Soc Psychol Educ (2014) 17:357–381
DOI 10.1007/s11218-014-9248-8

Why "Gender" disappeared from the gender gap: (re-)introducing gender identity theory to educational gender gap research

Wendelien Vantieghem · Hans Vermeersch · Mieke Van Houtte

Received: 30 September 2013 / Accepted: 24 January 2014 / Published online: 7 May 2014

Abstract Educational gender gap research tries to explain the differential achievement of boys and girls at secondary school, which manifests in many western countries. Several explanatory frameworks are used for this purpose, such as masculinities theory. In this review article, the history of educational gender gap research in Anglo-Saxon literature and problems with the contemporary approach are discussed. It is argued that gender identity theory could prove valuable both in furthering educational gender gap research and mitigating several problems with masculinities theory. To this end, an overview of the history and recent developments in gender identity theory is given. The scarce research combining educational gender gap research with gender identity theory is reviewed. Possible contributions, assets and research questions from gender identity theory to educational gender gap research are discussed.

Keywords Educational gender gap · Gender identity · Review

1 Introduction

Ever since boys and girls shared a single classroom, their differences in interests and achievement have been the focus of both popular discussion and scientific research in western countries. While the original emphasis was on the low achievement of girls and their underrepresentation in mathematics and science (Byrne 1978; Foster et al.

W. Vantieghem (✉) · H. Vermeersch · M. Van Houtte
Department of Sociology, University of Ghent, Korte Meer 5, 9000 Ghent, Belgium
e-mail: wendelien.vantieghem@ugent.be

H. Vermeersch
e-mail: hans.vermeersch@ugent.be

M. Van Houtte
e-mail: mieke.vanhoutte@ugent.be

2001), the current focus has shifted to the 'boy problem'. This shift occurred in the nineties when studies in western industrialized countries showed that girls had started to outperform boys in several domains, including lower dropout, better test scores, and more enrolment in higher education (Connell 1996; Fergusson and Horwood 1997; Kleinfeld 1999; Younger and Warrington 1996).

Educational scientists have tried to explain the gender gap in education through several theories and frameworks, such as innate traits (Cole et al. 1999; Duckworth and Seligman 2006), a tendency for 'laddish' behaviour (Fergusson and Horwood 1997) and an overall masculinity culture (Connell 1989; Francis 2000; Jackson 2003). Nevertheless, in the past decades too little research has attempted to transcend the gender dichotomy and too many explanations have stayed firmly linked to sex categorizations and distinctions, thus ignoring unifying mechanisms across the sexes and reducing intra-sex differences.

As a new way of approaching the educational gender gap, this review article would like to propose gender identity theory. It is quite clear that the gender gap in education is an important problem. Too many boys become demotivated and uninterested in the course of their school career, thereby throwing away chances of self-development and acquiring diplomas. Only by understanding how this gender gap originates can we develop ways to combat this waste of talent in our societies. We believe that gender identity theory can be an extra pathway to explaining and therefore reducing the gender gap. Next to this community-oriented incentive, we believe that gender identity theory has been isolated from other research areas for too long and that the time has come for gender identity theory to contribute to applied research. After all, there are several strongpoints to this theory. For example, it classifies people according to how 'masculine' or 'feminine' they are, thereby paying attention to both sides of the gender order, while being able to separate this from purely biological sex. It can be used not only as a qualitative method, but a quantitative one as well. Furthermore, gender identity does not have to be utilized solely as a post-hoc clarification, but can have predictive and explanatory power as well. Where masculinity theories tend to explain processes on the social and cultural group level, gender identity theory can be used to discern inherent traits on the individual and personality level.

The goal of this review article is to show that gender identity theory could prove to be a valuable contribution to educational gender gap research. To this end, we will provide an overview of relevant scientific literature to support this argument. We start by giving a historical overview of the educational gender gap and its accompanying scientific literature. We will then discuss the history and recent developments in gender identity theory. Ultimately, we will make a case for linking both research fields by reviewing and unifying the scarce literature that exists up to this point and by indicating possible contributions, assets and research questions.

