure (providing students with clear expectations, consistent rules and instrumental help) a partial mediation of teacher gender effect through restricted emotionality is observed. Male teachers report higher levels of restricted emotionality which, in turn, relate to lower levels of provision of structure. The direct effect of teacher gender on structure provision remains significant when entering restricted emotionality to the model, so only a partial mediation is assumed.

4.4.2 Discussion

When presenting our research design we mentioned choosing a quantitative study because conclusions might have greater credibility and impact on many people in power such as administrators, politicians and funding agencies (Johnson & Onwuegbuzie, 2004). We wish to emphasise for these stakeholders that this study found no indications that the numerical feminisation of teaching is to be considered a problem based on the claim that male or female teachers’ relationship with students would be less qualitative. This finding is not surprising considering previous observation studies investigating actual teacher-student classroom interactions that have also found no significant teacher gender effects (Howe, 1997). This tempts us to conclude that the gendered character of teaching is heavily overrated. Indeed, we believe it is the case, but stating that the feminisation of teaching is a non-issue, however, would be an oversimplification. The multiple mediations of our structural model show that teaching is clearly gendered with men adhering to different strategies than women to achieve these qualitative relationships with their students. Men are found to rely more on their reflective skills in assessing their own teaching and their openness to try out new approaches when deemed to improve their teaching. Women draw on their greater emotionality to build qualitative relationships with students. We immediately nuance these findings by emphasising that effects are indeed significant but relatively small and not present in predicting all measures of teacher-student relationships. We do not advocate for a gender-neutral and disembodied teaching (Acker, 1990), nevertheless, an accumulation of subtle differences in teaching style between male and female teachers that reinforce traditional gender role stereotypes could surely affect students’ beliefs about
masculinity and femininity and their construction of beliefs regarding gendered professions. Gendered teaching styles are part of the hidden curriculum and should not be left unchallenged. Emotionality and experimentation might be gendered but they are not conflicting with each other. Teachers with a more extended emotionality are also more likely to have a more extended professionalism. Contrary to the fear expressed by Forrester (2005) the introduction of more rationality, theory and systematic inquiry into teaching should not as such lead to fundamental conflicts in teachers’ professional identity. Also, both dimensions of teacher professionalism are found to positively correlate with teacher-student relationships. We therefore advise teacher education and teacher professional development programs to pay equal attention to both emotionality and rationality as important dimensions of teaching. We also suggest supporting the rise and development of extended professionalism in teachers as means to bridge the theory-practice and rationality-emotionality gap. We believe this study provides empirical support for Christianakis’ (2008) claim that the movement towards ‘teacher research’ and systematic and rational teacher inquiry into their own practice is a feminist act, and not the uncritical and unconscious acceptance of the supremacy of rationality over emotionality. Recently, Consuegra, Engels, and Struyven (2014), however, found that schools tend to hinder rather than serve teacher inquiry and workplace learning. Future research thus needs to investigate which school-level characteristics facilitate the development of teachers’ extended professionalism and their co-construction of knowledge on teaching, since bridging the gap between emotionality and rationalism and theory and practice by developing both to high extents is found to be an important predictor of teacher-student relationship quality.

4.4.3 Limitations

A strength but also a limitation of this study is the focus on teacher perceptions. In 2014 Pulsford comments that the voices of teachers themselves should receive more attention in theorizing about models of gendered professionalism, a recurring critique in feminist literature. Our study meets this need but we immediately point to the limitations of teacher perceptions of teacher-student relationships as an indicator of actual teaching
quality. The review of Stroet et al. (2013) on the effects of need supportive teaching shows only small or even no associations between teacher perceptions of autonomy support and provision of structure on adolescents’ motivation and engagement. Student reported measures of autonomy and structure (which do to some extent correlate with teacher perceptions) are much more accurate predictors. Further research would benefit from a triangulation of perspectives (including student perceptions) in measuring the effects of teacher gender (and its mediators) on teacher-student relationships.

A second limitation of this study is that it does not engage in actual interaction with teachers to co-construct knowledge on teacher professionalism. This is, however, a recurrent recommendation in teacher professionalization research (Cochran-Smith & Lytle, 1999). We note, though, that fundamental research is not to be omitted and replaced entirely by practitioner and teacher research. Also, this study was only one within a greater research project and other studies within the project did involve teacher research. Teachers acted as researchers analysing their own video-recorded lessons in search of gender differences in pupil’s classroom interactions (Consuegra et al., in press). We thus tried to meet both the need for testing a theoretical model using a large-scale quantitative design, and the need for teacher research to co-constructing knowledge on gendered teaching and teacher professionalism.

Acknowledgement
We would like to gratefully acknowledge the support of our colleague Gert Vanthournout for his significant contribution in the preliminary analysis and CFAs.
Chapter 5
Effectiveness of Teacher Professionalisation
Chapter adapted from:
Abstract

There have been numerous studies investigating the extent to which teacher-student classroom interactions differ between boys and girls and the results of these studies suggest that teacher negative feedback is higher for boys which in turn leads to lower levels of on-task behaviour. The article describes the results of a quasi-experimental study, involving 30 teachers and a selection of 129 of their students, aimed at improving teacher feedback patterns, student behavioural responses and student perception of equity in the classroom environment. Fifteen teachers took part in a five-session professional development programme in which they investigated their own teaching practice by means of collaborative appreciative inquiry. The pre-test and post-test data in the control group illustrate that girls who receive very high levels of positive feedback at the beginning of the year succeed in maintaining and even strengthening their position as favoured students while at the same time lowering their level of raising hands and increasing their level of calling out to the teacher. By contrast, the data in the intervention group show a drop of very high levels of positive feedback and we see the levels of misbehaviour in boys and girls rise. Students’ sense of equity increases for all conditions to a same extent. The paper considers two alternative explanations for the results: some students actively resist changes in teacher treatment and various confounding factors might have been omitted. Implications and suggestions for further research and the design of professional development programs on gendered classroom interactions are discussed.
5.1 Introduction

Previous research has convincingly shown that teacher-student classroom interactions are gendered (Beaman et al., 2006; Howe, 1997; Jones & Dindia, 2004). One recurrent finding is that - in general - boys receive more negative feedback from their teachers than girls. The meta-analysis of Kelly (1988) arrived at this conclusion in the 1980s and the review of Beaman et al. in 2006 still confirms it. Patterns of gendered negative feedback might have remained stable throughout the years, interpretations of the results have not. In the 1980s a teacher gender bias was theorised explaining that teachers have differential expectations toward boys and girls and that these are self-fulfilling by means of gendered treatment (Brophy, 1985). In the 1990s the existence of teacher gender bias was called into question. It was suggested that boys bring criticism upon themselves as a consequence of their higher levels of disruptive behaviour (Howe, 1997).

Recent evidence shows that both theories are probably – at least in part – correct. Consuegra, Halimi and Engels (under review) found that girls indeed get away with some of their classroom misbehaviour (e.g. talking to other students), while other misbehaviours (e.g. calling out) are related to teacher criticism in boys and girls alike. Teacher gender bias is thus subtle, nonetheless, it is affecting students. Boys argue that teachers are more lenient toward girls and they report higher levels of discrimination in class (Younger et al., 1999; Consuegra, Halimi, & Engels, under review). Teachers, on the other hand, are found to be relatively unaware of their gender differential treatment, sticking to the belief that they treat boys and girls equal (Consuegra et al., in press; Younger et al., 1999; Garrahy, 2001).

With regard to positive feedback empirical findings are inconsistent. Some review studies conclude that boys and girls receive equal amounts (e.g. Jones & Dindia, 2004) while others report that boys occasionally receive more positive feedback than girls (e.g. Howe and Abedin, 2013). Howe in 1997 concludes that research on positive interactions is inconclusive and she suggests studying the proportion of positive to negative interactions rather than the separate frequencies, since it may be more significant in terms of impact on pupils. Correlational analyses of Myhill (2002) indeed suggest that the higher the ratio of positive to negative interactions the more pupil on-task behaviour is displayed. White (2010) recommends a range from 3:1 to 10:1
for optimal promotion of appropriate student behaviours. Such ratios are, however, seldom reached. Levels of praise and reinforcement in general are relatively low (Drudy and Uí Chatháin, 2002; Merrett & Wheldall, 1992; Jones & Dindia, 2004). Swinson and Harrop in 2005 trained teachers to increase their amount of positive feedback to pupils. They found that the levels of on-task behaviour in students increased from 77 to 94 percent when positive feedback was three times more frequent than negative feedback. Only few studies have, however, confirmed this – or any - causal relationship between teacher treatment and student behavioural responses by means of experimental research (White, 2010; Howe, 1997).

This study is aimed at doing exactly that and investigates the effect of a professional development programme on gendered teacher-student classroom interactions and students’ perception of equity in the classroom environment.

5.2 Theoretical framework

5.2.1 Effectiveness of teacher professionalisation

In 2009 Desimone argues that a common conceptual framework is needed in professional development impact studies to enable researchers to synthesise, refine and expand the knowledge base on teacher professionalisation. In their review van Veen, Zwart, Meirink, and Verloop (2010) discuss several suggested models for assessing the effectiveness of teacher professionalisation. They note that recurring in all models is the central importance of the hypothesised causal relationships between (1) characteristics of the professional development intervention, (2) teacher knowledge and attitudes, (3) teacher behaviour and teaching practice, and (4) student outcomes (see Figure 11). Wayne, Yoon, Zhu, Cronen and Garet (2008) distinguish between two theories: the ‘theory of change’ which addresses the relationship between the first two elements, and the ‘theory of instruction’ that discusses the relationship between the two latter. Beside these theories, van Veen et al. (2010) also identify the effects of school conditions that can facilitate or obstruct the impact of professional development. In the following sections we present the theory of instruction and the theory of change that are hypothesised in this study.
5.2.2 Theory of instruction: evidence-informed appreciative approach

Optimal instruction, Brophy argues in 1985, implies some degree of individualisation of instruction. It is not always desirable to treat all students equal (Tomlinson et al., 2003). In order to avoid, however, that differentiated instruction is negatively influenced by incorrect and stereotyped expectations of teachers, we suggest an evidence-informed and an appreciative approach.

5.2.2.1 Evidence-informed approach
Noticing pupil diversity and differentiating instruction according to different student needs is key to establishing equal learning opportunities. The knowledge to notice and to interpret significant features of classroom situations is called ‘professional vision’ (Seidel & Stürmer, 2014). A variety of studies have shown that developing teacher professional vision impacts teachers’ classroom practice (Sherin & Han, 2004; Sherin & van Es, 2005; van Es & Sherin, 2010). In developing teacher professional vision, the use of video has become a prominent tool (van Es & Sherin, 2008; Seidel & Stürmer, 2014). Video allows displaying the complexity of real classroom interactions by capturing voices and body language. Video-stimulated recall interviewing (Tripp & Rich, 2012), in which teachers watch videos of their own teaching and discuss these with others, enables teachers to gain a perspective on their practice that they can otherwise not take. It provides them an opportunity to become aware of incorrect perceptions of students and biased interactions.
5.2.2.2 **Appreciative approach**
Swinson and Harrop (2012) suggest a positive and appreciative pedagogy to increase student on-task behaviour. Positive psychology is a term used to address the study of positive interactions and emotions, positive personality traits and positive institutions (Seligman et al., 2005). It is a relatively young research field that has flourished the past decades (Seligman & Csikzentmihalyi, 2000). Of course, positive approaches to learning and development have existed long before being labelled ‘positive’. Appreciative inquiry (Cooperrider & Srivasta, 1987), flow theory (Csikzentmihalyi, 1990), self-determination theory (Ryan and Deci, 2000) and the U-theory (Senge et al., 2004a) all stress positive motivations and appreciation of strengths as an important factor that drives learning. These appreciative approaches can be distinguished from deficiency-based approaches that depart from problems.

5.2.3 **Theory of change: teacher collaboration and teacher inquiry**

Based on the work of Timperley, Wilson, Barrar and Fung (2007), van Veen et al. (2010), Wayne et al., (2008), Hammerness et al. (2007), and Desimone (2009) we identify the following essential characteristics of effective teacher professional development:

1. Focus on student learning
2. Taking into account the subject field and prior knowledge of teachers
3. Extended and intensive programs
4. School-based programmes that are incorporated into the daily work of teachers
5. Collaboration with peers
6. Integrating theory and practice
7. Production of knowledge by means of teacher inquiry
8. Teacher ownership of content and process, responding to needs and interests

Within the scope of this article two elements in particular are relevant: teacher collaboration and teacher inquiry.

5.2.3.1 **Teacher collaboration**
In an evidence-informed approach teacher collaboration is key since it allows for information sharing and intersubjectivity between interpretations. Teacher
collaboration is, however, not self-evident. In Flanders (Dutch speaking part of Belgium where this study is performed) teachers are still very much isolated and working on an individual basis (Deneire, Vanhoof, Faddar, and Van Petegem, 2013). Flemish teacher education programs do not prepare teachers for collaboration with colleagues (Valcke, Rots, & Struyven, 2012) and as a consequence teachers’ self-efficacy with regard to collaboration is very low (Schraepen, Lebeer, & Vanpeperstraete, 2010). In order for successful and sustainable teacher collaboration, Vescio, Ross and Adams (2008) conclude their review, collaborative learning should always be student-centred and directed towards improving learning outcomes and well-being of pupils. Also, teachers should be provided with the time to plan and structure their own collaboration (Santoli, Sachs, & Romey, 2008).

5.2.3.2 Appreciative inquiry
It is often stated that the driving force for reflection is the recognition of uncertainty involving a problematic situation (Hoban & Hastings, 2006; Loughran, 2002). Consequently, many models for teacher inquiry start from problem statements. Appreciative inquiry (Cooperrider & Srivasta, 1987), however, suggests a strength-based alternative, starting from the best of what is and the desired future state that could be. This ‘positive principle’ is assumed to increase positive emotions, which in turn leads to people being more flexible, integrative and open to information (Bushe, 2011). This appreciative approach to inquiry is promising, since teachers can be very resistant to critical and systematic inquiry into their own practice. Many teachers are unconcerned about theory and their teaching is often driven by intuition (Hoyle, 1980). Also, teachers are not familiar with seeing themselves as the producers of knowledge. Most teachers believe that the knowledge base for teaching is constructed by ‘experts’ and they merely have to apply the knowledge produced by others (Cochran-Smith & Lytle, 1992). The appreciative approach to inquiry is hypothesised to increase teachers’ motivation to step outside of their comfort-zone and increase their self-efficacy and optimism to engage in collaborative teacher inquiry (Verleysen et al., 2015).
5.3 Research focus

This study investigates the effect of a professional development programme aimed at improving teachers’ feedback patterns toward boys and girls. The theory of change stresses the importance of collaboration and appreciative inquiry as a way to support teacher professional development. The theory of improvement emphasises the provision of positive feedback to students and differentiated instruction based on actual differences in student needs rather than differentiated instruction based on stereotyped expectations. The effects of the intervention will be assessed on several levels: teacher observed instruction, student behavioural responses and student perceptions of equity.

5.4 Method

A quasi-experimental study design (Harris et al., 2006) was used to evaluate the effects of the intervention. A preintervention and a postintervention measurement design with a nonrandomly selected control and intervention group was used. Data collection was carried out at two occurrences at the beginning of the school year in October-November 2012 and at the end of the year in April-June 2013. At each time point a questionnaire was administered to and video-recorded classroom observations were performed.

5.4.1 Sample

A purposeful heterogeneous sample of six secondary schools was selected for participation in the quasi-experimental study. Schools were selected from a representative sample of 59 secondary schools in Flanders (Dutch speaking part of Belgium) that were enrolled in a larger research programme in which this study is situated. The six schools varied in a range of conditions: public and private schools; urban and suburban areas; small, medium and large-sized schools; general, vocational and technical education schools. With support from the principal in each school five teachers were selected and asked to participate in the intervention study. Teachers had to be teaching the first year of secondary education in general subjects and those teachers with most contact hours with the first-year students were addressed first for
participation in the study. For each teacher, one of their first-year secondary classes was selected and one lesson was video-recorded. Classes were selected in such a way that those classes with most contact hours with the teachers involved in the study were selected. Teachers were told that any lesson would do for the video-recording unless the entire hour would be devoted to an examination or test. In the one or two cases where the observed lessons involved something other than whole-class instruction, we would revisit the teacher at another time for filming an extra lesson (which resulted in observations of whole-class teaching for all teachers). In total, 30 teachers and their 500 pupils were video-recorded twice (pre-test and post-test). Half of the 500 pupils were randomly selected for focal coding. Due to practical reasons (out of sight, quality of recordings) the teacher-student interactions of 180 of these 250 pupils were coded. In the post-test recordings only 129 of these 180 pupils could be coded (dropout rate 28%). The mean age of the observed students at the pre-test and post-test was twelve years old. The six schools were divided randomly over the control and intervention condition in such a way that a similar diversity of school size and school type (urban/suburban, public/private, tracks in higher grades) was present in both groups. In this regard it is important to note that the control group was actually a wait group, since these teachers would participate in the intervention the following year. Each group counted 15 teachers and Table 19 provides an overview of the number of boys and girls in the control and intervention groups for the pre- and post-test.

Table 19. Control and intervention group samples, by gender

<table>
<thead>
<tr>
<th></th>
<th>Pre-test</th>
<th>Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N Total</td>
<td>N Male</td>
</tr>
<tr>
<td>Control</td>
<td>103</td>
<td>44</td>
</tr>
<tr>
<td>Intervention</td>
<td>77</td>
<td>42</td>
</tr>
<tr>
<td>Total</td>
<td>180</td>
<td>86</td>
</tr>
</tbody>
</table>
5.4.2 Instruments

5.4.2.1 Survey
Questionnaires were administered by the research team during class. Pupils were asked about their socio-demographic background and rated the extent to which students are treated equal by their teachers. To measure this sense of equity the 6-item equity subscale of the What Is Happening In This Class? (WIHIC) survey was used (Dorman, 2003). The WIHIC is a recently-developed and widely validated instrument that assesses classroom psychosocial environment in high schools. The scale is developed in such a way that it does not pre-define the basis for the discrimination that pupils experience. Sample item: “I am treated the same as other students in class”. All items are scored 0, 1, 2, 3, and 4, respectively, for the responses of almost never, seldom, sometimes, often and almost always. A principal component analysis confirmed the one-component underlying structure of the scale and reliability at the pre- and post-test was good (respectively $\alpha = .894$ and $\alpha = .935$).

5.4.2.2 Classroom observation schedule
Two educational researchers acted as the observers, using continuous focal coding for 15-minute video-recorded fragments. The middle 15 minutes of all video-recorded lessons were selected for coding teacher-student interactions to ensure that the episode would represent whole-class instruction. The first and last couple of minutes of a 50-minute lesson are traditionally used to enter and leave the classroom, fill in the school diary, hand out corrected assignments, and take quizzes. A content analysis of pupils' classroom interactions was performed using nine codes and a code for missing/out of sight (see Table 11). The codebook was based on the recurring codes in three recent studies (cf. Harrop & Swinson, 2011; Myhill, 2002; Younger & Warrington, 2002). Teacher-initiated behaviours as well as student-initiated behaviours were coded. Detailed selection criteria were consistently applied to determine if a behaviour could be assigned to a certain category. Several iterative cycles of coding, assessment of reliability and codebook modification were conducted by the two coders until reliability for all codes was high (Hruschka et al., 2004). To ensure that high levels of observer agreement would not be mere reflections of high chance agreement, a Krippendorf's alpha (Krippendorff, 2013) was calculated. As Krippendorff's
alpha is known to be a rather conservative index (Lombard, Snyder-Duch & Bracken, 2002), a more liberal criterion of .60 was used to determine if the coding of a variable was considered reliable. Training continued until kalpha's were above .60 and from that point on the two coders scored the remaining tapes, being aware that a selection of lessons, taken at random, would be scored by the other observer at the end. Table 11 gives an overview of the mean kalpha's for each code for both reliability checks combined. In line with recommendations of Harrop and Swinson (2007) we present a contingency table so that agreements and disagreements on occurrences and non-occurrences of behaviour recorded by the two observers can be inspected. The C, H, T, R, N and P categories were analysed as events, generating the frequency of each behaviour for each student. The O, F, and S categories were analysed as a state, generating the total duration of each behaviour as a proportion of time in sight for each student. The positive to negative feedback ratio (P/N) was calculated by dividing the frequency of positive feedback by the frequency of negative feedback. All values were first raised with 1 in order to avoid values of zero which could not be used as a divisor. Coding was performed using JWatcher (Blumstein et al., 2012). More information on the observation schedule and procedure can be found in Appendix 2.