The reader should note that this review will primarily consider secondary school students, unless stated otherwise. Even though aspects of the educational gender gap are noticeable in higher education (Buchmann et al. 2008; Marrs and Sigler 2012; Van Woensel 2007) and as early as primary education (Buchmann et al. 2008; Derks and Vermeersch 2001; Fergusson and Horwood 1997), most research has focused on gender differences in secondary educational achievement. Furthermore, although

aspects of the educational gender gap manifest throughout the western world (Martínez et al. 2012), we have decided to limit our scope for practical and consistency reasons. So, unless stated otherwise, we will mainly consider Anglo-Saxon literature, with studies based in Canada, Australia, New-Zealand, the UK and the USA. We will also cite several Belgian-based studies. Belgium occupies place 12 on the gender inequality index, whereas the Anglo-Saxon countries occupy places 18 to 42 (UNDP 2011). Belgium is thus more gender egalitarian than the Anglo-Saxon countries and provides an example demonstrating the pervasiveness of the educational gender gap, despite progressive gender beliefs. Since we focus on literature from these countries, the reader should thus always keep in mind that effects may differ in other countries and cultures.

2 Gender gap in education

2.1 Historical overview

For several decades, differences between boys and girls in educational achievement have been the focus of scientific attention. In the seventies, the emphasis was on the low achievement of girls. Girls had lower scores and participation in science and mathematics when compared to boys (Byrne 1978; Foster et al. 2001). Researchers posited several explanations for these findings, which could be divided into two categories. Firstly, girls were positioned within a 'deficit framework' (Anyon 1983; Foster et al. 2001; Hodgetts 2008; Spence et al. 1975). Their poor performances were attributed to a lack of certain qualities, such as low confidence, high anxiety and fear of success. Furthermore, girls' tendency to be compliant in class, work hard and hand in homework that was neat and on time, was taken as a sign of a passive, compliant and malleable learning style (Hodgetts 2008). This was contrasted with boys' active and curious learning style, which would be focused on understanding rather than on achievement or teacher expectations. This way, girls' attitudes and behaviour in school were taken to be a sign of an inferior learning style (Byrne 1978).

Secondly, several researchers pointed to factors other than innate traits that impeded girls' academic success. They suggested studying the impact of societal norms and expectations of gendered behaviour (Byrne 1978; Skelton and Francis 2011). For instance, research pointed out that textbooks lacked positive role models for girls, and that boys dominated the classroom and teacher attention. These researchers declared that it was therefore no surprise that girls would be less confident than boys. Policy recommendations based on this literature focused on changing learning material and raising teachers' awareness of inequalities in their classroom management by letting girls answer questions as often as boys, debunking gender stereotypes concerning mathematics and science, and so on.

However, from the nineties on the attention shifted from girls' low achievement to that of boys. Research revealed that, contrary to expectations, girls were outperforming boys in several domains of schooling, with boys repeating grades more often (Fergusson and Horwood 1997; Van Landeghem et al. 2010), having lower grades (Fergusson and Horwood 1997; Gaer et al. 2006; Duckworth and Seligman 2006; Epstein et al.

1998; Jackson 1998; Younger and Warrington 1996), dropping out more often (Buchmann et al. 2008; Fergusson and Horwood 1997; Van Landeghem et al. 2010), having lower enrolment in higher education (Buchmann et al. 2008; Van Woensel 2007) and being overrepresented in special education services and remedial classes in all OECD-countries (Benjamin 2003). Even in the traditional 'masculine' fields of mathematics and science, girls were gradually catching up to boys, resulting in very small or even insignificant gender differences in these courses (Fergusson and Horwood 1997; Gunderson et al. 2012; Jackson 1998; Younger and Warrington 1996). These revelations instantly turned 'the boy problem' into a hot topic for both policymakers and scientists.