5.4.3 Intervention

5.4.3.1 Intervention design
The intervention was performed in the period November 2012 to April 2013 and was made up of a video-stimulated recall interview with the main researcher and five two-hour professional development sessions. The five teachers that were video-recorded in each school would make up a professional learning community and an external facilitator supported the community. The individual video-stimulated recall interview and the five monthly sessions took place in the school itself during, before, in between or after school hours depending on the schedules of the teachers. Several cycles of inquiry of formulating goals, collecting data, analysing data and changing practice were planned. An appreciative approach was used, for example with teachers being asked to reflect and comment on good practices and strengths in their own or colleagues’ practices or in students.
5.4.3.2 Process evaluation

The following topics of discussion were predominant: (1) becoming aware of gender bias mechanisms and sharing experiences and information on students and teaching practices, (2) learning to collaborate, to appreciate and to inquire. In the first sessions teachers were mainly involved with this latter. Many meta-level discussions occurred about how to make use of the sessions and debates arose about the sense and nonsense of collaborating with colleagues (as opposed to autonomy and each teacher having an individual and uncomparable style). Teachers would also disagree on the need to read professional literature and inquire into their own practice (as opposed to experts providing clear directions on how to interact with boys and girls). Also, all teachers made explicit their appreciation for the appreciative approach. They commented that they seldom discuss their practice from a positive point of view. As acceptance with the professionalisation approach and trust among the teachers increased, the depth of discussions concerning gender bias and classroom interactions would rise. Teachers discussed their experiences of the video-stimulated recall interviews, they debated about empirical evidence and whether their own practice is biased as well, they compared good practices (e.g. successful experiences with positive and collective rewards to motivate students to do their homework) and shared positive perceptions and information on students (e.g. certain students failing subjects but excelling in others, students who ‘against all odds’ blossom in extra-curricular activities). No phase of actually trying out new and different practices and systematically evaluating the effects was reached.

5.4.4 Analysis: generalised linear mixed models (GLMMs)

A graphical assessment of the observation data (histograms and QQ-plots) clearly shows that the data are Poisson distributed (except for being on-task and sense of equity). Transforming count data to be used in parametric tests is what Cincinnato (personal communication May 19, 2015) identifies as a ‘statistical sin’. Advanced statistical techniques remove the need for transforming count data in order to be able to analyse them (O’Hara & Kotze, 2010). Generalised Linear Mixed Models (GLMM) are flexible and increasingly popular tools for analysing data with various distributions and permit analysis with hierarchical structure, spatial or temporal dependence by
means of the inclusion of random effects in the model (Thiele & Markussen, 2012). GLMMs remarkable flexibility, however, does not come without a price: “they are surprisingly challenging even for statisticians” (Bolker et al., 2009). Also, GLMM computational algorithms regularly fail to converge (Thiele & Markussen, 2012) since a large number of fixed and especially random effects compared to sample size can burden the analysis. No simulation studies exist to our knowledge that suggest criteria for sample sizes and model complexity. Still, we opt for GLMMs following O’Hara and Kotze (2010) plea to not log-transform count data to be used in parametric tests. In their simulation study they show that GLMMs consistently performed well while log-transformed data performed poorly.

5.5 Results

5.5.1 Descriptive results and bivariate statistics

Table 20 and Figure 12 present the descriptive results for all dependent variables at the pre- and post-test for control boys, control girls, intervention boys and intervention girls. We note that the means should be interpreted with caution since they are pulled in the direction of extreme observations (except for on-task and sense of equity, which are normally distributed).
Table 20. Descriptive results pre-test and post-test for teacher-student interactions and student sense of equity

<table>
<thead>
<tr>
<th></th>
<th>Pre-test</th>
<th>Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td><strong>C</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boys control</td>
<td>1.97</td>
<td>6.16</td>
</tr>
<tr>
<td>Girls control</td>
<td>1.14</td>
<td>2.72</td>
</tr>
<tr>
<td>Boys intervention</td>
<td>1.90</td>
<td>2.86</td>
</tr>
<tr>
<td>Girls intervention</td>
<td>1.91</td>
<td>3.33</td>
</tr>
<tr>
<td><strong>H</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boys control</td>
<td>5.70</td>
<td>6.15</td>
</tr>
<tr>
<td>Girls control</td>
<td>6.14</td>
<td>6.51</td>
</tr>
<tr>
<td>Boys intervention</td>
<td>2.83</td>
<td>3.34</td>
</tr>
<tr>
<td>Girls intervention</td>
<td>3.22</td>
<td>3.85</td>
</tr>
<tr>
<td><strong>T</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boys control</td>
<td>1.82</td>
<td>2.23</td>
</tr>
<tr>
<td>Girls control</td>
<td>1.64</td>
<td>2.26</td>
</tr>
<tr>
<td>Boys intervention</td>
<td>2.00</td>
<td>2.85</td>
</tr>
<tr>
<td>Girls intervention</td>
<td>2.00</td>
<td>1.86</td>
</tr>
<tr>
<td><strong>R</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boys control</td>
<td>1.85</td>
<td>2.21</td>
</tr>
<tr>
<td>Girls control</td>
<td>1.67</td>
<td>2.28</td>
</tr>
<tr>
<td>Boys intervention</td>
<td>1.83</td>
<td>2.76</td>
</tr>
<tr>
<td>Girls intervention</td>
<td>2.13</td>
<td>2.01</td>
</tr>
<tr>
<td><strong>N</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boys control</td>
<td>.82</td>
<td>2.17</td>
</tr>
<tr>
<td>Girls control</td>
<td>.45</td>
<td>.83</td>
</tr>
<tr>
<td>Boys intervention</td>
<td>1.33</td>
<td>1.83</td>
</tr>
<tr>
<td>Girls intervention</td>
<td>.78</td>
<td>.95</td>
</tr>
<tr>
<td><strong>P</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boys control</td>
<td>.88</td>
<td>1.58</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>H</td>
</tr>
<tr>
<td>--------------------------</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>Girls control</td>
<td>.71</td>
<td>1.00</td>
</tr>
<tr>
<td>Boys intervention</td>
<td>.57</td>
<td>.68</td>
</tr>
<tr>
<td>Girls intervention</td>
<td>1.30</td>
<td>1.64</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>P/N</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Boys control</td>
<td>1.35</td>
<td>.95</td>
<td>1</td>
<td>.33</td>
<td>5</td>
<td>1.72</td>
<td>1.24</td>
</tr>
<tr>
<td>Girls control</td>
<td>1.37</td>
<td>.94</td>
<td>1</td>
<td>.33</td>
<td>5</td>
<td>1.74</td>
<td>1.06</td>
</tr>
<tr>
<td>Boys intervention</td>
<td>1.00</td>
<td>.80</td>
<td>.83</td>
<td>.13</td>
<td>4</td>
<td>1.33</td>
<td>.83</td>
</tr>
<tr>
<td>Girls intervention</td>
<td>1.55</td>
<td>1.41</td>
<td>1</td>
<td>.25</td>
<td>7</td>
<td>2.00</td>
<td>1.59</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>S</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Boys control</td>
<td>4.27</td>
<td>8.52</td>
<td>1</td>
<td>0</td>
<td>39</td>
<td>4.97</td>
<td>7.13</td>
</tr>
<tr>
<td>Girls control</td>
<td>2.43</td>
<td>2.69</td>
<td>1</td>
<td>0</td>
<td>10</td>
<td>3.62</td>
<td>4.18</td>
</tr>
<tr>
<td>Boys intervention</td>
<td>4.16</td>
<td>5.23</td>
<td>2</td>
<td>0</td>
<td>18</td>
<td>7.85</td>
<td>10.27</td>
</tr>
<tr>
<td>Girls intervention</td>
<td>5.09</td>
<td>7.31</td>
<td>2</td>
<td>0</td>
<td>32</td>
<td>5.79</td>
<td>5.90</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>O</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Boys control</td>
<td>85.18</td>
<td>14.40</td>
<td>91</td>
<td>52</td>
<td>99</td>
<td>81.19</td>
<td>14.73</td>
</tr>
<tr>
<td>Girls control</td>
<td>88.83</td>
<td>8.11</td>
<td>92</td>
<td>70</td>
<td>100</td>
<td>87</td>
<td>10.03</td>
</tr>
<tr>
<td>Boys intervention</td>
<td>81.10</td>
<td>14.05</td>
<td>83</td>
<td>46</td>
<td>98</td>
<td>74.74</td>
<td>17.79</td>
</tr>
<tr>
<td>Girls intervention</td>
<td>76.91</td>
<td>15.56</td>
<td>80</td>
<td>31</td>
<td>98</td>
<td>76.08</td>
<td>16.78</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>F</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Boys control</td>
<td>5.48</td>
<td>8.49</td>
<td>2</td>
<td>0</td>
<td>34</td>
<td>6.97</td>
<td>7.52</td>
</tr>
<tr>
<td>Girls control</td>
<td>3.19</td>
<td>4.35</td>
<td>1.50</td>
<td>0</td>
<td>18</td>
<td>4.62</td>
<td>5.63</td>
</tr>
<tr>
<td>Boys intervention</td>
<td>7.97</td>
<td>8.43</td>
<td>6.50</td>
<td>0</td>
<td>36</td>
<td>11.00</td>
<td>9.89</td>
</tr>
<tr>
<td>Girls intervention</td>
<td>3.96</td>
<td>12.19</td>
<td>1</td>
<td>0</td>
<td>59</td>
<td>12.79</td>
<td>15.82</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>EQ</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Boys control</td>
<td>2.93</td>
<td>1.02</td>
<td>3</td>
<td>.40</td>
<td>4.00</td>
<td>2.68</td>
<td>1.10</td>
</tr>
<tr>
<td>Girls control</td>
<td>3.11</td>
<td>.76</td>
<td>3.17</td>
<td>1.33</td>
<td>4.00</td>
<td>2.87</td>
<td>.85</td>
</tr>
<tr>
<td>Boys intervention</td>
<td>3.31</td>
<td>.67</td>
<td>3.42</td>
<td>1.17</td>
<td>4.00</td>
<td>3.20</td>
<td>.80</td>
</tr>
<tr>
<td>Girls intervention</td>
<td>3.33</td>
<td>.86</td>
<td>3.83</td>
<td>1.17</td>
<td>4.00</td>
<td>2.97</td>
<td>.83</td>
</tr>
</tbody>
</table>

*Note.* C, H, T, R, N, P and P/N are occurrences during an observed episode. S, O and F are a percentage*100 of time during an observed episode. EQ is measured on a 5-point Likert-scale ranging from 0 (indicating low sense of equity) to 4 (indicating high sense of equity).
5.5.1.1 Pre-test differences
We performed T-tests and Mann-Whitney U tests to assess the statistical significance of differences between boys and girls and intervention and control groups. At the pre-test students in the control group raised their hand significantly more often than students in the intervention group (U=1449.5, p=.006, MRank_{ctr}=72.7, MRank_{int}=54.3), they were more on-task (t(95.33)=3.45, p=.001, M_{ctr}=87.2, SD_{ctr}=11.4, M_{int}=78.9, SD_{int}=14.8) and they received less negative feedback (U=2556, p=.004, MRank_{ctr}=57.9, MRank_{int}=74.8). As to gender differences, the only significant effect was boys being more off-task than girls (U=1517, p=.007, MRank_{boys}=73.9, MRank_{girls}=56.5).

We performed One-Way ANOVAs and Kruskal-Wallis H tests to assess the statistical significance of differences between the four crossed groups. A significant difference exists for on-task behaviour (F(3,125)=5.49, p=.001). A Turkey post-hoc test reveals that girls in the control group are significantly more on-task (M=88.8, SD=8.1) than girls in the intervention group (M=76.3, SD=15.6, p=.001). There was also a significant difference between groups for negative feedback ($\chi^2(3)=9.038$, p=.029) and being off-task ($\chi^2(3)=9.537$, p=.023). Post hoc Mann-Whitney U tests show that negative feedback is significantly higher in intervention boys than in control girls (U=419.500, p=.007, MRank_{ctr,girls}=31.5, MRank_{int,boys}=43.5) and in control boys (U=352.500, p=.030, MRank_{ctr,boys}=27.7, MRank_{int,boys}=36.8). Off-task behaviour is significantly higher for intervention boys than for intervention girls (U=221.000, p=.013, MRank_{int,boys}=21.7, MRank_{int,boys}=32.1) and control girls (U=411.000, p=.010, MRank_{ctr,boys}=31.3, MRank_{int,boys}=43.8).

No gender differences in teacher feedback are observed in the pre-test. Differences in the pre-test between control and intervention groups might point to confounding variables.

5.5.1.2 Post-test differences
At the post-test the intervention students are significantly more off-task (U=2639, p=.003, MRank_{ctr}=56.8, MRank_{int}=76.4), less on-task (t(127)=3.92, p=.000, M_{ctr}=84.6, SD_{ctr}=12.3, M_{int}=74.5, SD_{int}=17.1) and they are more often interacting with other students without permission (U=2592, p=.006, MRank_{ctr}=57.4, MRank_{int}=75.5) than students in the control group. Also, boys
in general are less on-task than girls (t(127)=-2.04, p=.043, M_{boys}=77.6, SD_{boys}=16.4, M_{girls}=83.0, SD_{girls}=13.8).

One-Way ANOVAs for the four crossed groups shows significant differences at the post-test for on-task behaviour (F(3,125)=6.21, p=.001). A Turkey post-hoc test reveals that girls in the control group are significantly more on-task (M=87.0, SD=10.0) than girls in the intervention group (M=76.1, SD=16.8, p=.020) and than boys in the intervention group (M=73.2, SD=17.6, p=.001). Also, significant differences between the four groups are found for off-task behaviour ($\chi^2(3)=11.240, p=.010$). Post hoc Mann-Whitney U tests show that off-task behaviour is no longer higher for intervention boys than for intervention girls (p=.463), but is higher than control girls (U=336.500, p=.001, MRank_{ctr_girls}=29.5, MRank_{int_boys}=46.1) and control boys (U=341.000, p=.033, MRank_{ctr_boys}=27.3, MRank_{int_boys}=37.1). We will return to these differences in the discussion, after presenting the results of the generalised linear mixed model analyses.

5.5.2 Best models

Table 21 provides the results of the ‘best models’ predicting the different dependent variables. Only fixed effects are entered in the models. Ideally, we would specify gender and intervention as fixed effects and we would specify two types of within-subject variability and random effects: multilevel grouping (student*classroom*school), and temporal dependence (pre-test*post-test). These models did not converge, however, even after trying all techniques suggested by Thiele and Markussen (2012) to solve convergence problems. We hypothesise that these models are too complex to converge given our small sample size, and we perform analyses with simplified models without random effects.

In order to find the ‘best model’ Thiele and Markussen (2012) recommend that model building for inference is done by backward elimination. We start by fitting a model with all three main variables (pre-test, gender, intervention) and all of their two-way and three-way interactions. Then the higher order variables are dropped, so long as they are not significant at the .05 critical level. We continue by successively re-fitting reduced models until no more interaction variables can be dropped. None of the three main variables will be deleted, even if they are not significant. The IBM SPSS
Statistics Version 22 software package was used and the default restricted maximum likelihood (REML) estimation for GLMM converged easily (at least, for these simplified models). Since there was evidence of over-dispersion in our count data (too many zeros), we opted to use a negative binomial distribution (Berk & MacDonald, 2007). For on-task and sense of equity which were approximately normally distributed in each of the four groups (skewness ranging from -.3 to -1.4 and kurtosis ranging from .06 to 1.9) a normal distribution was specified.
Table 21. Results generalized linear mixed modelling analyses (best models)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>C post</th>
<th>H post</th>
<th>T post</th>
<th>R post</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est.</td>
<td>SE</td>
<td>Sig.</td>
<td>Est.</td>
</tr>
<tr>
<td>Intercept</td>
<td>.579</td>
<td>.353</td>
<td>.104</td>
<td>.157</td>
</tr>
<tr>
<td>Pre</td>
<td>.210</td>
<td>.069</td>
<td>.003</td>
<td>.127</td>
</tr>
<tr>
<td>Gender</td>
<td>-.496</td>
<td>.367</td>
<td>.179</td>
<td>.514</td>
</tr>
<tr>
<td>Int</td>
<td>.052</td>
<td>.360</td>
<td>.886</td>
<td>.066</td>
</tr>
<tr>
<td>Distribution</td>
<td>Negative binomial</td>
<td>Negative binomial</td>
<td>Negative binomial</td>
<td>Negative binomial</td>
</tr>
<tr>
<td>Link function</td>
<td>Log</td>
<td>Log</td>
<td>Log</td>
<td>Log</td>
</tr>
<tr>
<td>AIC Corrected</td>
<td>482.28</td>
<td>584.68</td>
<td>498.78</td>
<td>502.04</td>
</tr>
<tr>
<td>BIC</td>
<td>496.09</td>
<td>598.49</td>
<td>512.59</td>
<td>502.04</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>N post</th>
<th>P post</th>
<th>P/N post</th>
<th>S post</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sig.</td>
<td>Est.</td>
<td>SE</td>
<td>Est.</td>
</tr>
<tr>
<td>Intercept</td>
<td>-1.384</td>
<td>.396</td>
<td>.001</td>
<td>.686</td>
</tr>
<tr>
<td>Pre</td>
<td>.861</td>
<td>.198</td>
<td>.000</td>
<td>-.379</td>
</tr>
<tr>
<td>Gender</td>
<td>.399</td>
<td>.403</td>
<td>.324</td>
<td>-.763</td>
</tr>
<tr>
<td>Int</td>
<td>.035</td>
<td>.374</td>
<td>.925</td>
<td>-.803</td>
</tr>
<tr>
<td>Pre*Gender</td>
<td>-.402</td>
<td>.190</td>
<td>.036</td>
<td>.699</td>
</tr>
<tr>
<td>Pre*Int</td>
<td>-.253</td>
<td>.110</td>
<td>.023</td>
<td>.892</td>
</tr>
<tr>
<td>Pre<em>Gender</em>Int</td>
<td>-.985</td>
<td>.441</td>
<td>.027</td>
<td>.779</td>
</tr>
<tr>
<td>Distribution</td>
<td>Negative binomial</td>
<td>Negative binomial</td>
<td>Negative binomial</td>
<td>Negative binomial</td>
</tr>
<tr>
<td>Link function</td>
<td>Log</td>
<td>Log</td>
<td>Log</td>
<td>Log</td>
</tr>
<tr>
<td>AIC Corrected</td>
<td>264.46</td>
<td>407.15</td>
<td>949.74</td>
<td>688.99</td>
</tr>
<tr>
<td>BIC</td>
<td>283.55</td>
<td>431.37</td>
<td>512.59</td>
<td>502.04</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>O post</th>
<th>F post</th>
<th>EQ post</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est.</td>
<td>SE</td>
<td>Sig.</td>
</tr>
<tr>
<td>Intercept</td>
<td>2.484</td>
<td>.665</td>
<td>.000</td>
</tr>
<tr>
<td>Pre</td>
<td>.659</td>
<td>.080</td>
<td>.000</td>
</tr>
<tr>
<td>Gender</td>
<td>-.430</td>
<td>.207</td>
<td>.040</td>
</tr>
<tr>
<td>Int</td>
<td>.423</td>
<td>.220</td>
<td>.057</td>
</tr>
<tr>
<td>Distribution</td>
<td>Normal</td>
<td>Negative binomial</td>
<td>Normal</td>
</tr>
<tr>
<td>Link function</td>
<td>Identity</td>
<td>Log</td>
<td>Identity</td>
</tr>
<tr>
<td>AIC Corrected</td>
<td>412.35</td>
<td>346.90</td>
<td>289.47</td>
</tr>
<tr>
<td>BIC</td>
<td>415.14</td>
<td>349.70</td>
<td>710.67</td>
</tr>
</tbody>
</table>


5.5.2.1 **Intervention effects on teacher behaviours**

Significant intervention effects are observed in the provision of positive and negative feedback.

*Positive feedback.* Overall, a rise in teachers’ levels of positive feedback towards the end of the year is observed and in the control group the rise is slightly larger than in the intervention group. These general effects mask an important three-way interaction effect. For positive feedback the pre-test observations in general are no significant predictors for the post-test observations. This is due to the varying strength of the predicting effect in the different groups and the direction of the effect being positive in some groups and negative in others. Careful examination of the data shows that in all groups low scoring students at the pre-test tend to see their levels of positive feedback increase slightly. Differences exist in the changes for students who score high for positive feedback in the pre-test in the different groups. In the control group equal changes in the outliers and lows can be observed for boys (a moderate rise). For girls in the control group we see a steeper increase of positive feedback for outlier girls than for low girls. In the intervention group the outlier boys, however, experience a decrease of positive feedback and the same goes for outlier girls who experience an even steeper decrease.

*Negative feedback.* As to the negative feedback we see that in general students who receive much negative feedback at the start of the year are more likely to receive more negative feedback at the end of the year. For boys this effect is significantly stronger than for girls. For control students the effect is significantly stronger than for intervention students.

*Positive to negative feedback ratio.* No significant intervention effect exists for the ratio of positive to negative feedback.

*Giving students a turn.* No significant intervention effect was found for giving students a turn.

5.5.2.2 **Intervention effects on student behaviours**

Significant intervention effects are found in two student behaviours: being off-task and social interactions with other students.

*Off-task.* In both the intervention and the control group we observe an overall and significant increase of off-task behaviour. In the intervention group, however, the rise is steeper. At the post-test no significant difference exists
for off-task behaviour between intervention boys and girls, but intervention boys do score significantly higher than control girls and control boys.  

Social interactions. Overall, a rise in students’ social interactions towards the end of the year is observed in all groups and in the intervention group the rise is significantly larger than in the intervention group. The post-test observations of social interactions are in general not significantly predicted by the pre-test observations. In the control group, however, we do see significant positive effects of the pre-test level of social interactions on the post-test. In the intervention group we do not see this effect. In total, the level of social interactions increases more in the intervention than in the control group, but it is not the students who are misbehaving in the pre-test who account for this rise.

5.5.2.3 Intervention effects on student sense of equity  
No significant differences were found for sense of equity in any of the groups.

5.6 Discussion and conclusion  
In the introduction we started from the recurring finding in previous research that boys receive more negative feedback than girls (Beaman et al., 2006). As to the levels of positive feedback, research has been inconclusive and Howe in 1997 suggested to focus on the ratio of positive to negative feedback instead. In the 2000s Myhill (2002), Swinson and Harrop (2005) and White (2010) found a positive correlation between the ratio of positive to negative feedback and students’ on-task behaviour. Experimental research confirming this causal relationship is, however, scarce (White, 2010; Howe, 1997). This quasi-experimental study was set out to investigate the effects of a professional development programme on teachers’ feedback provided to students and student responses to the changed instruction. Lack of random assignment and inability to sufficiently control for important confounding variables in quasi-experimental designs is known to be a major threat to establishing causality. Follow the suggestion of Harris et al. (2006) – we therefore present two alternative explanations for the apparent causal
associations. The first concerns a hypothesis of counterproductive effects and the second explanation involves confounding factors.

5.6.1 Positive approach with negative effects

Based on the model of van Veen et al. (2010) we assume that the professional development has an impact on students through its effects on teacher thinking and teacher behaviour. Indeed, we find effects of the intervention on teacher behaviour and the most striking effect can be found in the patterns of positive feedback towards students. The strength based approach of the professional development programme might have made a big impression on the teachers who – energised by the effects of positive attention on themselves – started experimenting with their patterns of positive feedback towards students. Previous empirical research on appreciative approaches to inquiry report that its positive nature supports change by fostering a sense of excitement in one’s professional practice (Chapman & Giles, 2009). In their excitement, however, the teachers might have failed to monitor unwanted effects of their changing feedback practices. Teachers involved in the intervention seem to turn around their positive feedback patterns abruptly: students who received lots of praise at the beginning of the year experience a drastic drop in their levels of positive feedback toward the end of the year and students who received hardly any at the beginning of the year see a steady growth. This drop of positive attention for the high outliers is significantly stronger in girls than in boys and it might be an overcompensation of the intervention teachers who have become aware of the fact that teachers might be more lenient towards girls (Öhrn, 1993). These changes in practice are quite a big difference with what we observe in the control group (and the type of changes students are more likely to be expecting). In the control group we saw that the highly praised boys succeed in maintaining their higher positive feedback and highly praised girls even strengthen their position and the gap with girls who receive less positive feedback grows. The difference is not significant, but we also see control girls’ calling out tripling (they overtake boys) and their hand raising drops more strongly than in other groups. This is not in line with Myhill’s (2002) finding that calling out is more prevalent in boys. It gives the impression that control girls are trading their hand raising for calling out and they are getting away with it since their levels of positive feedback grow
Effectiveness of Teacher Professionalisation

significantly more than in other groups and their levels of negative feedback drop more than in the intervention group. This does fit Myhill’s (2002) description of ‘good girls’ reinforcing their position as favoured students and getting away with more misbehaviour than boys (Öhrn, 1993, Younger et al., 1999). The intervention might have been successful in changing the levels of positive and negative feedback, but the ratio of positive to negative feedback was not affected by the intervention. The ratio increases for all groups, but the mean ratio is far from near the minimum 3 that was suggested by Swinson and Harrop (2005) and White (2010). No significant differences are found between any of the groups. Changes at both sides of the equation very much cancelled each other out. We therefore suggest to focus not only on the ratio but also on the absolute frequencies of positive and negative feedback. The ratio alone can give a false impression of nothing having changed.

Now, what are the effects of these changing feedback patterns in student behaviour? We observed some unexpected changes. Levels of off-task and unauthorised social interactions increased significantly more in the intervention group. What might have happened is that students who are used to receiving lots of positive attention, felt rejected or neglected when teachers’ positive feedback decreases and as a consequence they disengaged and turned to misbehaviour. Our results support this hypothesis in that sense that the post-test social interactions in the intervention group are not predicted by the pre-test. It is, thus, not the same students who misbehave in the beginning and at the end of the year. The qualitative work of some authors supports this hypothesis of student reaction against teacher treatment they perceive as being unfair. Taino (2011) shows how students can resist the gendered character of teachers addressing them by teasing or even openly criticising teachers on their practices. Öhrn (1993) in her qualitative study as well describes confrontational mechanisms. She describes how girls turn against their teachers to make them feel uncomfortable when they are unhappy with the treatment they receive. Resistance to teachers is found to spread quickly through the networks of students. Teachers, in the midst of trying to implement an appreciative approach, might have failed to notice or respond adequately to the counterproductive responses of students to the changing feedback patterns. Still, in this scenario we would expect to see a drop of students’ sense of equity in the ‘rejected’ group. No such effect could
be observed, however. We did not observe any differences in sense of equity between boys and girls. This is not in line with previous research showing that boys feel more discriminated against (e.g. Younger et al., 1999). Our data suggest sense of equity to be a quite stable construct, not easily influenced by the changing interaction patterns in the classroom. Previous research of ours (Consuegra, Halimi, & Engels, under review) confirms this stability showing that in boys with lower sense of equity the measure is uncorrelated to behavioural patterns and teacher feedback. Boys who experienced more positive feedback or a higher ratio of positive to negative feedback did not feel any more equally treated. Student sense of equity might conceptually be closely related to student sense of futility, measuring a stance of perceived deprivation and discrimination. Sense of futility has a large correlation with achievement (Brookover, Schweitzer, Schneider, Beady, Flood, & Wisenbaker, 1978) and sense of futility is proven to be an important predictor of school misbehaviour (Van Houtte & Stevens, 2008). This would turn our hypothesised relationship upside down. Sense of equity might not be the result of teacher equitable treatment but rather it is students entering class with a pre-defined ‘sense of equity’ which is very hard to influence for teachers. Garrahy (2001) identified teachers as being ‘gender-blind’ and unaware of gendered interactions. Students might be ‘gender-blind’ as well, but unaware of equitable interactions. Brophy in 1985 already emphasised the active role of student beliefs and student expectations in teacher-student interactions. Negative expectancy effects can originate in incorrect teacher expectations, but in incorrect student expectations as well. This intervention might have addressed students too much as being the passive recipients of interactions and teacher treatment. We suggest future professional development programmes to include a closer monitoring of the reponses of students to changing teacher practices. In the cycle of teacher inquire, the phase of trying out new teaching strategies and collecting and analysing data on the effects could have been effective in adjusting and correcting counterproductive effects. In the intervention programme, however, this phase was never fully reached. At the start of the programme teachers’ self-efficacy in regard to collaborating in professional learning communities indeed was very low (Schraepen et al., 2010) resulting in much time being spent on learning how to collaborate and learning how to inquire. It is nevertheless crucial to find ways in which teachers can be quickly engaged in
systematic monitoring of student outcomes and student beliefs and responses. Unsystematic experimentation can clearly generate counterproductive effects.

5.6.2 Confounding factors and suggestions for future research

An important limitation of this study is the absence of controls for confounding variables. The fact that we did not check for the multilevel effects of students being nested within classes and schools is a limitation we already mentioned in the results. Also, we did not enter covariates such as socio-economic background or migration background into the models to assess effects of other variables than gender alone (Beaman et al., 2006) nor did we check effects of gender composition of the classrooms which Drudy and Uí Chatháin (2002) found to have a significant influence on gendered interaction patterns. It is a real pity we did not succeed in getting more complex models to converge, since the significant differences in the pre-test observations between the intervention and control group indicate that such unaccounted for confounding variables might be in play.

Another limitation of our study is that we did not systematically analyse data on changes in teacher thinking at the post-test measurement which would help interpret the changes in feedback patterns. The overcompensation of teachers we hypothesise reminds of what Brophy in 1985 identified as teachers being overreactive and rigidly treating students as stereotypes rather than individuals. An explicit focus on differentiated instruction towards boys and girls might have unconsciously affirmed rather than challenged gender stereotypical thinking in teachers.

For a final limitation we return to van Veen et al.’s (2010) model and the environmental and macro conditions influencing the effects of professional development. This level has been identified as heavily underaddressed in research. Implementing interventions in several settings in order to assess which conditions need to be satisfied in order to generate effective teacher learning is essential (van Veen et al., 2010). Increasing the scale of research by implementing not one but several professional development programmes in various settings is needed. The increase of interventions, cases and observations would serve the investigation of school-level effects as well as allowing for a more fine-grained analysis of effects such as the ones described earlier in this section. More collaboration and tuning is needed
between research teams to allow for these larger-scale studies or to improve possibilities for comparison of results in multiple settings.
Chapter 6
Predicting Student Achievement
Chapter adapted from
Abstract
Teacher-student interactions have received increasing attention as a predictor of student educational outcomes. With regards to the underachievement of boys, observational studies have shown that boys receive more negative attention from their teachers than girls. In this mixed method study the relationship between teacher-student classroom interactions, students' perception of equity in the classroom, initial ability and grade retention were analysed. Questionnaires and standardised ability tests for math and reading were administered to 6380 first year secondary students. For a sub-sample of 129 students, fifteen-minute fragments of video-recorded lessons were coded. Information was gathered about students’ certification (grade retention). Logistic regression analyses provide evidence for the higher likelihood of students with a lower sense of equity to receive a certificate forcing them to repeat the current year. The odds to repeat a year were also significantly higher in students that are more off-task. Gender was no significant predictor and based on the regression results of multiple models it is hypothesised that gender effects are mediated through sense of equity. Interaction terms between gender and other dimensions of social identity show that no gender difference exists in students from low socio-economic background (SES). High SES girls are less likely than high SES boys to repeat a year. Results are discussed by means of relating them to the findings in the previous chapters and suggestions for further research are formulated.
6.1 Introduction

Review studies have indicated the ineffectiveness of grade retention to promote students’ learning in comparison with a similar level of academic achievement but moving up to the next grade (Hattie, 2009; Jimerson, 2001). Recent studies even point to negative effects of grade retention, for example on academic self-concept, early school leaving and participation and success in higher education (Lamote et al., 2013; Lamote et al., 2014). Students who are retained in a grade are significantly more likely to dropout (Jimerson, 1999; Roderick, 1994), to develop a lower self-efficacy and less self-regulated learning strategies (Rosario, Nunez, Valle, Gonzalez-Pienda, & Lourenco, 2013). Despite the controversy, grade retention is a wide-spread practice in Flemish compulsory education with more than one third of all students having repeated a year or more at the end of secondary education (Lamote et al., 2013). Several predictors of grade retention have been identified. In the present paper we review three categories of predictors: socio-demographic background variables, student classroom behaviour, and class and school environment context.

6.1.1 Socio-demographic predictors

A first socio-demographic predictor of grade retention is gender. As early as kindergarten, boys are more at risk for retention (Goldstein, Eastwood, & Behuniak, 2014). In secondary education, Lamote et al. (2013) showed that in the first grade one fourth of boys (25%) and one fifth of girls (20%) are grade retained. This gap increases towards the end of secondary education with one fourth more boys (39%) than girls (30%) that have doubled at least one year. Other students, which have been identified to be at greater risk for grade retention, are those from ethnic minority groups and those from low socio-economic backgrounds (Gottfredson, Fink, & Graham, 1994; Meisels & Liaw, 1993; Reynolds, 1991). A large-scale longitudinal study of Klapproth and Schaltz (2015) in secondary schools found that these three socio-demographic background variables’ effects hold when controlled for student achievement (operationalised by their grade point average). Students with similar school achievement are thus not at an equal risk of grade retention.
Klapproth and Schaltz (2015) in their discussion relate these results to teachers’ biased decision-making processes when deliberating on whether or not to retain a student. They explain, based on social cognition research, that decision-making about students occurs – in part – without much cognitive effort, relying “on social categories rather than on all information available” (Klapproth & Schaltz, 2015, p. 131). They continue with stating that the primary categories that are applied in these social-category-based judgements are gender and ethnicity. The study of Goldstein et al. (2014) confirms the strong predictor effects of teacher judgements and teacher biased expectations of students, and in specific for boys of low SES. Negative teacher expectation effects have also been observed in other assessment practices with easier grading for girls than for boys (Bonesronning, 2008; Lavy, 2004; Marcenaro-Gutierrez, & Vignoles, 2015). Lavy (2004) however does note that in comparison to blind grading or standardised tests grading, the regular assessment practices of teachers could reflect student effort and attitudes in addition to cognitive competence. If girls put systematically more effort into their work or if they display more desirable attitudes than boys, it could be argued that boys’ underachievement is the fair result of their less appropriate behaviour and attitudes.

6.1.2 Student behaviour and attitudes

Klapproth and Schaltz (2015) pointed out the limitation in their large-scale retention study that they have not taken into account the possible role of students’ misbehaviour in predicting the likelihood of students being retained. Hardly any study, they state, has investigated this relationship in the past and it is therefore impossible to draw any conclusions. Previous research (Howe, 1997), however, shows that boys display more disruptive behaviour than girls. Howe (1997) concludes her review on gendered classroom interactions that boys are more likely to call out in class and that they are more restless. Recent study results (Consuegra, Halimi, & Engels, under review) confirm these findings showing that boys are more off-task and boys call out more often than girls. However, evidence was also found for some categories of misbehaviour in which no significant gender differences were observed (e.g. talking to other students). Teacher bias favouring girls was hypothesised based on the observation that girls were not being reprimanded
for such misbehaviour, contrary to boys. Girls seemed to be getting away with this type of misbehaviour. It could be theorised that this subtle, nevertheless clearly present, teacher bias could lie at the base of boys’ higher levels of sense of discrimination. Teachers’ increased levels of reprimands to boys were explained both by boys being more disruptive as well as by an unconscious teacher bias (Consuegra et al., under review). Jussim and Harber (2005) recently reviewed the controversy about teacher expectation effects on student achievement and they concluded that negative expectancy effects should not cause for too much concern, since they are usually small and do not accumulate from year to year or from teacher to teacher. Nevertheless, in a quasi-experimental study aimed at decreasing teacher unconscious gender bias, Consuegra et al. (under review) concluded differently. Student sense of equity appeared to be a surprisingly stable construct (uninfluenced by changing teacher feedback patterns), and it was hypothesised that the sense of being discriminated against might be the result of subtle, yet continued and systematic teacher bias.

Student attitude effects have been studied more frequently than student behaviour effects. A recent study performed in South Africa shows that students’ positive attitude towards learning is associated with successful transition to the next grade. Such attitudes towards school and learning were also found to be gendered and more negative in boys than in girls (OECD, 2015). The PISA 2012 data showed that boys are for example more likely than girls to report that school is a waste of time and boys are less likely to agree that trying hard at school is important or that they enjoy receiving good marks. Warrington et al. (2000) found that boys’ study motivation is considerably lower than girls’, whereas Wang et al. (2011) showed that boys’ emotional school engagement (e.g., valueing of school education) was lower compared to girls. Wang et al. (2011) also found boys to report lower levels of attentiveness and school compliance.

6.1.3 School and class-level predictors of grade retention

Blaming individual students for their negative attitudes towards learning and school is, however, too simplistic. Research shows that students’ attitudes towards school are very much influenced by the attitudes of their peers as the result of (unconscious) socialisation processes. Younger and Warrington
(2000) found boys in comparison to girls to be more concerned with the way they are perceived by their peers. A study in Flanders confirms this higher sensitivity for peer pressure in boys (Van de Gaer, 2006).

Furthermore, boys and girls prefer friends of the same gender (Younger & Warrington, 2000), especially during adolescence. These same-sex subcultures emerge in early secondary education. Van Houtte (2004) identified these gendered subcultures as important vehicles for the transmission and reinforcement of boys’ negative attitudes towards school. Putting effort into school is perceived to be uncool and results in boys displaying more disruptive classroom behaviour as a way to establish their reputation among peers (Myhill, 2002).

When discussing the effects of peers on academic achievement, debates about class composition effects come into focus. Lamote et al. (2013) recently criticised past research on student achievement as being methodologically flawed. They point to – among others – the hierarchical structure of school education effects which has been disregarded in previous studies. They assessed school composition effects on math achievement in primary education, but found no direct school composition effects with respect to prior achievement, SES, ethnicity and sex. De Witte and Rogge (2013) confirm that little research has investigated institutional school and class-level effects on student dropout and achievement. They did however find evidence for compositional effects in Dutch middle schools. Students’ gender and ethnicity were found to have no considerable influence on students’ dropout probability after accounting for class and school influences such as the mean ability of students in the class, the class size and ethnic composition of the school. They concluded that especially in the first class of secondary education institutional environment effects are “extremely important” (De Witte & Rogge, 2013, p. 143). Control of class composition could therefore decrease dropout by adding some high achieving students in different classrooms. Based on their work it can be hypothesized that de facto heavily segregated schools or classes (along the lines of ability, ethnicity) are not beneficial for minority or underprivileged students. These findings are, however, also contested. In Flemish primary education, Agirdag, Van Houtte, and Van Avermaet (2012) found no effects of ethnic school composition on immigrant students’ self-esteem, for example. For native students, however, they did find that higher self-esteem was reported in schools with a greater
share of immigrant pupils and in those with less ethnic heterogeneity. They explain these results by referring to social comparison processes: “native pupils are likely to experience relative gratification as they compare themselves with their less esteemed ethnic minority peers” (Agirdag et al., p. 1152). Negative effects of ethnic minority concentration on immigrant students’ self-esteem were suppressed by the experience of a supportive relationship with the teacher. In immigrant students, the perception of a positive teacher treatment thus counters the negative effects of social comparisons with peers in segregated classrooms.

6.2 Research focus

Arguments can be provided for the hypotheses that boys are more likely to be grade retained due to (A) biased teacher decision making, (B) boys’ higher levels of misbehaviour and (C) social comparisons with peers and teacher treatment.

The present study aims at investigating these hypotheses in first grade Flemish secondary education students. The key research questions are:

a) To what extent do student background variables such as gender, SES, language background and migration background predict grade retention, when accounted for initial ability and previous school attainment?

b) To what extent does classroom misbehaviour such as calling out, being off-task and talking to other students predict grade retention?

c) To what extent does students’ perception of equity in the classroom predict grade retention?

d) To what extent does actual teacher treatment such as giving students a turn and providing negative and positive feedback predict grade retention?

6.3 Method

For this study, student surveys, video-recorded observations and information on student certification collected at the end of the first school year of secondary education in April-June 2013 have been used. Also, students’
scores on a standardised math and reading test administered at the beginning of the school year in September-October 2012 have been used in the analysis to account for effects of initial cognitive ability. Student tracking was also included as a measure of initial attainment. The regular track and the V-track which prepares for vocational education were compared. It is important to note that in this V-track the requirements to pass the grade [eindtermen] are less academically demanding than in the regular track. Part of the students enrolled in this track are students who are promoted to secondary education, without obtaining their primary education degree.

6.3.1 Participants

Survey and standardised tests sample. A representative sample of 59 schools in Flanders (Dutch speaking part of Belgium) was selected to participate in the study. Schools were selected using disproportionate stratified sampling with three strata: (1) geographical location in one of five provinces and the capital city, (2) rural and urban area, and (3) public and private educational network. All first-year secondary students of these 59 schools were selected to participate in the survey. In total, 6380 pupils (response rate 97%) completed the questionnaire in the beginning of the school year and at the end of the school year 2012-2013. The mean age of all pupils when taking the survey and the tests was twelve years old (SD=.51). In total, 6139 pupils participated in a mathematics test (response rate of 93%) in September-October 2012 and 5897 pupils participated in a reading ability test (response rate 89%) in the same period.