2.2 Explanations for the gender gap

Just as was the case when trying to explain why girls underperformed, researchers have investigated both individual-level and societal factors. When considering the societal factors, two important but interlocking processes appear to be important. Firstly, there is the expansion of education, which has taken place worldwide on all levels of schooling during the second half of the last century (Meyer et al. 1977). Through this expansion of education, women were increasingly able to finish secondary education and enrol in higher education. Hence, women were gradually catching up to boys (Buchmann and DiPrete 2006; Derks and Vermeersch 2001), who historically have had more chances of education (Buchmann and DiPrete 2006; Buchmann et al. 2008; Byrne 1978). Secondly, a decline in gender discrimination during this period (Brooks and Bolzendahl 2004; McHugh and Frieze 1997) reinforced the effects of the educational expansion (Buchmann and DiPrete 2006; Buchmann et al. 2008). These egalitarian processes not only ensured women's increased participation in education, but also opened up more options that were previously considered improper or inappropriate. Women were no longer limited to home economics, but increasingly able to study A-levels or 'masculine' courses such as mathematics or sciences (Byrne 1978). Nowadays in industrialised countries, girls' historical disadvantage in educational participation has been largely resolved (Buchmann and DiPrete 2006; Derks and Vermeersch 2001). Thus, attention has shifted to gender differences *during* the school career (Derks and Vermeersch 2001), which has prompted investigation into grades, drop-out, and school track and individual-level factors that could explain these differences.

When considering individual level-factors, researchers have investigated certain traits, qualities and behaviours of boys that could lead to their low achievement. Since most researchers would agree that intelligence does not vary between the sexes in a manner that is sufficiently consistent to influence academic performance (Duckworth and Seligman 2006; Fergusson and Horwood 1997), research has focused on non-cognitive skills that would explain the gender gap in school performances. For instance, Belgian and British research found that boys had more negative attitudes towards school and were less motivated than girls (Warrington et al. 2000; Houtte 2004; Gaer et al. 2006). Furthermore, Anglo-Saxon research has consistently shown that boys were more inattentive and exhibited more disruptive behaviour in the classroom (Fergusson and Horwood 1997; Francis 2000; Warrington et al. 2000; Younger et al. 1999). They

were also less self-disciplined when it came to homework (Duckworth and Seligman 2006), overestimated their own abilities and tended to attribute their successes to talent and ability, rather than hard work (Cole et al. 1999; Gunderson et al. 2012; Meece et al. 2006). This demotivation, disruptive behaviour, overestimation of own abilities and lack of self-discipline would then all contribute to the gender gap in school performances.

While this research does provide an answer to why boys are underachieving, the focus on traits and behaviour tends to keep the explanation on the descriptive level and does not explain the origins of these behaviours and attitudes. Masculinity theories, however, put the underachievement of boys in a broader context by situating their behaviour and attitudes in a culture of masculinity. The famous starting point hereof is the British ethnography by Willis (1977) that showed how the rebellious and anti-academic attitude of working class boys, which impeded their success at school, fit within a larger working class culture. This work inspired various investigations into the 'lads-culture' by authors like Connell (1989) and Mac an Ghaill (1994). They demonstrated how these boys asserted their masculinity through several practices, one of which was defying the authority and goals of school. Other common practices were 'having a laugh', placing importance on the physical (most prominently on sexuality and sports), acting tough, and displaying a sexist and homophobic attitude (Francis 2000; Jackson 2003; Swain 2005). These practices would impact negatively upon boys' school-related attitudes and behaviour. For instance, 'having a laugh' by being the class-clown diverts attention from the content of the lessons (Francis 2000). It also results in more negative interactions with teachers, who need to intervene to get students back on task (Younger et al. 1999). Furthermore, The emphasis on physicality, such as being a great athlete, is in contrast with necessary behaviour for academic excellence, such as sitting down to read and study (Martino 1999). Moreover, challenging the authority of schools and displaying a tough demeanour impede positive relationships with teachers and school staff.

Since it soon became clear to researchers that not only working class boys enacted 'laddish' behaviour at school, the original focus on working class culture broadened to a more general masculinity culture which influenced all boys (Houtte 2004). In these theories, the concept of hegemonic masculinity takes central stage. Hegemonic masculinity refers to the dominant form of masculinity in a certain context, which is superior in the gender order (Connell 1996). In order for hegemonic masculinity to exist, it must define itself against what it is not, what it is superior to. Therefore, masculinity is constructed within this gender order against subordinated 'others' (Brutsaert 2006; Connell 1996; Epstein 1997; Schippers 2007; Herek 1987). These 'others' include not only femininity, but also marginalized and subordinated masculinities, such as homosexuality, thus creating a hierarchy of masculinities in the process. Often, these masculinities would be conflated with femininity, thereby firmly ensuring their subordinate position (Connell 1996; Schippers 2007).