Classroom observation subsample. A purposeful heterogeneous subsample of six secondary schools was selected for participation in the quasi-experimental study in which additional observation data were collected. Schools were selected from the representative sample of 59 secondary schools The six schools varied in a range of conditions: public and private schools; urban and suburban areas; small, medium and large-sized schools; general, vocational and technical education schools. With support from the principal in each school five teachers were selected and asked to participate in the intervention study. Teachers had to be teaching the first year of secondary education in general subjects and those teachers with most contact
hours with the first-year students were addressed first for participation in the study. For each teacher, one of their first-year secondary classes was selected and one lesson was video-recorded. Classes were selected in such a way that those classes with most contact hours with the teachers involved in the study were selected. Teachers were told that any lesson would do for the video-recording unless the entire hour would be devoted to an examination or test. In the one or two cases that the observed lessons involved something other than whole-class instruction, we would revisit the teacher at another time for filming an extra lesson (which resulted in observations of whole-class teaching for all teachers). In total, 30 teachers and their 500 pupils were video-recorded twice (pre-test and post-test). Half of the 500 pupils were randomly selected for focal coding. Due to practical reasons (out of sight, quality of recordings) the teacher-student interactions of 180 of these 250 pupils were coded. In the post-test recordings only 129 of these 180 pupils could be coded (dropout rate 28%). The mean age of the observed students at the pre-test and post-test was twelve years old. The six schools were divided randomly over the control and intervention condition in such a way that a similar diversity of school size and school type (urban/suburban, public/private, tracks in higher grades) was present in both groups. In this regard it is important to note that the control group was actually a wait group, since these teachers would participate in the intervention the following year. It is the post-test sample of 129 students that will be used for the analyses in this study since the sense of equity at the end of the school year is more likely to be a result of actual teacher treatment than the sense of equity at the beginning of the school year.

6.3.2 Instruments

Survey. Questionnaires were administered by the research team during class. Pupils were asked questions about their socio-demographic background and were asked to rate the extent to which students are treated equally by the teachers. To measure socio-economic status, pupils were asked for the last occupation of their mother and father and the answers were converted into 8 categories of occupational prestige based on the Erikson, Goldthorpe and Portocarreiro (1979) classification. The highest of both parents was used as an indicator of the SES of pupils. For the analyses, SES was dichotomised into
the following categories: (a) unemployed or manual workers and (b) routine non-manual employees up to professionals, entrepreneurs and large proprietors. Migrant background was operationalised by segregating pupils whose grandmothers from mothers' side originated from a non-Western-European country. Foreign language background was registered by asking pupils which language they usually speak with their parents. If no Dutch was spoken with either of the parents, students were categorised as being from a foreign language background. To measure sense of equity the 6-item equity subscale of the What Is Happening In This Class? (WIHIC) survey was used (Dorman, 2003). The WIHIC is a recently developed and widely validated instrument that assesses classroom psychosocial environment in high schools. The scale is developed in such a way that it does not pre-define the basis for the discrimination that pupils' experience. Sample item: “I am treated the same as other students in this class”. All items are scored 0, 1, 2, 3, and 4, respectively, for the responses of almost never, seldom, sometimes, often and almost always. A principal component analysis was performed to verify the underlying structure of the construct and the reliability of the scale was checked (α = .893 in the large sample and α = .935 in the small subsample).

*Standardised tests.* Math ability was measured using the Dudal-orientation test Middle 6th Grade (Dudal, 2003). The test comprises 50 exercises and measures the knowledge of numbers, mental calculation, arithmetic, mathematical problems, and measures. Reading ability was measured using the CITO-test Group 8 (CITO, 2010). The test presents students with text fragments, followed by multiple-choice questions. The test is comprised of 25 questions that required students to retrieve and summarise information from the text, identify the goal of the text, describe relationships and define concepts. Both tests are aimed at assessing student ability at the end of primary education, and thus the initial ability of students when entering the first year of secondary school.

The standardised tests were provided to the 59 schools along with short manuals with clear instructions on how to correctly administer the tests. Mathematics and Dutch teachers were asked to administer the tests during regular classes. Some teachers were unable, or refused, to participate.
Classroom observation schedule. Two educational researchers acted as the observers, using continuous focal coding for 15-minute video-recorded fragments. The middle 15 minutes of all video-recorded lessons were selected for coding teacher-student interactions since these minutes were guaranteed to be instruction time. The first and last minutes of a 50-minute lesson are traditionally used to enter and leave the classroom, fill in the school diary, hand out corrected assignments, and take quizzes. A content analysis of pupils’ classroom interactions was performed using ten codes and a code for missing/out of sight (see Table 11). The codebook was based on the recurring codes in three recent studies (cf. Harrop & Swinson, 2011; Myhill, 2002; Younger & Warrington, 2002). Teacher-initiated behaviours as well as student-initiated behaviours were coded. Detailed selection criteria were consistently applied to determine if a behaviour could be assigned to a certain category. Several iterative cycles of coding, assessment of reliability and codebook modification were conducted by the two coders until reliability for all codes was high (Hruschka et al., 2004). To ensure that high levels of observer agreement would not be mere reflections of high chance agreement, a Krippendorf’s alpha (Krippendorff, 2013) was calculated. As Krippendorff’s alpha is known to be a more conservative index (Lombard et al., 2002), a more liberal criterion of .60 was used to determine if the coding of a variable was considered reliable. Training continued until kalpha's were above .60 and from that point on the two coders scored the remaining tapes, being aware that a selection of lessons, taken at random, would be scored by the other observer at the end. Table 11 gives an overview of the mean kalpha's for each code for both reliability checks combined. In line with recommendations of Harrop and Swinson (2007) we present a contingency table so that agreements and disagreements on occurrences and nonoccurrences of behaviour recorded by the two observers can be inspected. The C, H, T, R, N and P categories were analysed as events, generating the frequency or occurrence of each behaviour for each student. The O, F, I, and s categories were analysed as a state, generating the total duration of each behaviour as a proportion of time in sight for each student. The h code was also coded as a mean duration state (mean duration of hand up). A P/N code was calculated by subtracting the frequency of negative feedback from the positive feedback. Coding was performed using JWatcher (Blumstein et al., 2012). More
information on the observation schedule and protocol can be found in Appendix 2.

_School Achievement._ School achievement was operationalised by collecting data on the certificate that students receive at the end of the year. An A-certificate grants access to the next year without restrictions, a B-certificate implies two options: repeating the current year or promotion with restrictions from certain tracks. A C-certificate forces the student to repeat the current year. The U-certificate is granted when a student has re-examinations and an _uitgestelde_ or postponed decision. Since we are only interested in predicting the odds that a student will receive a C-certificate we dichotomised the variable with multiple categories into a dichotomous one: C-certificate (1) or A/B-certificate (0). The 13 cases with a U-certificate were recoded into missing values.

### 6.3.3 Analysis

Binomial logistic regression analyses are used to predict the probability that a student receives a C-certificate (1) rather than an A- or B-certificate (0). The independent variables need not be normally distributed, nor linearly related, nor of equal variance within each group. This allows us to predict grade retention based on the skewed observational data. However, the small sample size of the observation dataset (N=129) and the limiting sample size (10 observed pupils with C-certificate) provide serious restrictions to the number of predictors that can be included in the model. Baybak (2004) reviews simulation studies on overfitting in regression models and concludes that logistic models show reasonably stable results if the limiting sample size allows a ratio of approximately 10 to 15 observations per predictor. This implies that we can only predict the relationship between one behaviour and grade retention at a time. We will therefore perform separate regression analyses in the larger sample which allows for testing the effects of about 10 predictors (limiting sample size being 118).

All analyses are performed in the IBM SPSS Statistics Version 22 software package. The binomial assumption was tested in both datasets with a binomial test that indicates that the proportion of students receiving a C-
certification of .02 (large dataset) and .08 (small dataset) is lower than the expected .50, p = .000 (2-tailed).

### 6.4 Results

#### 6.4.1 Descriptive analyses

An overview of the descriptive results for all variables in both the large and the small sample is provided in Table 22.

**Table 22. Descriptive statistics for post-test survey and observation sample**

<table>
<thead>
<tr>
<th>Categorical variables</th>
<th>Large sample</th>
<th></th>
<th>Small sample</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td>Certificate</td>
<td>5830</td>
<td>91.4</td>
<td>129</td>
<td>100</td>
</tr>
<tr>
<td>A/B-certificate (0)</td>
<td>5712</td>
<td>89.5</td>
<td>119</td>
<td>92.2</td>
</tr>
<tr>
<td>C-certificate (1)</td>
<td>118</td>
<td>1.8</td>
<td>10</td>
<td>7.8</td>
</tr>
<tr>
<td>Gender</td>
<td>6380</td>
<td>100</td>
<td>129</td>
<td>100</td>
</tr>
<tr>
<td>Boys (0)</td>
<td>3435</td>
<td>53.8</td>
<td>63</td>
<td>48.8</td>
</tr>
<tr>
<td>Girls (1)</td>
<td>2945</td>
<td>46.2</td>
<td>66</td>
<td>51.2</td>
</tr>
<tr>
<td>Socio-economic status</td>
<td>6194</td>
<td>97.1</td>
<td>126</td>
<td>97.7</td>
</tr>
<tr>
<td>Low SES (0)</td>
<td>1363</td>
<td>21.4</td>
<td>27</td>
<td>20.9</td>
</tr>
<tr>
<td>Middle/High SES (1)</td>
<td>4831</td>
<td>75.7</td>
<td>99</td>
<td>76.7</td>
</tr>
<tr>
<td>Language background</td>
<td>6356</td>
<td>99.6</td>
<td>129</td>
<td>100</td>
</tr>
<tr>
<td>Dutch with at least one parent</td>
<td>5718</td>
<td>89.6</td>
<td>105</td>
<td>81.4</td>
</tr>
<tr>
<td>No Dutch with either parents</td>
<td>639</td>
<td>10.0</td>
<td>24</td>
<td>18.6</td>
</tr>
<tr>
<td>Migration background</td>
<td>6170</td>
<td>96.7</td>
<td>120</td>
<td>93.0</td>
</tr>
<tr>
<td>West-European (0)</td>
<td>4929</td>
<td>77.3</td>
<td>88</td>
<td>68.2</td>
</tr>
<tr>
<td>Non West-European (1)</td>
<td>1241</td>
<td>19.5</td>
<td>32</td>
<td>24.8</td>
</tr>
<tr>
<td>Track</td>
<td>6380</td>
<td>100</td>
<td>129</td>
<td>100</td>
</tr>
<tr>
<td>Regular track</td>
<td>5640</td>
<td>88.4</td>
<td>118</td>
<td>91.5</td>
</tr>
<tr>
<td>Preparing for vocational track</td>
<td>740</td>
<td>11.6</td>
<td>11</td>
<td>8.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Continuous variables</th>
<th>Large sample</th>
<th></th>
<th>Small sample</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Math initial ability</td>
<td>26.79</td>
<td>10.77</td>
<td>26.62</td>
<td>11.01</td>
</tr>
<tr>
<td>N=5982</td>
<td></td>
<td></td>
<td>N=123</td>
<td></td>
</tr>
<tr>
<td>Reading initial ability</td>
<td>14.41</td>
<td>5.37</td>
<td>14.04</td>
<td>14.04</td>
</tr>
<tr>
<td>N=5749</td>
<td></td>
<td></td>
<td>N=117</td>
<td></td>
</tr>
<tr>
<td>Sense of equity</td>
<td>3.12</td>
<td>.79</td>
<td>2.92</td>
<td>.91</td>
</tr>
<tr>
<td>N=5888</td>
<td></td>
<td></td>
<td>N=124</td>
<td></td>
</tr>
<tr>
<td>On-task</td>
<td>80.34</td>
<td>15.32</td>
<td>80.34</td>
<td>15.32</td>
</tr>
<tr>
<td>N=129</td>
<td></td>
<td></td>
<td>N=129</td>
<td></td>
</tr>
</tbody>
</table>
6.4.2 Predicting grade retention based on survey data

We adopt a stepwise approach in which we first predict grade retention based on student demographic background variables alone (Model 1). Second we add the independent variables measuring initial ability and achievement (Model 2). Third, we include student sense of equity (Model 3) and finally we include the three interactions terms of gender and SES, language background and migration background (Model 4). All models are statistically significant (see Table 23) and the fourth model explains most variance in certification. The Cox & Snell $R^2$ squared indicates that 3% of variance is explained and the Nagelkerke $R^2$ squared indicates 23% of variance explained. “Low $R^2$ values in logistic regression are the norm and this presents a problem when reporting their values to an audience accustomed to seeing linear regression values. (...) However, they may be helpful in the model building state as a statistic to evaluate competing models.” (Hosmer & Lemeshow, 2005, p.167). Hosmer and Lemeshow (2005) argue that the true measure of fit is one based strictly on a comparison of observed to predicted values from the fitted model. Our final model succeeded in correctly classifying 98.9% of cases and indicates that the likelihood that a student is forced to repeat the current year is significantly related to math ability, track, sense of equity, migration background and the interaction of SES and gender. Students scoring high on the mathematics ability test at the beginning of the year and students enrolled in the V-track (preparing for vocational education) are less likely to double a year. Students reporting a higher sense of equity are also less likely to receive a C-certificate. The odds that a student with a non-Western-European migration background has to repeat a year are three times as high than for other students. As to the interaction effect of gender and SES we see that no gender difference exists in the low SES group. The gender effect is only present in the high SES group, with girls from high SES being less likely than boys from high SES to receive a C-certificate. Figure 13 shows the percentage of students receiving a C-certificate in the different groups.
Table 23. Odds of grade retention (C-certificate) results from logistic regression analyses

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Model 1 Exp(B)</th>
<th>Model 2 Exp(B)</th>
<th>Model 3 Exp(B)</th>
<th>Model 4 Exp(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>.021***</td>
<td>.231**</td>
<td>1.052</td>
<td>.583</td>
</tr>
<tr>
<td>Gender</td>
<td>.545**</td>
<td>.590*</td>
<td>.703</td>
<td>2.494</td>
</tr>
<tr>
<td>SES</td>
<td>.574*</td>
<td>.858</td>
<td>1.195</td>
<td>2.457</td>
</tr>
<tr>
<td>Language background</td>
<td>1.623</td>
<td>1.751</td>
<td>1.601</td>
<td>1.415</td>
</tr>
<tr>
<td>Migration background</td>
<td>3.234***</td>
<td>2.476**</td>
<td>3.113**</td>
<td>3.785**</td>
</tr>
<tr>
<td>Track</td>
<td></td>
<td>.127***</td>
<td>.095***</td>
<td>.099***</td>
</tr>
<tr>
<td>Math initial ability</td>
<td>.920***</td>
<td>.901***</td>
<td>.901***</td>
<td>.901***</td>
</tr>
<tr>
<td>Reading initial ability</td>
<td>.945</td>
<td>.955</td>
<td></td>
<td>.957</td>
</tr>
<tr>
<td>Sense of equity</td>
<td></td>
<td></td>
<td>.520***</td>
<td>.516***</td>
</tr>
<tr>
<td>Gender*SES</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender*language</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender*migration</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N pupils</td>
<td>5516</td>
<td>4891</td>
<td>4658</td>
<td>4658</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>76.056***</td>
<td>97.108***</td>
<td>118.564***</td>
<td>123.314***</td>
</tr>
<tr>
<td>df</td>
<td>4</td>
<td>7</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>Cox &amp; Snell R²</td>
<td>.01</td>
<td>.02</td>
<td>.03</td>
<td>.03</td>
</tr>
<tr>
<td>Nagelkerke R²</td>
<td>.08</td>
<td>.15</td>
<td>.22</td>
<td>.23</td>
</tr>
</tbody>
</table>

Note. *p<.05, **p<.01, ***p<.001

Figure 12. Percentage of students receiving a C-certificate at the end of the first year of secondary education, by gender and SES
6.4.3 Predicting grade retention based on observation data

The model predicting certification based on levels of off-task behaviour is significant ($\chi^2(1)=9.129; \ p=.003$) and explains 6.8% (Cox & Snell $R^2$) to 16.3% (Nagelkerke $R^2$) of variance in certification. The model correctly classified 92% of cases. Students who are more off-task are more likely to receive a C-certificate (Exp(B)=1.079, $p=.002$)(see Figure 14 for boxplots by certification).

![Boxplot for being off-task, by certification](image)

Figure 13. Boxplots for being off-task, by certification

The model predicting certification based on levels of on-task behaviour is borderline significant ($\chi^2(1)=3.627; \ p=.057$) and explains 2.8% (Cox & Snell $R^2$) to 6.6% (Nagelkerke $R^2$) percent of variance in certification. Students who are more on-task are less likely to receive a C-certificate (Exp(B)=.965, $p=.048$).

Other student behaviours such as talking to other students ($\chi^2(1)=.095; \ p=.758$), calling out ($\chi^2(1)=.398; \ p=.528$), raising hands ($\chi^2(1)=1.330; \ p=.247$),
p=.249), responding to teachers ($\chi^2(1)=.509; p=.475$) do not significantly add to the prediction of grade retention.

None of the models including teacher behaviours are significant in predicting grade retention. The model with as a predictor the ratio of positive to negative teacher feedback is borderline significant ($\chi^2(1)=3.587; p=.058$) but the individual variable is not statistically significant ($p=.118$). The models for positive feedback ($\chi^2(1)=2.811; p=.094$), negative teacher feedback ($\chi^2(1)=.206; p=.650$), and giving a turn ($\chi^2(1)=.152; p=.697$) are not statistically significant.

### 6.5 Discussion

This study was set out to investigate gendered grade retention in the first year of Flemish secondary education. We theorised that several mechanisms could lie at the basis of gender disparities in grade retention. First, teacher decision making whether or not to retain a student could be bias as with gender being a key social category used for (unconsciously) judging students (Klapproth & Schaltz, 2015). Students with equal ability might be judged differently based on stereotyped expectations of teachers. Second, boys might be bringing grade retention upon themselves by being more disruptive in class. Previous research has shown that boys call out more and are more off-task than girls (Howe, 1997). Third, boys might be underachieving as a result of lower self-esteem resulting from social comparison processes with their peer girls who are favoured.

The study shows that gender is no significant predictor of grade retention when accounted for student initial ability and sense of equity. The interaction of gender with SES is significant, though. In low SES groups, boys and girls are at equal risk for grade retention. In high SES groups, girls are significantly less likely than boys to be retained. Lower initial ability in mathematics increased the likelihood of grade retention (reading ability was not significant) and students with a non-Western-European migration background were also at higher risk for grade retention. Students reporting lower levels of perceived equity and student being more off-task are more likely to be retained. Teacher feedback patterns were no significant predictor
of being grade retention and students in the V-track were less likely to be retained.

6.5.1 Teacher expectations: track and migration background effects?

Students enrolled in the track preparing for vocational education are less likely to be retained. This finding is contrary to the findings of De Witte and Rogge (2013) in the Netherlands that, when controlled for initial ability, found no school track effect. They explain this absence of effect by referring to the Dutch situation in which there is a relatively great mobility between the tracks. This is quite opposite to the situation in Flanders. Mobility between tracks in Flanders is small and teachers might be more lenient in promoting V-track students in comparison to students in the regular track. Further research is needed into the effects of tracking in combination with the observed mobility between tracks because they affect teacher expectations which might be the underlying predictors of grade retention. Besides the effect of V-track we also see that migration background remains to be a significant predictor of grade retention, despite accounting for track, SES, initial ability and sense of equity. Immigrant students in comparison to native students from equal ability, sense of equity and socio-economic status are still at higher risk for grade retention. This could point to teacher bias effects with regards to ethnic background, which is in line with the findings of Klapproth and Schaltz (2015).

6.5.2 The mediation of gender effects

When observing the effect of gender in the four prediction models we see that gender loses its predictive power when sense of equity is entered (Model 3). In the final model, only the interaction between gender and SES is significant. This might indicate a mediation with the effect of gender being carried by sense of discrimination. Direct gender effect might only exist in high SES students. Agirdag et al. (2012) also identifies student perceptions of teacher treatment (support) as a mediator of effects in the association between socio-demographic background (composition) and student self-esteem (a predictor of achievement). Agirdag et al. also report different explanatory
mechanisms in different socio-demographic groups (they compared immigrant and native students).
A recent study of Van Maele, Huyge, Vantieghem and Van Houtte (2004) on the same Procrustes dataset also presents evidence of mediated effects of gender. A drop in significance values for student gender is observed in the prediction grade retention when adding gender typicality and school engagement in the equation. A student who feels like an atypical boy or girl and students who feel less engaged in the school are more at risk for grade retention and when entering these variables into the model, the p-value of gender drops (but remains significant). Again these variables bear some element of social comparison in them (feeling typical in relation to the perceived norm, feeling engaged as feeling part of the school culture and fitting in). All of these results emphasise the importance of students’ individual perceptions of the classroom and school environment for predicting academic achievement. It might not be class and school composition as such that influence student achievement, however, students’ perceptions and experiences that are the result of composition might be the true active ingredient in predicting academic achievement. This could explain why Lamote et al. (2013) found no significant direct effects of composition on dropout.