These theories have proven valuable in understanding the gender gap, since they show how the culture of masculinity has a profound impact on the way boys enact masculinity and 'do gender' in their everyday lives at school. For instance, certain subjects are considered to be 'feminine', such as language, arts or home economics (Connell 1996; Martino 1996). Consequently, boys do not want to be caught paying

attention or even liking these courses for fear of being associated with femininity. Moreover, studying itself is construed as passive and therefore as a devalued and feminine activity (Epstein 1998; Martino 1999). Boys who do work hard at school or cannot compensate good grades with 'appropriate masculine' behaviour, such as excelling in sports or being the class clown, get ridiculed and called names such as 'poofter' or 'fag' (demonstrating how homosexuality is construed as both subordinate and akin to femininity) (Epstein 1998; Martino 1999; Warrington et al. 2000; Stoudt 2006; Swain 2005). The impact of this culture of masculinity on boys' lives should not be underestimated. Various ethnographic studies show how boys actively try to negotiate maintaining a masculine image with getting good grades. Boys who fail to do so get harassed, while others ultimately place more importance on peer acceptance and popularity and thus end up neglecting their studies. One clear illustration of this mechanism is how the classroom behaviour of British high-achieving boys' changes remarkably when they get older. At first, all through primary school, their classroom behaviour resembles most closely that of high-achieving girls: being enthusiastic, providing answers and having positive interactions with the teacher (Jones and Myhill 2004). By age 14, however, this behaviour changes dramatically. High-achieving boys are the least likely to answer questions in class, even less likely than low achieving boys. Researchers attribute this change to the emerging male culture, in which being seen as hard-working or enthusiastic about school is not 'cool'.

2.2.1 Criticisms of masculinity theory

Regardless of how enlightening this hegemonic masculinity theory proved to be, it still received criticism. For instance, several researchers stated that the concept of hegemonic masculinity was too static and ignored the real-world variance in masculinities in classes, regions and cultures (Connell 1996; Connell and Messerschmidt 2005; Smiler 2004). They felt that the theory would benefit from acknowledging several possible masculinities, which would still be in a hierarchical gender order in reference to each other as well as to multiple femininities.

These critiques were swiftly incorporated into the masculinity theory and proved beneficial for research into the gender gap in school as well. Researchers were then able to show how several types of masculinity existed next to each other in schools, such as the 'golden boy', 'macho boy', 'geek' or 'nerd' (Lyng 2009) or the 'real Englishmen', 'macho lads' and 'new enterprisers' (Mac an Ghaill 1994). Acknowledging these different types of masculinity helped to explain why some groups of boys were more or less able to incorporate good grades into their masculine identity, thus recognizing the fact that some boys were in fact high-achievers and that not all boys were so-called "lads" who failed at school. Especially illuminating in these masculinities-typologies is the way class, race and gender intersect. For instance, the working-class "macho lads" from Mac an Ghaill's ethnography (1994) prided themselves on physical dominance. The "real Englishmen" on the other hand, who had a predominantly upper middle-class background, emphasized mental dominance and 'effortless achievement'. These intersectional masculinities are an improvement to research because it acknowledges the fact that some boys, especially those from higher socio-economic background, tend to do well in school (Epstein et al. 1998). This, however, does not imply that

laddish masculinity conceptions and anti-school cultures are a marginal phenomenon, impacting little on the overall achievement of boys and thus not explaining the educational gender gap. Despite variations in masculinities, Francis (2000) has shown that core masculinity-conceptions have changed little over the past decades and that the laddish construction of masculinity continues to be the most accepted form among secondary school pupils.