6.5.3 Are high SES boys bringing it upon themselves?

We found that being off-task increases the likelihood to be retained. Given the small sample in which we had to perform analyses, we were not able to assess simultaneously the effects of other variables or possible interactions. In a previous study, however, we observed that high SES boys, in particular, are significantly more off-task than others (Consuegra, Halimi, & Engels, under review). We related this to anti-schoolish cultures that foster disruptive behaviours in students and which are more predominant in boys than in girls (Myhill, 2002). We hypothesised that high SES boys get away with being off-task more easily than low SES boys, since we found no parallel significant differences in teacher reprimands (which are in general correlated to student misbehaviour). We interpreted this as a favouring of high SES boys. However, we would like to nuance this finding based on the present results. High SES boys might be getting away with their misbehaviour in that
sense that they are not being reprimanded for in during class. This study clearly shows that, despite the absence of reprimands, these boys still are at an increased risk for grade retention in comparison to high SES girls. The interaction between gender and SES is significant despite the other general gender effect disappeared when sense of equity was entered. Are high SES boys under the impression that they are being favoured and do they feel privileged or equally treated, while only in the end of the year they are being presented with the costs of their misbehaviour? This could explain why the high SES boys’ increased risk for grade retention even when including into the model the effects of sense of equity and initial ability. It is regularly stated that boys are bringing underachievement upon themselves by being more disruptive. We would suggest that this might indeed be the case for high SES boys, with teachers not correcting them or warning them in advance for the possible consequences of their ‘cool’ misbehaviour.

6.6 Conclusion

This study provides support for several explanatory mechanisms in the effects of gender on grade retention. Teacher expectations, student misbehaviour and student social comparison effects might all be in play simultaneously. Also, explanatory mechanisms appear to differ in different groups of boys and girls, with the intersection of gender and SES being an important one to take into consideration in further research. One of the recurrent critiques in previous research was that studies on academic achievement were methodologically flawed, due to the omission of class- and school-effects. We did not check for composition effects as such, but for students perception of equity which is the result of a social comparison with peers in their class or school. We found that this perceived sense of equity is a significant predictor of early grade retention. We discuss several other studies pointing to the mediation of gender effects on academic achievement as well and we observe that measures of self-concept which are the result of social comparison are also found to mediate gender effects on grade retention in these studies (Van Maele et al., 2014). We believe that great potential lies in the further exploration of student perceptions of equal teacher treatment. We argue that in this regard, it is student perceptions of
interactions rather than actual interactions that are the true predictors of academic achievement.

Bearing this in mind, we formulate the following suggestions for policy makers to decrease grade retention and academic underperformance of boys: (1) stimulating a more information-based judgement in teachers’ deliberation whether or not to retain a student (rather than a stereotyped expectations judgement), (2) providing teachers with the opportunity to investigate their own teaching practice and their students’ actual behaviour in order to identify and re-engage students who are displaying at risk behaviour (such as high levels of being off-task), (3) providing opportunities for teachers and students to discuss feelings of privilege and discrimination and to allow for collaborative inquiry into the possible discrepancy between reality and perception. We understand that these are no clear-cut, easy-to-implement tips and tricks. These would do no justice to the complex reality of gendered academic achievement. Tackling one mechanism without simultaneously tackling the other mechanisms as well, is unlikely to yield effects. Plugging one hole in a sinking ship will not prevent it from sinking. Plugging a considerable amount of holes, however, will keep it afloat for several more miles.
Chapter 7
General Conclusion and Discussion
7.1 Introduction

This doctoral dissertation research project sets out to investigate gendered teacher-student interactions in secondary education.

In the general introduction we illustrated that boys perform lower than girls on several indicators of school achievement such as early school leaving, grade retention and representation in lower tracks. The gap emerges in early secondary education and continues to grow towards the end of compulsory education. This lower performance is not only a relative underachievement when compared to girls, but also it is an absolute underachievement with regards to boys’ own potential. OECD in 2015 concludes her extensive report “The ABC of Gender Equality in Education” that the gender gap that is observed in PISA scores is much smaller than the gender gap observed in school marks. Boys and girls with similar scores on the PISA tests are not performing equally well in school. Something in school is thus happening which impedes boys to perform to their full potential.

Previous research has identified several predictors of boys’ scholastic underachievement. Risman and Davis (2012, p. 747) state “gender inequality is produced, maintained and reproduced at each level of social analysis”: individual, interactional and institutional. Bearing the valorisation potential of research findings in mind, we chose as the main focus of this dissertation research project the interaction between teachers and students. The mediation of teachers’ gendered expectations through classroom interactions receives our special attention. Brophy in 1985 successfully illustrated the great complexity of such teacher expectation effects, which involve and interplay between teacher and student identity, teacher and student perceptions, and teacher and student behaviours.

Six limitations in previous research on gendered teacher-student interactions were identified and six studies were designed to address the gaps in existing research. The following paragraphs will summarise the studies and their takeaway messages. Afterwards we will discuss the limitations of our own research and formulate suggestions for future research. Finally, we conclude with suggestions for practitioners and policy makers.
7.2 Summary of results

Study 1. Reviewing European Research

The first limitation we identified was the supremacy of USA research in the field of gendered teacher-student classroom interactions. We performed a systematic search and review of European studies and discussed the results of 18 studies. The fact that boys receive more negative feedback from their teachers than girls is confirmed. We discuss three particularities in the European research. First a topic, which has produced inconsistent findings in both USA and European studies: positive feedback. Our review shows conflicting results with studies in the UK studies pointing to boys receiving more positive feedback than girls while the studies performed in continental Western-Europe show the opposite. We hypothesise that cultural differences between Anglo-Saxon and other western cultures might exist with regards to positive and appreciative approaches to learning (as opposed to deficit-based approaches). Second we discuss a topic, which has emerged as a new focus of interest in European research: sexual temptation in teacher-student relationships. Indications are found for romanticised pressure in the higher grades of secondary education. Male teachers are found to pay more attention to girls as they get older and girls report using their ‘feminine wiles’ to influence their male teachers. Previous research has disregarded the issue of sexuality in secondary education. Contextual and cultural differences could explain this lack of attention to sexuality in previous research since it has mostly been USA based and in comparison to continental Western-European countries sexuality is more of a taboo there. In many European countries governments support sex education and public health policies support widespread public education campaigns through internet, television, films, radio, billboards etc. with a focus on safe and pleasurable sexuality (Alford & Hauser, 2000). Further research into sexuality in teacher-student relationships is needed. Third, we discuss the scarcity of experimental and intervention research aiming at change and not only describing gendered interactions in both USA and European research. Alex Harrop and Jeremy Swinson (2000, 2011) who also extensively studied gendered teacher-student interactions were the only ones performing intervention research in which they trained teachers to increase positive interactions. The study was not focused on gender but their suggestions are promising, showing that an appreciative and
positive pedagogy can be used by teachers to improve the ratios of positive to negative feedback to students (which are found to be lowest for boys). A higher ratio of positive to negative interactions is related to more on-task behaviour, but little empirical research has actually investigated the causal relationship between both. We conclude that more research is needed on the effects of ‘positive pedagogy’ as a strategy to counter teacher gender bias. Also, we suggest appreciative approaches to teacher professionalization to facilitate teachers’ to question their own practices and to become aware of possible gender bias in their teaching.

**Study 2. Increasing teacher awareness**

A second major limitation that was identified in previous research is the contradictory findings concerning the existence of teacher gender bias (e.g. Howe, 1997; Younger et al., 1999). Some authors claim that teacher thoughts are not (or not so much) gendered, and that boys bring negative feedback and attention upon themselves by being more disruptive. The claim that teacher biased expectation effects exist, is based on the assumption that teacher expectations and thoughts are biased. We critique the methods based on self report used in earlier research (e.g. teacher focus groups, self report Likert items scales) which might not have been the best approaches to capture teacher gender bias it is not self-evident to be conscious of and recognise one’s own bias and teachers might cover up their bias due to social desirability. We have investigated teachers’ implicit and explicit thoughts when recalling on a video-taped lesson and have investigated the gendered character of utterances concerning student character traits and student misbehaviour. Qualitative analyses of the transcripts of these video-stimulated recall interviews show a clear awareness gap in teacher thoughts. Teachers’ explicit recall is unbiased, while implicit recall of problematic traits and behaviour is three times more likely to concern boys than girls. We have also observed some qualitative differences in positive attributions to boys and girls with the latter being praised almost exclusively for their good academic behaviour while boys are also being praised for their social traits such as having a sense of humour, being extrovert, or being able to build good relationships with other students. We have also investigated whether video-stimulated recall interviewing increased teacher awareness of gender
bias. This was the case for some teacheres who experienced moments of ‘becoming aware’ while many remained blind to their gendered interpretation of student behaviour and according reactions. They showed the same bias during and when reliving the lesson. In depth reflection and inquiry is not triggered by merely watching and recalling own practices. We suggest future research to examine what types of facilitation are needed for video-stimulated recall to become video-stimulated reflection. The teachers involved in our study who did become aware of their gender bias mentioned the importance of the safe environment with an unknown researcher interviewing them and the low-stakes character of the interview (no evaluation purpose). We suggest recall on an individual basis or with an independent facilitator might be a needed first step before continuing video based reflection with peers or teacher educators.

**Study 3. Gendered classroom interactions**

A long tradition of over 40 years of research into gendered teacher-student interactions exists. Research agrees with regards to one recurrent finding: boys in general receive more negative feedback from their teachers than girls. Two remarks are usually added: not all boys and not only boys receive negative feedback. Despite the fact that this commentary has been formulated over several decades already, still it is unclear which boys (and which girls) account for the increased levels of negative feedback. Part of the reason why so little is known about which boys and which girls are the methodological choices made in previous studies. Classroom observation research has relied on mostly on teacher-focal coding procedures registering which behaviours the observed teacher addresses to boys or girls (e.g. Younger & Warrington, 2002; Jones & Dindia, 2004). In order to investigate more than the gender dichotomy alone (Beaman, Wheldall & Kemp, 2006) we opted for a mixed methods design in which students were both surveyed and observed. Also students completed a standardised test for mathematical and reading ability. This allowed us to connect student socio-demographic background variables to student perception of equity, student (mis)behaviour and teacher feedback. Differences in classroom interaction patterns at intersections between the genders and several other dimensions of social identity (SES, migration background, language background) were assessed. Our findings confirm that
boys receive significantly more negative feedback from their teachers than girls and they report lower levels of perceived equity in the classroom. No significant gender differences were found with regards to positive (and neutral) teacher feedback. Boys increased negative feedback appears to be – at least in part – justified by their calling out and being off-task more often. However, we also observe some indicators of negative teacher bias towards boys. Boys and girls talk to their neighbours just as much. However, for girls this misbehaviour is not related to criticism from the teacher while for boys it is. The findings from the video-stimulated recall interviews point in a similar direction showing that teachers mistake boys’ talk for task unrelated chat while they interpret girls’ talk more easily as task-related conversations. Teachers indeed appear to favour girls. Not all girls, however. Our intersectionality-approach brings to the surface one group of girls that is exceptionally vulnerable, namely girls who do not speak the instruction language with either of their parents. We observe that these girls (unlike girls in general) are off-task significantly more often and they raise their hand less often and for less long, they get the turn from teachers less often and they respond to questions less. Language background was also a major predictor of student sense of equity, with a Cohen’s d effect size of .90 indicating a high practical relevance. Another intersect with gender which shows different relationships in boys and girls is initial ability in maths and reading. In girls we see that high ability is related to higher levels of activity in teacher-student interactions. In boys no such relationship exists. This could confirm Myhill’s (2002) findings that high performing boys in middle school tend to join the pattern of passivity and indifference of underachievers. She relates this to boys’ construction of masculinity and the peer group pressure not to appear too eager or enthusiastic in class. The relationship between student gendered anti-schoolish attitudes and classroom interaction patterns can be explored in further research. Finally we see an interesting interaction between SES and gender as well, showing that high SES boys are most disruptive in class, while they do not receive more negative feedback. High SES boys might get away with misbehaviour more easily too. We suggest future research on intersection of gender with other dimensions of social identity. Language background in specific deserves more attention, especially since this variable has been disregarded in previous research and we find it to be
one of the most important variables to help explain variance in perceived sense of equity and actual inequities in interactions.

**Study 4. Feminised Teacher Professionalism**

When presenting the above-cited results concerning negative teacher bias towards boys, usually the question follows whether it is related to the feminisation of teaching. We refer to previous research on the topic showing no teacher gender effects on teacher treatment of boys and girls (Howe, 1997) or student achievement (Helbig, 2012). The absence of gender differences, however, does not imply that teaching is not gendered. We refer to Acker (1990) when she critiques that much research on gendered and feminised professions studies gender as one of the many variables and often lacks a solid embedding in gender theory. We set out this fifth study as an attempt at connecting theories on teacher professionalism in the field of teacher education research and theories on gendered organisations in the field of feminist research. Britton (2000) explains that three definitions of ‘feminised’ professions can be used: the numerical feminisation (number of women in the job), the inherent feminisation (soft and caring profession as stereotypically feminine), and the hegemonic feminisation and masculinisation (the celebration of stereotypically feminine or masculine traits). Our focus of research concerns this hegemonic feminisation or masculinisation. We identify interesting analogies and contradictions between the discourses in both fields in regard to the ‘norm’ and the ideals in teacher professionalism. The move towards ‘teachers as researchers’ is both framed as (A) a conservative movement in which stereotypically masculine traits such as rationality, management and objectivity are emphasised and overshadow the stereotypically female traits such as emotionality and subjectivity, and as (B) a progressive movement in which the gap between theory and practice is contested and teachers are stimulated to engage in knowledge production of teaching rather than only consuming the knowledge of teaching which is produced by ‘experts’. Since a large amount of literature was found concerning these issues of teacher professionalism, we opted for a study design aimed at the assessment and confirmation of a hypothesised model. Teacher gender was assumed to have an indirect and mediated effect on measures of self-reported teacher-student relationship quality. The mediators
that were assumed to carry the effect of gender are: (a) teacher restricted emotionality and the disability to express and ventilate emotions, and (b) teacher extended professionalism and the degree to which teachers rely not only on intuition but also on professional literature, systematic inquiry and reflection and see their role as teachers broader than merely teaching students in the class. The structural equation model shows a full mediation for gender effects on the provision of autonomy and a partial mediation on the provision of structure (with women scoring slightly higher on structure provision). The effects of restricted professionalism and extended professionalism are of similar power with opposite signs, therefore cancelling each other out (partially). We conclude that the gender of teachers does not really matter that much with regard to the (self-reported) quality of teacher-student relationship. It does matter, nevertheless, when bearing in mind that teachers are modelling gendered paths to reach qualitative relationships with their students. The results of our structural equation model also show that extended emotionality and extended professionalism are both positively related to the measures of teacher-student relationship. Both are also not negatively related to each other, no indications of conflict between the more ‘rational’ and more ‘emotional’ approaches to teaching thus are observed. The claim (A) that ‘teachers as researchers’ would result in a decline of professionalism is not supported. Our findings are more in line with hypothesis (B) showing that teachers who report high quality relationships with their students can successfully combine high scores on both dimensions of professionalism. We suggest further research is needed to check, however, if teacher perceptions of professionalism and teacher-student relationships are in line with student perceptions.

**Study 5. Effectiveness of Teacher Professionalisation**

The fifth central limitation in previous research we address is the lack of intervention research that tries to improve and change practice instead of only describing and explaining it (Beaman et al., 2006). In the systematic literature review we suggested that the appreciative approach to learning is promising for reducing negative teacher gender bias in two ways: first as an appreciative pedagogy that increases attention and sensitivity for positive traits and behaviour instead of a general deficiency-based approach and negativity bias,
second as an appreciative approach to teacher professional development and inquiry of teachers into their own and their peers’ practice. The video-stimulated recall study hinted to the importance of a safe and appreciative environment to facilitate teacher inquiry into their own practice since it is not self-evident for teacher to question their own teaching and to become aware of possible bias in their practice. Based on the work of Swinson and Harrop (2005) and White (2010) we hypothesise an effect of the ratio of positive to negative teacher feedback on student on-task behaviour. A pre-test post-test quasi-experimental design with control and experimental group was adopted. The effectiveness of an appreciative collaborative inquiry-based teacher professional development programme was assessed on several levels as proposed by van Veen et al. (2010): teacher behaviour, student behaviour, and student perception of equity in the classroom environment. Generalised linear mixed models are used to assess inference in the strongly skewed count data. With regards to positive teacher feedback a significant three-way interaction effect is found for the interaction between gender, pre-test level of positive feedback and intervention. We see that in all groups the students receiving little praise at the beginning of the year see their levels of positive feedback increase slightly. For the students with high levels of positive feedback at the beginning of the year we see that for girls who were already favoured with positive feedback from the start, there is a steep increase of positive feedback towards the end of the school year while for boys who were favoured at the beginning the amount of positive feedback remains the same or increases slightly. In the intervention group the pattern is dramatically different. Girls favoured with positive feedback at the start, saw their positive feedback level drop steeply and boys favoured at the start experienced a more moderate drop. The gap in positive feedback increases in the control group, in the intervention group it decreases (but also drops in general since we see a downward equalisation). With regard to student behaviour responses to the changed teacher feedback pattern we see that the levels of off-task behaviour and talking to peers in both the intervention group and the control group increase, with a significantly steeper rise in the intervention group. No effects were observed in student sense of equity; these remained stable throughout the school year, regardless of the intervention, regardless of gender. Since quasi-experimental designs lack random assignment of and inability to sufficiently control for important confounding variables, we present two
alternative explanations for the apparent causal associations. First we discuss
the lack of systematic monitoring of student responses to changed teacher
practices and experiments in the intervention group. Brophy in 1985 already
pointed to students’ active role in the reproduction of or resistance to teacher
treatment. Privileged boys and girls (in terms of much praise at the beginning
of the year) might feel rejected by the teachers when their privilege drops
drastically and those students might respond by losing interest. Still, we
would expect to see changes in students’ sense of equity in such scenario. No
such effects are observed, however. Might the behavioural response in
students be of an unconscious nature (in line with teachers’ gender
blindness)? Sense of equity in any case appears to be a very stable construct
that does not easily change, regardless of changed teacher feedback patterns.
Students are likely to enter secondary education with a certain sense of
discrimination, which is the result of many years of subtle and continuous
teacher gender bias in primary education and kindergarten. Effects of
changed teacher feedback might only generate changes in students’ sense of
equity after a longer period of continued efforts of changed practice and our
intervention of one school year might not have been long enough to capture
any effects. A second explanation we present for the results is related to the
limitations of the study to account for confounding variables, in specific
possible multilevel effects concerning the nestedness of the observed students
in classes in schools. The highly skewed observation data and relatively small
number of cases (especially in the post-test) presented us with serious
challenges for analysis. We suggest future research to implement the
intervention in a much bigger scale in a practitioner research project and to
collect observation data by means of self-analysis of the teachers. Drudy and
Uí Chatháin’s (2002) work can serve as an interesting example in this
respect. Taking into account a system for close monitoring of student
behavioural responses to changed teacher treatment is crucial to prevent
counterproductive effects of the intervention.

**Study 6. Predicting Grade Retention**

The final gap in the existing evidence-base we aim to fill is the scarcity of
research linking observed student classroom interactions to academic
achievement. In this final study we investigate the likelihood of boys and
girls to be grade retained at the end of the first year of secondary education. Early grade retention is proven to be a predictor of dropout and participation and success in higher education. We assess the predictive power of several variables: gender (and its interactions with other dimensions of social identity), initial mathematical and reading ability, student misbehaviour, teacher feedback patterns, and students’ perception of equity of teacher treatment. Binomial logistic regression is used and again, the skewed observation data present us with challenges for analysis since the samples are small and models with more than one independent variables would not meet the prescribed ratio of predictors to limiting sample size (cases with event = 1, number of students that are grade retained). For testing the effects of the behaviours we performed separate regressions. The findings indicate that the effect of gender disappears when entering sense of equity into the model. This could indicate a mediation effect, which is also suggested in previous research (e.g. Van Maele et al., 2014; Agirdag et al., 2012). The interaction between gender and SES is significant and points to no gender difference in low SES students and a higher likelihood of high SES boys to be retained in comparison to high SES girls. Students that are more off-task are also more likely to be grade retained (high SES boys are most off-task). Neither positive teacher feedback, nor negative feedback or the ratio are significant predictors. This final study shows that several mechanisms simultaneously affect student achievement: student perceptions of discrimination, student off-task behaviour, and some student demographic variable effects that are not accounted for when entering sense of equity and initial ability measures. For future research we suggest focusing on other measures of student achievement since the number of students in early secondary education being grade retained is very small, making it difficult to capture and assess the complexity of predictory mechanisms that are at play.

7.3 Limitations and suggestions for future research

Our research was faced with several limitations. We focus on two limitations and discuss them in depth. We explain their impact on our findings and on our ability to effectively answer our research questions. We justify the choices we made when confronted with them and we reflect on suggestions to overcome the limitations in future research.
The first limitation we will discuss is the overwhelming complexity of teacher-student interactions and the apparent impossibility to capture all relevant variables, leaving the researcher with multiple doubts about confounding variable effects.

The second fundamental self-critique on failing to take into account in the professional development programme the active role of students in gendered classroom interactions.

### 7.3.1 Methodological catch up

The general introduction gives a clear overview of the many theories that exist with regards to explaining gendered student achievement. Improving teacher-student interactions is only one of the many possible strategies to reduce gendered underachievement. Sure, one has to make choices when designing research, but we regret that we were unable take into account some institutional-level effects which have proven to be underaddressed in previous research (De Witte & Robbe, 2013). Especially in our intervention study it would have been very informative to control for confounding variables at the classroom and school level. Gender composition, SES composition, teacher (gendered) professional identity, school inclusive policy and culture ... all variables, which are very likely to explain variance in the effects of the professional development programme. Including such effects in the model might have resulted in even more fine-grained information on how to effectively support teacher professional development to tackle teacher bias.

In the systematic review we discussed that several methodological shifts have occurred in the field of classroom observation research over the past decades, all aimed at better grasping the great complexity of gendered interactions:

- A move from studying the quantity to studying the quality of interactions
- A shift towards the study of other variables than student gender alone
- The use of mixed methods research investigating student perceptions, teacher perceptions and actual behaviour observed in classrooms

Theoretic explanatory models grew in complexity. However, the nature of classroom observation data does not always allow for testing such complex models. The observation data we want to analyse are hierarchically structured
(students are nested within classrooms within schools), temporally dependent (pre-test and post-test data), non-normally distributed and in addition labour-intensive to collect and analyse resulting in relatively small samples. Our systematic review shows that, apart from the self-analysis study of Drudy and Uí Chatháin (2002) the student sample sizes exceed 200 in only two studies. Several studies use similar behavioural codes and observation techniques, but when reviewing previous research, we did not find any discussion of the topic of statistical challenges, which, in retrospect, is quite peculiar.

When using student focal coding and registering the occurrence of behaviours such as calling out, raising hands and response to teachers, the distribution of all these behaviours proves to be strongly skewed (for example from our dataset, see Figure 15).
Figure 14. Histograms for calling out and raising hand
Clearly, the data are nowhere near normally distributed. However, in classroom observation research that adopted similar procedures and coding categories, the distribution of the data and ways of dealing with it is not mentioned (Myhill, 2002). We find it difficult to assume that the same type of observation data would follow a bell curve in other studies. All the same, means are reported as the indicators of centrality and it is not specified which particular “test for statistical significance” is used to assess the differences between groups. The studies of Younger and Warrington (2002), Swinson and Harrop (2009) and Drudy and Ui Chatain (2002) all report using parametric tests to analyse their data. The data differ from ours since these studies adopt teacher-focal coding instead of student focal coding. Still, it concerns count or proportional data which are generally assumed not to be normally distributed. All three studies fail to discuss the assumptions that need to be met in order for parametric tests to produce valid results.

We urge future research to discuss the nature of their data and the selection of statistical tests more extensively, in order for other researchers to be able to build on the knowledge of others. Our search into appropriate techniques has led us to generalised linear mixed modelling (GLMMs). GLMMs are relatively new and only have been implemented in statistical software packages since the 2000s. GLMMs are a very powerful and flexible technique to analyse data with different distributions which are geographically and temporally dependent. Two reviews on the method in the field of clinical medicine (Casals et al., 2014) and ecology and evolution (Bolker et al., 2009) show that the method is rapidly gaining popularity. O’Hara and Kotze (2010) show that the method performs better than analyses with log-transformed data when handling count data. The metaphor of the Bed of Procrustes can very well be re-used in this context. We should not change our data to be able to use inappropriate tests, but we should find appropriate tests to handle our complex data.

Another strategy we suggest is the development of robust practitioner research self-analysis instruments for (student) teachers that can serve both as an instrument for data collection as well as an instrument for professional learning. Drudy and Uí Chatháin (2002) show that it is indeed possible to
develop such an instrument and the expanding of data sets would make computation of GLMMs (and the testing of complex models) more feasible.

7.3.2 Students as active participants rather than passive recipients

The second central reflection we want to share is our relative lack of attention for the active role of students in teacher-student interactions. In the intervention research clearly students show resistance to changes in teacher treatment and the teacher learning communities did not engage in data collection and monitoring of student responses to their changed practice. This has resulted in an increase of misbehaviour in the intervention group, which in turn increases students’ likelihood for grade retention. In retrospect, we would design the professional development programme in such a way that it involves inquiry into privilege and discrimination by teachers in collaboration with not only their colleagues but also with the students themselves. Research has shown that students do not always experience a learning environment in the way it was intended by the designers. Yet, pupils’ perceptions of a learning environment (and bias in it) determine how much they will learn and how effective a learning environment will be (Könings, Brand-Gruwel, & van Merriënboer, 2005). Könings et al. (2005) discuss a new approach to changing learning environments in which designers, teachers and students collaborate.

Despite our study failing to take the active role of students into full consideration in the design of the professional development program, we do feel that we have much added to the student-focused analysis of gendered classroom interactions.

One of the recurrent recommendations in review studies in the field of gendered classroom interactions was the need to investigate differential classroom interactions along the lines of more than student gender alone (Beaman et al., 2006): not all boys and not only boys are being disadvantaged in classroom interactions. Apart from attainment and ethnicity not many additional background variables have received attention (Howe & Abedin, 2013). This gap in research was - in part - the result of the way in which much previous research was performed, namely by using teacher-focal
observations. Younger & Warrington (2002) observed 13 teachers on four to six occasions. The number of reprimands and statements of praise uttered by teachers was registered and, in addition, for each utterance it was recorded whether the feedback was directed to a boy or a girl. Also it was registered how many times boys or girls request for help or direct public comments to the teacher. Swinson and Harrop (2009) adopt a similar teacher-focal coding system, counting the occurrences of questions, and the approval and disapproval directed at either boys or girls. These teacher-focused observation schedules allow for commenting on the total amount of feedback directed toward the genders, but nothing is known about the distribution of the feedback: do some boys or all boys in general receive more negative feedback? Also, the teacher-focal observation methods do not allow for registering multiple dimensions of student background at the same time since this would be practically infeasible. Myhill (2002) adopts a student-focal observation approach. In each observed class a high achieving boy and girl and a low achieving boy and girl were selected for observation and the observation schedule recorded the number of times the children were seen engaging in a variety of interactions such as calling out or raising hands. This approach provides detailed information on students-initiated behaviours. However, in this approach no information was collected on teacher behaviours so the dynamic relationship between individual student and teacher behaviours could not be assessed.

In this dissertation we aimed at combining the best of both techniques. We sampled students and we registered teacher-initiated behaviour directed at those students in specific, and we coded student behaviours as well. The fact that we also surveyed the observed students enabled us to connect the student behaviour data to the student background and student perception data. Clearly, the approach has paid off showing some important interaction effects between gender and ability, gender and SES and gender and language background. The effect of not speaking the instruction language with either of the parents is both surprising and not surprising. Clearly, classroom interactions are mostly verbal and language ability appears to be a self-evident predictor of variance in students. However, the eight review studies we reviewed in the systematic literature review do not mention language background as a variable that has been studied previously. We strongly advise future research to take this variable into account since it showed to be
the predictor of sense of equity with the highest practical relevance of all predictors (Cohen’s $d = .90$). Such effects cannot be neglected, and an important interaction was found with gender, with girls from a foreign language background being very underrepresented in teacher-student interactions, while we know from linguistic research that teacher-student interactions are of vital importance for learners’ language development especially when home and instruction languages do not coincide (Gibbons, 2003).

We suggest linguists, educational scientists and gender researchers to join forces in the study of classroom interactions.

Another suggestion for further research we would like to present is the study of gender identity and gendered anti-schoolish cultures on classroom interactions, rather than the gender dichotomy. Vantieghem et al. (2014b) and Huyge et al. (2015) have studied more flexible operationalisations of gender with quantitative and large-scale techniques in the survey datasets of the Procrustes programme. Further research is needed to link these concepts of gender identity to the observational data and to assess how gender identity can add to understanding intra-gender differences in interactions.

### 7.4 Recommendations for practitioners and policy makers

We structure the discussion of recommendations for practitioners and policy makers along the lines of three questions we have repeatedly received from professionals themselves.

#### 7.4.1 Tips and tricks for handling boys and girls?

We are often asked to provide concrete teacher-proof tips and tricks with regards to teaching boys and girls. We have to disappoint the ones asking, since no such advice can be formulated based on our research findings. Should boys and girls be treated equally at any time? No, Brophy in 1985 already proposed that it is not always desirable to treat all students as equal. Optimal instruction, he argued, implies some degree of individualisation (and
gendered) instruction. Differentiated instruction, however, should not be based on generalisations and stereotypes but on the nuanced reality as much as possible.

In 1993 Bailey suggested – based on mainly descriptive studies – techniques such as deliberate turn-taking and increasing wait-time before calling on students (which would increase the number of female volunteers) in order to achieve equal classroom interactions. Our critique on Bailey’s suggestions would be that it fails to acknowledge the multidimensional identity of students. Our findings show that students from a foreign language background (and girls in specific) are hesitant to raise their hand and they tend to drop their hand quickly after raising it. Introducing a longer waiting time might be interesting for girls in general, but this specific group of girls and other students from a foreign language background would most certainly not be served with it. Hence, each tip or trick for teaching “boys and girls” will fail to take into account the complexity of reality. It is important to move beyond gender as an isolated category of analysis and teacher instruction and to take into account simultaneous interactions between multiple dimensions of diversity.

Brophy (1985b) suggests – and we follow his suggestions – that teachers should counter negative teacher expectancy effects by keeping expectations for individual students current and up to date by monitoring student progress closely and stressing present performance over past history.

A suggestion that Bailey (1993) formulates, which fits very well with this idea of increased monitoring that Brophy describes, is sensitising teachers about the mechanisms of gender bias and getting feedback on interaction patterns in their own classrooms. Olivares and Rosenthal (1992) propose similar interventions of teachers developing skills to become aware of their biased perceptions and elements in the learning environment which inhibit students’ development. The above recommendations advocate for an increased professionalism in the form of teacher awareness and monitoring skills rather than teacher-proof tips and tricks on how to handle boys and girls. We believe this is indeed the way to proceed given the gaps in evidence, since implementing changes without proof of effectiveness is a risky business and could result in unsought effects. The review by Plumm (2008) on technology in education and gender-biased interactions very well illustrates these risks. They explain how it had been assumed that technology
for the most part was gender-neutral and that its introduction would undo gender-bias. However, as the review shows, biases have simply been reshaped and have in some cases even increased gender-bias in student perceptions, teacher perceptions and curricular materials. Plumm (2008) concludes that educators and researchers need to be very cautious when attempting to reduce or eliminate gender-bias from classroom interactions with clear-cut remedies, since the mechanisms are so complex that without solid empirical data it is hard to predict effectiveness (or counter-productivity) of interventions.

7.4.2 Introducing standardised testing

In Flanders no central or standardised testing exists to evaluate student school performance. Based on the observation of OECD (2015) that great discrepancies exist between PISA test scores and school performance, one could argue that standardised tests should be introduced in school evaluation to reduce teacher bias effects. We do not feel that such an introduction, however, would solve the problem. It might be effective in tackling direct teacher expectation effects such as teachers scoring equal tests higher or lower when they are respectively the work of boys or girls. The indirect expectation effects, however, remain unchallenged. Also, standardised test are often mistaken for being ‘unbiased’. They might simply substitute old biases with new biases. Curricular materials and evaluation instruments themselves are reproducing gender inequality. Previous research has demonstrated that evaluation instruments affect boys’ or girls’ performance (Stobard, 1994; Stobard, Elwood, & Quinlan, 1992). It has been suggested, for example, that men perform better on multiple-choice tests since they rely more the ‘masculine’ trait of risk taking (Livingston & Rupp, 2004). Beside the instruments being biased themselves, the introduction of standardised tests might unfairly claim to rule out teacher bias while neglecting the many indirect bias effects and thus blaming (or meriting) individual students for their good or bad performance. The self-efficacy of students who suffer from discrimination might drop resulting in lower academic achievement in a vicious cycle. We suggest tackling both direct and indirect teacher gender biases by supporting teachers in the development of their professionalism.
7.4.3 We are only human, is this realistic?

We can not deny that tackling gendered underperformance is complex and not self-evident. Our intervention study shows that developing effective professional development programmes is not self-evident and changed teacher behaviours will not automatically result in improved student response with many risks for unwanted effects in students (who can actively resist the changes). Our professional development programme focused on teacher collaborative appreciative inquiry but teachers did not succeed in completing the inquisitive cycle. The last phase in the inquisitive cycle in which teachers systematically try out new practices and monitor and evaluate student effects was never reached. A more intensive programme that can extensively address both the competences to inquire, appreciate and collaborate and the topic of biased interactions should be introduced. Such an intensive program, however, needs to be facilitated by school leaders providing teachers with time off from their regular assignment to ensure continued engagement. Still, even in such conditions we expect the shift towards teacher researchers to be extremely challenging. Teacher professionalism research has been battling the gap between theory and practice for more than a century (Dewey, 1907), and only minor bridges have been established so far (Zeichner, 2010; Grossman, 1989; Vanderlinde & Van Braak, 2010). The paradigm shift in teacher professionalism that is proposed is fundamental and it involves the abolishment – or at least upsetting – of power hierarchy in educational knowledge production (Christianakis, 2008). Researchers and teachers must both arise and break with the chains of tradition. In order to accelerate the shift we suggest a 360-degree-approach with intervention in the entire spectrum of teacher professionalisation (Feiman-Nemser, 2001): from pre-service teacher education programmes to continued in-service professional development. If we want teachers to develop a research stance, we need to prepare them for such practices as much as possible in teacher education (where students should have ample opportunity to switch from theory to practice and back to integrate theoretical and practical knowledge). If we want student teachers to develop a research stance, teacher-educators should develop such a research stance as well. This concept has been underexposed in literature, however (Tack & Vanderlinde, 2014). The great challenges and resistance to ‘teacher research’ should not prevent us from investigating the path, though. Feminist literature shows that utopian thinking might be
frowned upon as irrational or naive but it has an important role in theory development and scientific progress (Acker, 1990). Teachers and educational researchers must join forces and not give up on continuing to puzzle out which strategies are effective in eradicating – or at least minimising – unconscious biases in education. In utopian education, all students (and teachers) perform to their full potential.
References


Tack, H., & Vanderlinde, R. (2014). Teacher educators’ professional
development: towards a typology of teacher educators’ researcherly

Taino, L. (2011). Gendered address terms in reproach sequences in classroom

invasive spread. *CAB Reviews, 7*(16), 1-10.

teachers: immersion in practice or reflection on action. *Teaching and
Teacher Education, 16*(5), 575-591.

professional learning and development. Best evidence synthesis
iteration. University of Auckland: New Zealand Ministery of
Education.

Tomlinson, C. A., Brighton, C., Hertberg, H., Callahan, C. M., Moon, T. R.,
response to student readiness, interest and learning profile in
academically diverse classrooms: a review of literature. *Journal for the

process of teacher change. *Teaching and Teacher Education, 28*, 728-
739.

Tsouroufli, M. (2002). Gender and Teachers’ Classroom Practice in a

http://www.unicef.org/earlychildhood/index_40749.html

what do they mean for girls’ and boys’ school achievement? *Journal of
Youth and Adolescence, 25*, 73-88.


ongekwalificeerde uitstroom tot 2007]. Leuven: Steunpunt SSL, rapport nr. SSL/OD1/2009.25


Appendices
Appendix 1.
Survey instruments

Student reported: Sense of Equity in the Classroom
Adapted subscale from What is Happening in this Class (WIHIC) questionnaire (Dorman, 2003)

Indicate how often you feel the following statements are applicable for teachers in your school (almost never, seldom, sometimes, often, almost always):
1. The teachers gives as much attention to my questions as to other students’ questions.
2. I get the same amount of help from the teachers as do other students.
3. I have the same amount of say in class as other students.
4. I am treated the same as other students in class.
5. I receive the same encouragement from the teachers as other students do.
6. I get the same opportunity to contribute to class discussions as other students.

Teacher reported: Teacher Provision of Structure
Adapted subscale from Teacher as social context (TASC) questionnaire (Wellborn, Connell, Skinner, & Pierson, 1992)
Indicate whether the following statements are good illustrations of your relationship to the pupils in this school. (Not at all true, not very true, sort of true, very true)

1. When I discipline a student, I always explain why.
2. I let students get away with things I normally wouldn’t allow.
3. I find it hard to be consistent with students.
4. I don’t always have time to follow through with students.
5. I talk with students about my expectations for them.
6. I try to be clear with students about what I expect of them in class.
7. I change the rules about school work for students.
8. Sometimes I feel I don’t make my expectations clear to students.
9. When students don’t comprehend the material, I take a different approach.
10. When students don’t understand something, I explain it in a lot of different ways.
11. I can’t tell when students are keeping up with me.
12. It’s hard to know when students are ready to go on to new material.
13. I show students different ways to solve problems.
14. I find it difficult to tell when students need help.
15. I find it hard to teach students in a way they can understand.

Kruis uw mening aan die het beste weergeeft hoe u uw relatie met de leerlingen op deze school aanziet. (Helemaal niet akkoord, niet akkoord, akkoord, helemaal akkoord)

1. Als ik een leerling straf, leg ik altijd uit waarom.
2. Mijn leerlingen komen weg met dingen die ik normaal gezien niet zou toelaten.
3. Ik vind het moeilijk om consequent te zijn tegenover mijn leerlingen.
4. Ik heb niet altijd tijd om mijn leerlingen op te volgen.
5. Ik praat met mijn leerlingen over mijn verwachtingen naar hen toe.
6. Ik probeer duidelijk te zijn tegenover mijn leerlingen over wat ik van hen verwacht in de klas.
7. Ik verander de regels over schoolwerk niet voor mijn leerlingen.
8. Soms heb ik het gevoel dat ik mijn verwachtingen niet duidelijk maak naar mijn leerlingen toe.
9. Wanneer mijn leerlingen de leerstof niet begrijpen, probeer ik een andere benadering.
10. Als mijn leerlingen iets niet begrijpen, leg ik het op verschillende manieren uit.
11. Ik kan niet zeggen of mijn leerlingen mij kunnen volgen (wanneer ik iets uitleg).
12. Het is moeilijk om te weten wanneer mijn leerlingen klaar zijn voor nieuwe leerstof.
13. Ik toon mijn leerlingen verschillende manieren om problemen op te lossen.
15. Ik vind het moeilijk om mijn leerlingen iets aan te leren op een manier dat ze het begrijpen.

**Teacher reported: Teacher Autonomy Support**
Adapted subscale from Teacher as social context (TASC) questionnaire (Wellborn, Connell, Skinner, & Pierson, 1992)

Indicate whether the following statements are good illustrations of your relationship to the pupils in this school. (Not at all true, not very true, sort of true, very true)

16. I try to give students a lot of choices about classroom assignments.
17. My general approach with students is to give them as few choices as possible.
18. It’s better not to give too many choices to students.
19. I have to lead student through their schoolwork step by step.
20. When it comes to assignments, I’m always having to tell students what to do.
21. I find myself telling students every step to make when it come to schoolwork.
22. I let students make a lot of their own decisions regarding schoolwork.
23. I can’t let students to do things their own way.
24. I can’t afford to let students decide too many things about schoolwork for themselves.
25. I explain to students why we learn certain things in school.
26. I encourage students to think about how schoolwork can be useful to them.
27. It is difficult to explain to students why what we do in school is important.

Kruis uw mening aan die het beste weergeeft hoe u uw relatie met de leerlingen op deze school aanziet. (Helemaal niet akkoord, niet akkoord, akkoord, helemaal akkoord.

16. Ik probeer mijn leerlingen veel keuzes te geven als het gaat over taken of opdrachten.
17. Mijn algemene aanpak is om mijn leerlingen zo weinig mogelijk keuzes te geven.
18. Het is beter om niet te veel keuzes te geven aan mijn leerlingen.
19. Ik moet mijn leerlingen stap voor stap doorheen hun schoolwerk begeleiden.
20. Als het over taken of opdrachten gaat, moet ik mijn leerlingen altijd zeggen wat ze moeten doen.
21. Ik merk dat ik aan mijn leerlingen elke stap moet toelichten als het gaat over schoolwerk.
22. Ik laat mijn leerlingen zelf veel beslissingen nemen omtrent hun schoolwerk.
23. Ik kan mijn leerlingen de dingen niet op hun eigen manier laten doen.
24. Ik kan het me niet veroorloven om mijn leerlingen zelf veel te laten beslissen over hun schoolwerk.
25. Ik leg uit aan mijn leerlingen waarom we bepaalde dingen leren op school.
26. Ik moedig mijn leerlingen aan om na te denken over hoe schoolwerk nuttig kan zijn voor hen.
27. Het is moeilijk om uit te leggen aan mijn leerlingen waarom het belangrijk is wat we op school doen.

**Teacher reported: Teacher Extra-Role Behaviour**
Adapted from Extra-role behaviour towards students scale (Somech & Drach-Zahavy, 2000).

To what extent do you, as a teacher of this school, behave as follows? (almost never, hardly ever, rather not, in between, rather do, sometimes, almost always)
1. Stay after school hours to help students with class materials.
2. Arrive early for class.
3. Acquire expertise in new subjects that contribute to my work.
4. Stay in class during breaks in order to listen to my students.
5. Go to school on my free days to prevent problems in my class.
6. Prepare special assignments for higher and lower level students.

Als leerkracht van deze school, in welke mate handel je als volgt?
(absoluut niet, niet, eerder niet, tussenin, eerder wel, wel, absoluut wel)
1. Na de schooluren op school blijven om leerlingen te helpen met hun leerstof.
2. Vroegtijdig in de klas aanwezig zijn.
3. Expertise verwerven over nieuwe onderwerpen die bijdragen tot mijn werk.
4. Tijdens pauzes in de klas blijven om naar mijn leerlingen te luisteren.
5. Naar school gaan op vrije dagen om problemen in mijn klas te voorkomen.
6. Speciale taken voorbereiden voor leerlingen die op een hoger en lager niveau werken.

**Teacher reported: Experimentation and Reflective Practice**
Adapted subscale from Dutch School Improvement Questionnaire (Gijsel, Sleegers, Stoel, & Krüger, 2009).

Indicate to what extent the following statements are applicable to your as a teacher in this school. ((almost) never, sometimes, often, (almost) always).
1. I observe colleagues’ lessons to learn from them.
2. I try out new knowledge and skills in my lessons.
3. I make my own teaching materials.
4. I use pupils’ reactions to improve my classroom teaching.
5. I discuss problems in my classroom teaching with others in order to learn.

Geef aan in welke mate onderstaande uitspraken voor u van toepassing zijn als leerkracht van deze school.

4-punten schaal: (bijna) nooit, soms, vaak, (bijna) altijd
1. Ik observeer lessen van mijn collega’s om zo van hen bij te leren.
2. Ik probeer nieuwe kennis en vaardigheden uit in mijn lessen.
3. Ik stel mijn eigen lesmateriaal op.
4. Ik gebruik de reacties van mijn leerlingen om mijn eigen klaspraktijken te verbeteren.
5. Ik besprek problemen in mijn klaspraktijk met anderen om zo bij te leren.

Teacher reported: Keeping Up To Date
Adapted subscale from Dutch School Improvement Questionnaire (Gijsel, Sleegers, Stoel, & Krüger, 2009).

Indicate to what extent the following statements are applicable to your as a teacher in this school. ((almost) never, sometimes, often, (almost) always).
1. I take the initiative to work on my own professional development.
2. I take part in further training and in-service training even if it is not compulsory.
3. I read professional literature.
4. I study textbooks and lesson material thoroughly and on a regular basis.

Geef aan in welke mate onderstaande uitspraken voor u van toepassing zijn als leerkracht van deze school. ((bijna) nooit, soms, vaak, (bijna) altijd)
1. Ik neem initiatief om aan mijn eigen professionalisering te werken.
2. Ik neem deel aan nascholingen en bijkomende opleidingen, zelfs als het niet verplicht wordt.
3. Ik lees vakliteratuur.
4. Ik bestudeer handboeken en lesmateriaal diepgaand en op een regelmatige basis.

Teacher reported: Teacher Trust in Students
Adapted subscale from Faculty Trust Scale (Hoy & Tschannen-Moran, 2003).

Indicate whether you agree or not with the following statements about the pupils in this school. (strongly disagree, don’t agree, in between, agree, strongly agree)
1. The pupils in this school care about each other.
2. I can count on students to to their work.
3. I believe the pupils in this school are able to learn.
4. I trust students in this school.
5. The pupils in this school are closed.

Geef aan in welke mate u akkoord kan gaan met volgende uitspraken over de leerlingen van deze school. (Helemaal niet akkoord, niet akkoord, tussenin, akkoord, helemaal akkoord)
1. De leerlingen van deze school geven om elkaar.
2. Ik kan erop rekenen dat de leerlingen hun werk doen.
3. Ik geloof dat de leerlingen op deze school competent zijn om te leren.
4. Ik heb vertrouwen in de leerlingen op deze school.
5. De leerlingen zijn gesloten.

Teacher reported: Restricted Emotionality
Adapted subscale from Gender Role Conflict scale (Blazina, Pisecco, & O’Neil, 2005).

To what extent do you agree with the following statements? (Strongly disagree, do not agree, not really agree, rather agree, agree, strongly agree).
1. I have difficulty telling others I care about them.
2. Strong emotions are difficult for me to understand.
3. Expressing feelings makes me feel open to attack by other people.
4. It’s hard for me to talk about my feelings with others.
5. It’s hard for me to express my emotional needs to others.
6. When I am personally involved with others, I do not express my strong feelings.
7. I often have trouble finding words that describe how I am feeling.
8. I do not like to show my emotions to other people.
9. Telling others about my strong feelings is difficult for me.

In welke mate gaat u akkoord met volgende uitspraken? [Kruis aan wat voor u het best past] (helemaal niet akkoord, niet akkoord, eerder niet akkoord, eerder akkoord, akkoord, helemaal akkoord)

1. Ik heb het moeilijk om aan anderen te vertellen dat ik om hen geef.
2. Intense emoties zijn voor mij moeilijk om te begrijpen.
3. Ik voel me kwetsbaar ten opzichte van anderen wanneer ik mijn gevoelens uit.
4. Ik heb het moeilijk om met anderen over mijn gevoelens te praten.
5. Ik heb het moeilijk met het uiten van mijn emotionele behoeftes tegen anderen.
6. Wanneer ik een persoonlijke band heb met anderen, geef ik geen uiting aan mijn intense gevoelens.
7. Ik heb vaak moeite om woorden te vinden die beschrijven hoe ik me voel.
8. Aan anderen toon ik niet graag mijn emoties.
9. Ik heb het moeilijk om mijn intense gevoelens aan anderen te vertellen.
Appendix 2.
Student-centered classroom observation schedule and protocol

This observation schedule and protocol was developed by Els Consuegra, Myriam Halimi and Nadine Engels. Please refer to the instrument as follows: Consuegra, E., Halimi, M., & Engels, N. (2015). Student-centered classroom observation schedule and protocol. In E. Consuegra, Gendered Teacher-Student Classroom Interactions: perception, reality and professionalism. Brussels, Belgium: Vrije Universiteit Brussel.

Camera positioning

Cameras were positioned before the start of the lesson that needed to be filmed. If students experience the set up of the cameras they might be more aware and preoccupied with being filmed than necessary. Observations were scheduled in such a way that the observed lesson was taking place just after a break or after a ‘jump-hour’ in which no lessons were taught in the classroom. This gave the researcher ample opportunity to set up the cameras.

Two cameras were used to film the lessons. Cameras were positioned in such a way that the faces of students, the teacher and the blackboard were recorded as much as possible. Since student-focal coding would be used, student faces received priority in camera positioning. In most of the classes one camera was placed in front of the classroom and one was placed in the back. In some classrooms, however, students sat with their backs against the wall and camera at the back would only capture the backs or the sides of students (with the students closest to the camera covering the others). In this case, both cameras were to be positioned in front of the classroom. Cameras (in the front and back) were aimed at capturing the diagonal of the classroom. If cameras were positioned straight along the length of the classrooms the students closest to the camera were covering-up the students behind them. In the diagonal direction this covering-up was less of a problem.

Researchers did not rely on the batteries of the cameras and cameras were always installed in such a way that they were plugged in and connected to the
power supply. Since many classrooms did not have enough electric plugs, researchers brought along extension cables and plug distributors.

**Explanation to teachers**

Teachers were told that no specific type of lessons was being targeted for the observations. We did ask to avoid lessons in which students would be engaged in a test throughout the entire lesson hour. If a lesson that was eventually observed did not hold enough instruction time for us to analyse, the researcher proposed to revisit the teacher and film another lesson.

Teachers were told that a lesson would be filmed and that student behaviours were going to be analysed afterward. They were guaranteed that no other people than themselves or the researchers would be able to view the recordings. Teachers were explicitly told that the aim of the observation was not a quality assessment of their own teaching practice. They would not receive any feedback from the researcher. The teachers involved in the experimental track knew that they would watch their own lesson recording relatively soon after the observation. They knew that the researcher would merely be there to assist in watching and to support the process of recalling what had happened during the lesson.

**Explanation to students**

At the time of the classroom observations, students were already introduced with the research programme since they already filled in the questionnaire. They were told that the project also included tape recordings of lessons in order for the researchers to get a detailed image of how students of their age experience school. Students were assured that no evaluation would follow and that nobody, except for the observed teachers and the researchers, would get to view to recordings. They were explicitly told that no feeding back of the recordings would happen to themselves, to parents, to unininvolved teachers, to principals or anyone else. Also it was stressed that the recordings would not be showed or published on social media (YouTube, Facebook), television or any other such medium. During the recording the researcher
would sit in the back of the classroom, out of sight for the students as much as possible.

**Field notes during filming**

The observing researcher made field notes during the recordings. When unexpected things happened that could not be recorded (events in the hallway for example) this would be noted down. Also the researcher live scored the interactions (any interaction) of students with the teacher on a paper schedule. The X-axis would hold all students and the Y-axis was divided in 30-second sections. The researcher would tally the number of interactions for each student in each 30-second fragment. This was done for two reasons. First, it served as a system for the researcher to gain some preliminary insight into the interactions (actual coding results would only be generated much later). This was important since clips needed to be selected for the video-stimulated recall interviews that quickly followed the observation. Second, the preliminary observations supported the later video-based coding when it was not clear from the video to whom the teacher was addressing a specific comment or question.

The researcher asked teachers to provide a plan of the classroom and the positioning of the students in order to be able to link each student to an ID and a student survey after observation was performed.

**Episode selection**

A typical lesson comprises of 50 minutes. In each observed lesson an episode of 15 minutes was selected for coding. For the coders it would have been difficult to code longer episodes without losing their concentration towards the end of the episode. The middle 15 minutes of lessons were selected to ensure that the episode would represent whole-class instruction and to ensure valid data of ‘regular’ classroom behaviour. During the first minutes of recording some students are distracted by the cameras and they are for example waving and laughing at the cameras. This behaviour vanishes after a couple of minutes when the lesson has progressed (as if students forget that they are being filmed). Also, the first and last couple of minutes of a 50-
minute lesson are traditionally used to enter and leave the classroom, fill in the school diary, hand out corrected assignments, and take quizzes. The selected episode was watched and rewatched several times since multiple students were coded for each lesson.

**Coding software**

The freeware software package JWatcher was used for coding classroom interactions and behaviours. JWatcher was designed to observe a single subject continuously. Continuous student focal coding was used in this study, in which each sampled student is observed for 15 minutes. Each behavioural category in the codebook is represented by a unique alphanumeric key on the keyboard. JWatcher logs the time at which keys are pressed. JWatcher is based on a “score once, analyse many times” principle. Each behaviour can be analysed both as an event (measuring frequency) and as a state (measuring total duration). Relationships between behaviours can be defined, for example two behaviours can be allowed to overlap and other are defined as mutually exclusive. Uses may also choose to exclude or ignore certain behaviours in analyses. The programme produces simple output files (spreadsheets) that can be readily imported in statistical software programs. We opted to limit the number of codes to approximately 10 since it would be difficult to code more types of behaviours simultaneously and it showed to be difficult to hit the correct alphanumeric key for a larger set of behaviours.

**Codebook**


We used the following categories to indicate the type of instruction:

<table>
<thead>
<tr>
<th>Letter code</th>
<th>Short description</th>
</tr>
</thead>
<tbody>
<tr>
<td>W</td>
<td>Whole-class teaching</td>
</tr>
</tbody>
</table>
Paired or group work

Individual instruction from teacher

We identified the following categories of behaviours to be coded:

<table>
<thead>
<tr>
<th>Letter code</th>
<th>Short description</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>Call out</td>
</tr>
<tr>
<td>H</td>
<td>Hand up</td>
</tr>
<tr>
<td>T</td>
<td>Turn from teacher</td>
</tr>
<tr>
<td>R</td>
<td>Response to teacher</td>
</tr>
<tr>
<td>N</td>
<td>Negative feedback from teacher</td>
</tr>
<tr>
<td>P</td>
<td>Positive and neutral feedback from teacher</td>
</tr>
<tr>
<td>O</td>
<td>On-task behaviour</td>
</tr>
<tr>
<td>F</td>
<td>Off-task behaviour</td>
</tr>
<tr>
<td>S</td>
<td>Social interactions with other students</td>
</tr>
<tr>
<td>M</td>
<td>Missing</td>
</tr>
</tbody>
</table>

We provide a more elaborate definition of the codes in the following paragraphs. Afterward, we explain how the different codes were calculated and analysed.

**Whole-class teaching**
Instruction to the whole class such as lecture, discussion, debate, teacher demonstration, giving directions.

**Paired or group-work**
Any type of episode in which students are given the permission to interact with each other. For example a set of exercises needs to be completed and student are allowed to discuss with each other.

**Individual instruction**
Teacher walks up to the student and provides individual instruction. Instruction is also coded as ‘individual’ when the teacher has approached two students or a small group of students for explaining or discussing something in particular.

**Missing**
The student is not in sight. The student is in sight but it is impossible to determine what the student is doing. For example, the students’ face is completely covered by long hair that has fallen down and the coder cannot know if the student is reading or talking to his or her neighbour.

**Call out**

*Inclusion criteria:*
Student calls out to the teacher, doesn’t matter if the teacher hears the call or not. If the student calls out several times with a break in between, several call-outs are coded. If a student calls out during hand up, it is a call out. If the student calls out to the teacher but also addresses the other students, it is a call out. If the student calls out to the students but also addresses the teacher, it is a call out. For example: “No Frank, the correct answer to that question is Brussels!” If the student asks the teacher a question without asking permission to speak.

*Exclusion criteria:*
Calling out to other students. For example: “Hey, pass on your ruler to me, will you?” If students are talking to themselves. Non-verbal behaviours that call out to catch the attention of the teacher (see off-task behaviour).

**Hand up**

*Inclusion criteria:*
Student raises hand. Several kinds of hand raising are coded: the very prominent arm-waving (“the Windmill”) as well as stretching out the arm timidly and raising one finger in an uncertain way.

*Exclusion criteria:*
If the student calls out during hand raising: this is a call out. When the student stops calling out to the teacher and his or her hand remains raised, hand up is coded again. An arm that remains raised for a very long time does not receive multiple hits. The code H is pressed once at the beginning of the raise and stops when other behaviour is coded as the hand is dropped and the student, for example, returns to being on-task or gets the turn to speak and responds to the teacher.

**Turn from teacher**

*Inclusion criteria:*

Individual invitation or permission to speak or act. Student gets a turn from the teacher to answer a question, to give an answer, to come to the blackboard. Being asked to hand out papers, books or folders. Implicit and non-verbal turns are also included, e.g. name-calling, a nod from the teacher, pointing to the student.

*Exclusion criteria:*
Students who take the turn themselves without permission. For example when a student calls out the answer without being given the turn to speak, this is coded as calling out.

**Response**

*Inclusion criteria:*
Student responds to question or turn from teacher, it doesn’t matter if the answer is correct or wrong, relevant or irrelevant. Each reply to a teachers’ individual invitation to speak or act in a specific way is coded as a response. Non-verbal response is also included. For example when a student shakes his or her head as a way to answer ‘no’ to the teacher. When the student asks a question to the teacher after receiving the permission to speak, this is also coded as response.

*Exclusion criteria:*
Participation in collective responses (these are coded as on-task). We already mentioned that only dyadic interactions are coded.

**Negative feedback**

*Inclusion criteria:*
Student receives negative feedback from the teacher (both classroom management and content). Any behaviour of the teacher directed towards the student indicating disapproval. Common examples are: ‘stop that’, ‘be quiet’, ‘no’, ‘Frank...’, ‘Now is not the time to be doing that’, “I won’t tell you again”. The teacher repeating a student answer in a critical or questioning manner is coded negative feedback. Sometimes a combination of positive and negative feedback was present: “Correct Frank, but ... someone else?” . This would be coded positive and negative. Non-verbal feedback is also coded. For example: the teacher freezes and eyeballs the student. If the teacher directs reprimands to an island or a little group of students of which the
observed student is one. Negative feedback is coded, regardless of students noticing it or not.

**Exclusion criteria:**
Collective reprimands from the teacher that are directed to all students in general. For example: the teacher is writing on the blackboard and says ‘shhhhhht’ without turning towards the students.

**Positive feedback**

**Inclusion criteria:**
Student receives positive feedback from the teacher (both classroom management and content). Any behaviour that indicates praise or satisfaction with the behaviour of the student, including comments such as ‘excellent’, ‘yes’, ‘good’, ‘correct’ and ‘well-done’. The repetition of a student’s answer in a positive, neutral but non-critical tone. Non-verbal feedback is also coded. For example: the teacher winks to a student or directs a ‘thumbs up’ to the student. If the teacher directs praise to a little group of students of which the observed student is one. Positive feedback is coded, regardless of students noticing it or not.

**Exclusion criteria:**
Collective praise from the teacher that is directed to all students in general is not included. For example: the teacher tells students that they all performed very well in the exercise.

**Social behaviour**
Student is interacting with other students. Student is talking to other students (listening or speaking), student is laughing with other students. Student is gesticulating at other student or is watching other student that is gesticulating to him or her. When a student is engaged in some form of communication with other students.

**On-task**
Residual category. Student is doing what is being asked from him or her by the teacher: e.g. paying attention, making exercises, reading a text, listening to teacher, writing instructions down.

**Off-task**
Residual category. Student is off-task. Student is clearly not following the instructions of the teacher: e.g. teacher asks students to take their books and the student is not following this instruction. Student is distracted and for example observing other students that are displaying misbehaviour (without the observed student actually being involved in interaction or communication with the misbehaving students). Some behaviours that are often reprimanded were also coded as off-task. Teachers often only reprimand some students displaying this behaviour and ‘good pupils’ get away with it. Nevertheless, we coded each of the following behaviours a off-task, whether the teacher reprimanded it or not: student balances chair on two legs, student yawns, student is staring outside the window, student lays head down on his/her desk or arm, student ostentatiously sighs.

Analysis

The following behaviours were analysed as frequencies:
- call out
- hand up
- turn from teacher
- response to teacher
- positive feedback
- negative feedback

The following behaviours were analysed as states (total duration as percentage of time in sight):
- on-task
- off-task
- individual instruction

The following behaviour was analysed as a mean duration (mean duration):
- hand up

The following behaviour was analysed as a conditional state (total duration as percentage of time in sight during whole class teaching and thus not including social interactions during episodes in which group or paired work are allowed):
social interactions

Codebook development and training

Two researchers acted as the observers and two weeks of full-time training were performed. Detailed selection criteria were consistently applied to determine if a behaviour could be assigned to a certain category. Several iterative cycles of coding, assessment of reliability and codebook modification were conducted by the two coders until reliability for all codes was high (Hruschka, Schwartz, Cobb St. John, Picone-Decaro, Jenkins, & Carey, 2004). The following steps were used:

1) The coders familiarised with the codes
2) The coders watched and coded 5-minute fragments. Immediately afterwards they discussed the coding and discrepancies in the interpretation of the codebook were discussed. When needed, extra inclusion and exclusion details for the codes were agreed upon.
3) The coders coded 10-minute fragments separately and calculated Krippendorff’s alpha measures of reliability (kalpha). Coders would discuss the discrepancies in coding. Short periods of coding intermixed with periods of discussing until high levels of intercoder reliability were reached for each code (kalpha > .60).
4) Coders coded on an individual basis.
5) An intermediate check of intercoder reliability and a final check were performed.

Reliability

To ensure that high levels of observer agreement would not be mere reflections of high chance agreement, a Krippendorff’s alpha (Krippendorff, 2013) was calculated. As Krippendorff’s alpha is known to be a rather conservative index (Lombard, Snyder-Duch & Bracken, 2002), a more liberal criterion of .60 was used to determine if the coding of a variable was considered reliable. Training continued until kalpha’s were above .60 and from that point on the two coders scored the remaining tapes, being aware that a set of lessons, taken at random, would be scored by the other observer at the middle and at the end of coding. Table 11 gives an overview of the
mean kalpha's for each code for all reliability checks combined. In line with recommendations of Harrop and Swinson (2007) we present a contingency table so that agreements and disagreements on occurrences and non-occurrences of behaviour recorded by the two observers can be inspected.
Academic Track Record
I. Academic and professional experience

Predoctoral Research, Department of Educational Sciences, EDWE, VUB, October 2011 – present (fulltime)

Internship and professional development coordinator, Department of Teacher Education, IDLO, VUB, July - September 2011 (fulltime, ad interim)

Consultant, Kessels & Smit The Learning Company, September 2010 - July 2011 (0.5 fte, intern)

Piano teacher, Mozaïek Bertem vzw, September 2008 – June 2009 (0.1 fte, student work)

II. Education

Master in Educational Sciences (120 ECTS), with highest honor, 2011, VUB

Bachelor in Adult Educational Sciences (180 ECTS), with high honor, 2009, VUB

General Secondary Education (Latin-Mathematics), 2006, Koninklijk Atheneum II Leuven

III. Academic Track Record

1. Education

- contributions in academic programs

Assisting Practica


Guest lectures
“Pedagogische vraagstukken” (3 ECTS), titularis: Nadine Engels, Teacher Education Programme, VUB, academic year 2014-2015: 7/3 (ca. 100 students), academic year 2013-2014: 29/3 (ca. 100 students)

“Introduction to Women Studies”, titularis: Machteld De Metsenaere, VUB, academic year 2012-2013: 22/4 (ca. 30 students)

- support in the supervision of master theses

Note: theses in preparation are italicised.

Laurence TOBBACK, Gendered Teacher-Student Classroom Interactions in STEM and language classes, supervisor Nadine Engels, MA Onderwijskunde, VUB, academic year 2015-2016.

Baokye Sandra FRIMPONG, Gender disparity in education, supervisor Nadine Engels, MA Educational Sciences, VUB, academic year 2015-2016.

Sarah SLEVEN, Pupils from ethnic minorities in Flemish secondary education and inequalities in their relationship with teachers, supervisor Nadine Engels, MA Adult Educational Sciences, VUB, academic year 2014-2015.


Caro CABUS, Influences of teacher expectations on boys’ school achievement, supervisor Nadine Engels, MA Adult Educational Sciences, VUB, academic year 2011-2012

- contributions in assessment of education

Member educational sciences domain expertise of External Quality Assessment Committee 4, Teacher Education Programmes Secondary Education, February 2014 - present
Preparation of Self Assessment Report BA Adult Educational Sciences, VUB, academic year 2014-2015

Preparation of Self Assessment Report MA Educational Sciences, VUB, academic year 2014-2015

Rehearsal Quality Assessment BA en MA Philosophy en Moral Sciences, VUB, as part of the practicum for Curriculum Development in cooperation with the External Quality Assurance Service, academic year 2012-2013

Rehearsal Quality Assessment BA Nursing, HUB, academic year 2012-2013

Rehearsal Quality Assessment MA Disaster Medicine, VUB, as part of the practicum Curriculum Development in samenwerking met Cel Externe Kwaliteitszorg en –bewaking, academiejaar 2011-2012

Proefvisitatie MA Specialistische Geneeskunde, VUB, in kader van WPO Curriculum Development in cooperation with the External Quality Assurance Service, academic year 2011-2012

2. Research

<table>
<thead>
<tr>
<th>Note: Publications are listed chronologically. Manuscripts in preparation, submitted or under review are italicised.</th>
</tr>
</thead>
</table>

- scientific publications, aimed at a wide forum of scholars

B (articles/contributions in scientific monografies / books with an international referee system)


C (articles in scientific journals with international referee system)

Consuegra, E., Engels, N. (under review). Effects of professional development on teachers’ gendered feedback patterns, students’


Consuegra, E., Willegems, V., & Engels, N. (under review). Gendered Emotionality and Rationality in Teachers and it’s Effects on Trust in Students and Need Supportive Teaching. Gender and Education.


I2 (contributions at international conferences/symposia not published or only available as an abstract)

Submitted for sixth international conference Classroom-oriented research: Towards effective learning and teaching, Konin, Poland.


Halimi, M., Consuegra, E., Engels, N., & Struyven, K. (2014, juni 11-13). *Anything boys can do, girls can do better?* Genderideologieën en
educationele verwachtingen van ouders en leerlingen. Presented at annual Onderwijs Research Dagen, Groningen, Nederland.


- contributions in national conferences


- **participation in research networks and collaborations with other research teams within and external to the university**

Elected Junior Assistant Coordinator of the Special Interest Group (SIG) 11 Teaching and Teacher Education, EARLI, European Association for Research on Learning and Instruction, 2013 - present

Member of Procrustes Interuniversity Research Team, IWT Strategisch Basisonderzoek, 2012 - present

Member of the interdisciplinary centre for expertise RHEA, Centre for Gender, Diversity and Intersectionality, VUB, 2012 - present

Member of research team BILD, Brussels Research Centre for Innovation in Learning and Diversity, VUB, 2011 - present

- **referee-work for and membership of editorial and advisory boards of journals, books and series**

Editorial Board member “Journal for Higher Education” [*Tijdschrift voor Hoger Onderwijs*], Garant, 2012 - present

External referee for “Teaching and Teacher Education”, Elsevier, 2014


- **scientific contributions and governance functions at scientific events (meetings, seminars, workshops, conferences, summer schools)**

Member Organising committee biennial EARLI SIG 11 Conference (in cooperation with Tina Seidel, Technische Universität München, School of Education, and Kari Smith, University of Bergen, Department of Education), Frauenchiemsee, Duitsland, academic year 2013-2014
3. Social and scientific valorisation

- Professional and practitioner publications

S (article)


- Popular publications aimed at wider audience


Consuegra, E. (2013). Researching teachers and teaching researchers. [Onderzoekende onderwijzers en onderwijzende onderzoekers.] In J. Van Thienen (Eds.) Girls are such and boys are so? How to deal with gender in school. [Meisjes zus, jongens zo? Hoe omgaan met gender op school?] Leuven: LannooCampus.

- **External membership of governance and advisory boards**

Founding member and secretary of Mysterie van Onderwijs vzw, interdisciplinary platform promoting educational innovation, September 2014 - present

- **Activities in internal and external communication, contributions in sensitizing campaigns, events and media coverage**

Bel 10 “Onderwijs”, Radio1, June 22, 2015.

“Mannekes” Column, Henri, June 2015


“Boys Boys Boys” Opinion article De Standaard, Saturday June 6, 2015


Study and Information days (SID-in), Brabanthal, Leuven: 9/1/15, 20/1/12

Autumn Camp, VUB, Campus Etterbeek, VUB: 29/10/12, 30/10/12.


“Teachers reprimand boys more often than girls” [Leraren berispen jongens vaker dan meisjes]. Klasse for Teachers. February 27, 2014.

“Teachers are disadvantaging boys” [Leerkrachten hebben meer pik op jongens]. Het Laatste Nieuws. February 27, 2014.


“Girls and boys talk equally, but boys are reprimanded” [Meisjes babbelend evenveel, maar de jongens worden gestraft]. Het Nieuwsblad. February 27, 2014

“Girls are still the teachers’ favourite” [Meisjes nog steeds lieveling van Vlaamse leerkrachten]. www.nina.be. February 27, 2014

“Girls are still Flemish teachers’ favourite“ [Meisjes nog steeds lieveling van Vlaamse leerkrachten]. www.demorgen.be. February 27, 2014

“Meisjes babbelend evenveel, maar de jongens worden gestraft”. www.standaard.be. February 27, 2014
“Girls and boys talk equally, but boys are reprimanded” [Meisjes babbel evenveel, maar de jongens worden gestraft]. www.nieuwsblad.be. February 27, 2014

“Girls are still Flemish teachers’ favourite“ [Meisjes nog steeds lieveling van Vlaamse leerkrachten]. www.hln.be February 27, 2014

“Boys are being punished more easily than girls” [Jongens krijgen sneller straf op school dan meisjes]. Televisie-interview voor Het Journaal 1 en Het Journaal 7 op Eén. February 27, 2014

Radio-interview De Ochtend, Radio 1. February 27, 2014

Radio-interview BNR Nieuwsradio. February 27, 2014


4. Contributions in faculty and university governance

Observer Membership as representative of the Governance Board in the Strategic Research Programmes Assessment Committee, 2015 - present

Member of *Committee on Quality Assurance in doctoral program*, Committee of the Board of Assisting Academic staff in the faculty of Psychology and Educational Sciences, 2014 - present

Elected function *Vice-Chairman of the Governance Board of the Vrije Universiteit Brussel*, September 2014 - present
Elected member *Governance College [Bestuurscollege]*, VUB, September 2014 - present

Deputy member of the *Governance Committee [Governance Commissie]* VUB, September 2014 - present

Member of *Budget and Finance Committee [Plannings-en Begrotingscommissie]* VUB, September 2013 - present

Member of *Governance Board [Raad van Bestuur]* VUB, September 2013 - present

Member of Faculty Board of Faculty of Psychology and Educational Sciences [*Faculteitsraad*] PE VUB, September 2013 - present

Member of Governance Committee [*Governance Commissie*] VUB, September 2012 - August 2014

Deputy member of Faculty Board of Faculty of Psychology and Educational Sciences [*Faculteitsraad*] PE VUB, September 2012 - August 2013

Deputy member of *Governance Board [Raad van Bestuur]* VUB, September 2012 - August 2013

Member *Board of Assisting Academic staff* in the faculty of Psychology and Educational Sciences, 2012 - present
Nederlandse samenvatting

Uit verschillende indicatoren blijkt dat jongens op school slechter presteren dan meisjes: vroegtijdig schoolverlaten, zittenblijven, ondervertegenwoordiging in het algemeen secundair onderwijs, minder doorstroom naar het hoger onderwijs. De kloof tussen jongens en meisjes is reeds duidelijk zichtbaar in het begin van het secundair onderwijs en groeit verder naar het einde van het secundair onderwijs toe. Dit onderpresteren betreft niet alleen een *relatief* onderpresteren in vergelijking met meisjes, het betreft ook een *absoluut* onderpresteren in vergelijking met jongens hun eigen potentieel. De OOESE (2015) besluit dat de kloof tussen schoolse prestaties (zoals bv. zittenblijven) veel groter is dan de kloof die vastgesteld wordt tussen jongens en meisjes op de gestandaardiseerde PISA-testen wiskunde, lezen en wetenschappen. Jongens en meisjes die even goed scoren op de PISA-test, zullen niet even goed scoren op school. Jongens blijken op school verhinderd om zich ten volle te ontwikkelen.

Eerder onderzoek heeft verschillende voorspellers geïdentificeerd voor het schools onderpresteren van jongens. Risman & Davis (2012, p. 747) stellen dat genderongelijkheid ge(re)produceerd wordt op drie niveaus: het individuele, het interactionele, en het institutionele. Dit doctoraatsproefschrift richt zich op het interactionele niveau, met een focus op de klasinteracties tussen leerkrachten en hun leerlingen. Wij veronderstellen een model van gegenderde verwachtingseffecten van de leerkracht die gemedieerd worden door hun interacties met de leerlingen in de klas. In 1985 beschreeft Brophy reeds zeer uitgebreid de immense complexiteit van zulke verwachtingseffecten die een wisselwerking zijn tussen de identiteit van de leerkracht en de leerling, interpretaties en percepties van leerkrachten en leerlingen, en het gedrag dat gesteld wordt door leerkrachten en leerlingen.

Zes beperkingen werden vastgesteld in eerder onderzoek over gegenderde klasinteracties. Zes studies werden ontwikkeld om tegemoet te komen aan deze beperkingen in het bestaande onderzoek. In de volgende paragrafen wordt samengevat welke studies er uitgevoerd werden en welke de belangrijkste resultaten zijn.
**Studie 1**

De eerste beperking die we vaststellen is een oververtegenwoordiging van onderzoek uitgevoerd in de Verenigde Staten. Dit is best problematisch aangezien eerder onderzoek aangeeft dat het ‘onderpresteren’ van jongens niet in alle naties als een even groot probleem wordt ervaren (Moureau, 2009). Een systematische analyse van eerder onderzoek werd uitgevoerd en 18 empirische studies werden in detail besproken. We vinden indicaties die kunnen wijzen op culturele verschillen tussen Angelsaksische landen en continentaal Europese landen met betrekking tot het geven van positieve feedback aan leerlingen. Ook worden in continentaal Europa onderzoeksresultaten over de interacties tussen leerlingen en leerkrachten in het hoger secundair onderwijs vanuit theorieën over seksuele spanningen geïnterpreteerd. Deze aandacht voor seksualiteit in klasinteracties is nieuw. Een laatste bevinding is de hypothese dat het verhogen van de positieve feedback van leerkrachten een veelbelovende strategie is voor het verhogen van het opletten tijdens de les van leerlingen.

**Studie 2**

De tweede studie vertrekt vanuit de vaststelling dat veel onderzoekers nog twijfelen aan het bestaan van de zogenaamde teacher gender bias. Sommige onderzoekers stellen dat leerkrachten helemaal geen negatievere verwachtingen hebben voor jongens, maar dat jongens zelf schuld hebben aan hun hoger aandeel negatieve feedback doordat ze meer storend gedrag stellen. Deze studie bestudeerde de impliciete en expliciete gedachten van 13 leerkrachten aan de hand van video-gestimuleerde recall interviews waarbij leerkrachten opnames van hun eigen les herbekijken. Leerkrachten krijgen de opdracht om hun les te herbeleven en te expliciteren wat ze op bepaalde momenten zagen, dachten, deden, voelden. Uit de studie blijkt dat de bewuste gedachten van leerkrachten niet verschillen voor jongens en meisjes, maar hun onbewuste gedachten wel met bijna drie maal zoveel aandacht voor het negatieve gedrag en de negatieve eigenschappen van jongens dan die van meisjes.

**Studie 3**
De derde studie gaat verder op de vraag naar het bestaan van een negatief verwachtingseffect van de leerkrachten. In deze studie wordt het verband bestudeerd tussen percepties van leerlingen over discriminatie in de klas en de eigenlijke geobserveerde klasinteractiepatronen. Dit is een relatie die in eerder onderzoek zelden bestudeerd werd. Om niet te vervallen in een klassieke man-vrouw dichotomie wordt een intersectionele benadering gehanteerd voor de analyses waarbij naar de intersecten (de kruispunten) tussen verschillende assen van diversiteit gekeken wordt. Uit de analyses op de gedragspatronen van 180 leerlingen blijkt dat jongens inderdaad meer negatieve feedback krijgen als gevolg van het feit dat zo ook meer ongewenst gedrag vertonen (roepen, niet opletten). Echter, we vinden ook aanwijzingen voor oneerlijke behandeling van jongens en meisjes door leerkrachten. Meisjes en jongens praten even vaak in de klas, en jongens krijgen hiervoor ook meer negatieve feedback, terwijl dit voor meisjes niet het geval is. Verder valt op dat het hoge SES jongens zijn die het meest off-task zijn en dat het anderstalige meisjes zijn die het minst participeren in klasinteracties zoals hand opsteken, vragen stellen of de beurt krijgen van de leerkracht om te spreken.

**Studie 4**

In de vierde studie proberen we tegemoet te komen aan een kritiek van Acker (1990) dat onderzoek naar gender in beroepen zich vaak beperkt tot het meten van verschillen tussen mannen en vrouwen zonder veel inbedding in feministische of gender literatuur. Eerdere studies vinden geen effecten van het gender van de leerkracht op interactiepatronen met leerlingen of op de prestaties van leerlingen. Dit wil echter niet zeggen dat er geen sprake is van een ‘gegenderd’ beroep. De feminisering van een beroep kan namelijk op meer slaan dan alleen de verschillen tussen mannelijke en vrouwelijke uitvoerders van een job. Wij bestuderen de ‘vervrouwelijking’ van het lerarenberoep door ondervwijstheorieën en feministische theorieën te combineren in één kader. We hypothetiseren dat het effect van gender van de leerkracht op de relatie met leerlingen gemedieerd wordt door enerzijds emotionele aspecten van professionaliteit en anderzijds rationele aspecten van professionaliteit. Analyses van enquêtes bij 1225 leerkrachten tonen aan dat vrouwelijke en mannelijke leerkrachten een gelijkwaardige kwaliteit van relaties met hun leerlingen opbouwen. Echter, ze bereiken deze kwalitatieve
relaties op andere (gestereotypeerde) manieren waarbij vrouwen meer bouwen op hun emotionele competenties en mannen op hun bereidheid tot experiment en reflectie over hun klaspraktijk. Implicaties met betrekking tot de constructie van de genderidentiteit van de leerlingen worden besproken.

**Studie 5**

Een vijfde centrale beperking in eerder onderzoek was dat zelden interventieonderzoek of experimenteel onderzoek uitgevoerd wordt. Studies blijven op het beschrijvende en verklarende niveau, maar dat leert ons niet altijd veel over strategieën om negatieve verwachtingseffecten tegen te gaan. In deze quasi-experimentele studie bij 6 scholen en 30 leerkrachten wordt een eenjarig professionaliseringsstraject uitgerold dat bouwt op *inquiry-based*, samenwerkend en waarderend leren. De effecten worden gemeten op meerdere niveaus zoals door van Veen et al. (2010) aanbevolen: leraargedrag, leerlinggedrag, en leerlingpercepties. Uit de analyses blijkt dat meisjes in de controlegroep die voortgetrokken worden, naar het einde van het schooljaar hun geprivilegeerde positie versterken, terwijl deze groep in de interventiegroep haar privileges lijkt te verliezen. Echter, de geprivilegeerde leerlingen lijken zich tegen de veranderende feedbackpatronen van hun leerkrachten af te zetten. De leerlingen in de interventiegroep vertonen na de interventie gemiddeld genomen meer ongewenst gedrag en ze zijn minder bij de les. Wat ten slotte ook opvalt is dat het gevoel gediscrimineerd te worden onveranderd blijft in alle groepen (ondanks de wel degelijk veranderde interactiepatronen).

**Studie 6**

De laatste studie bestudeert de relatie tussen verschillende eerder aan bod gekomen elementen en de kans op zittenblijven (C-attest) aan het einde van het eerste jaar secundair onderwijs. Eerder onderzoek geeft geen uitsluitend over de verschillende mechanismen die simultaan de gegenderde prestaties van leerlingen beïnvloeden. Wij bestuderen de effecten van sociale achtergrond, eerdere onderwijsprestaties, ongewenst gedrag in de klas, feedbackpatronen van leerkrachten en het gevoel ongelijkwaardig behandeld te worden door de leerkrachten op leerlingen hun slaagkansen. Uit de binomiale logistische regressies blijkt dat gender alleen geen betekenisvolle voorspeller is van de kans op een C-attest. We veronderstellen een mediatie-
effect waarbij het effect van gender gedragen wordt door het gevoel
gediscrimineerd te zijn (want bij toevoeging van deze variabele aan het
model, verliest gender statistische significante). Ook zien we opnieuw een
interactie-effect tussen gender en SES waarbij voor lage SES leerlingen geen
genderschillen bestaan, maar voor hoge SES leerlingen jongens duidelijk een
hoger kans hebben dan meisjes om een C-attest te krijgen. Het verhoogde
risico van hoge SES jongens op zittenblijven kan te maken hebben met hun
hogere aandeel aan ongewenst gedrag (zie studie 4). Leerlingen die veel
ongewenst gedrag vertonen hebben inderdaad een verhoogd risico op
zittenblijven. De feedbackpatronen van leerkrachten (positieve en negatieve
feedback) hebben geen overtuigende voorspellende kracht op de kans om een
C-attest te krijgen.

Afluitend formuleren we op basis van de onderzoeksresultaten een aantal
suggesties voor verder onderzoek en voor professionals en beleidsmakers. Ook staan we kritisch stil bij de beperkingen van ons eigen onderzoek.

Wij erkennen dat we in onze analyses niet hebben kunnen controleren voor
effecten op klas- en schoolniveau (bv. gendercompositie, SES-compositie). De
studie van dergelijke effecten vereist in de eerste plaats een aparte
expertise die in dit onderzoek niet centraal stond maar wel in de onderzoeken
van overige Procrustes-onderzoekers aan bod komt. Ten tweede was het in
veel gevallen wegens de niet-normaal verdeelde data en kleine sample sizes
niet mogelijk om dergelijke analyses uit te voeren. We merken op dat
toekomstig onderzoek uitgebreider stil moet staan bij de aard van
observatiedata en de gevolgen voor mogelijkheden in analysetechnieken.

We stellen vast dat onze interventie niet op alle vlakken een succes was
aangezien sommige leerlingen meer wangedrag vertoonden aan het einde van
het interventietraject. Wij stellen vast dat we veel hypothese hebben om deze
onverwachte effecten te verklaren maar dat we nog onvoldoende verdere
analyses hebben kunnen doen op de verzamelde data om hypotheses uit te
sluiten en hard te maken.

Wij kunnen geen eenvoudige tips & tricks formuleren voor leerkrachten om
om te gaan met jongens en meisjes. Onze resultaten tonen aan dat jongens en
meisjes té diverse groepen zijn om als dichotome categorieën te benaderen. Voor de ene groep goed doen, betekent al snel andere groepen tekortschieten. Wij stellen een aanpak voor waarbij leerkrachten zoveel mogelijk differentiëren op basis van werkelijk waargenomen verschillen tussen leerlingen (en niet op basis van gestereotypeerde verwachtingen). Een gezonde dosis argwaan is aan te bevelen tegenover al te simplistische adviezen over het lesgeven ‘aan jongens zus en aan meisjes zo’.

We sluiten af met een hart onder de riem van alle onderwijsprofessionals en onderwijsonderzoekers. De weg naar (gender)gelijkheid in onderwijs is mogelijk nog lang gezien de diepgeworteldheid van sommige mechanismen die ongelijkheid in stand houden (zoals onbewuste bias). Niet opgeven en meer bruggen tussen praktijk en onderzoek zijn de boodschap. Blijven streven naar een verbetering van de omgeving die ons omringt, is wat ons menselijk maakt. Een onderwijssysteem waarin elke leerling zich maximaal kan ontwikkelen, is een utopie waar we ons met veel overtuiging voor blijven inzetten.