Nevertheless, masculinity theories remain insufficient for educational gender gap research on several accounts. For instance, there is a regrettable lack of attention for femininities in both research and theory (Connell and Messerschmidt 2005; Schippers 2007), thereby handicapping one half of the educational gender gap research (that is, the research into the achievement of girls). Furthermore, masculinity theory tends to stay firmly linked to biological sex (Francis 2000, 2010). In the framework of the theory, it is impossible for a boy to be a part of feminine culture or for a girl to be a part of masculine culture. In such cases, researchers tend to invent a marginalized masculinity or femininity in order to fit these people into a category which remains within their biological sex. What is more, even though the theory acknowledges multiple masculinities, this does not remedy the fact that differences between boys within each category tend to go unrecognized. As Francis argues on several occasions (e.g. Francis 2000, 2010), the use of typologies risks diminishing gender analysis to different 'sorts' of masculinity or femininity, thus ignoring the fact that people tend to do gender in opposite ways and that all performances of gender are characterized by contradictions and tensions. For example, people may behave differently depending on the context and the persons present (Francis 2000, 2010; Swain 2006). Just imagine a man sitting in the pub with his 'mates', versus that same man alone with his spouse. Categorizing this person as a 'lad' based on his interactions in the pub, would be to ignore the variety of ways he can behave, think and feel on other occasions and the ways in which he differs from other men in the 'lads' category. Moreover, masculinity theory is used as a post-hoc explanation of boys' achievement, whereby the researcher classifies those boys that underachieve into the 'lad'-group, but is unable to predict beforehand where each boy will go. Last but not least, masculinity theory has stayed firmly rooted in qualitative investigations, thereby limiting the findings to each specific time and setting.

An approach that might overcome these limitations is the gender identity theory. This theory and its history will be discussed in the following section.

3 Gender identity theory

Gender identity refers to 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 (Perry and Pauletti 2011; Stets and Burke 2000; Tobin et al. 2010; Wood and Eagly 2009). This concept clearly links to the 'doing gender' theory of sociology. 'Doing gender' refers to the ways people infuse their everyday behaviour and social interactions with gendered symbolic behaviour and signifiers (West and Zimmerman 1987). According to this theory, gender is a master identity. This means that gender cuts across situations and is 'omnirelevant', since any action can be interpreted as exemplifying it. As such,

people can always be held accountable for the gender appropriateness of their behaviour, whether at work, at home or in the street (West and Fenstermaker 1995). The differences between the fields of sociology and social psychology clearly come to play here. The sociological theory of doing gender focuses on interpersonal interaction and symbolic behaviour in the social sphere, whereas the gender identity concept from social psychology starts off at the intrapersonal level as a self-evaluation of masculinity or femininity. However, rather than being at odds with each other, these concepts complement each other and have a reciprocal influence. On the one hand, societal norms concerning ideal masculine and feminine natures may inform people's gender identity (Tobin et al. 2010; West and Zimmerman 1987), through a comparison of own characteristics with those from a gender category (for a more thorough discussion of this identity construction, we refer to Tobin et al. 2010). On the other hand, gender identity can form the rationale for the specific gendered behaviour people display in the social sphere (Tobin et al. 2010; West and Zimmerman 1987) (for a more thorough discussion of this stereotype emulation, we refer to Tobin et al. 2010). Via this way, gender identity influences how people perceive the world around them and how they behave.

Historically, most authors have based their gender identity research around the concepts of masculinity and femininity. Since masculinity and femininity have been dubbed 'one of the muddiest concepts in social sciences' (Constantinople 1973), it will come as no surprise that the last half century of research is characterized by a wavelike motion, where periods of consensus on theory and methodology alternate with moments of widespread discussion and disagreement. At the moment, we are at the crest of a wave, where discussion seems to be making way for a new consensus.

3.1 Unidimensionality before the 1970s

The first ones to develop a masculinity–femininity measure were Terman and Miles (Lippa 2001; Smiler 2004). Their instrument was called "The Attitude Interest Analysis Test" and they assumed that masculinity and femininity were a unidimensional and bipolar construct. This means that masculinity–femininity form a single continuum with masculinity on the one end of the continuum and femininity on the other. They also presumed that people's scores should be in accordance with their biological sex and that extreme scores were ideal (Constantinople 1973; Lippa 2001; Stets and Burke 2000). Therefore, to score 'atypically' was considered to be a sign of mental maladjustment. Their work inspired subsequent authors to develop their own instruments, which also assumed a unidimensional bipolar masculinity–femininity construct. Examples of these are the Guilford-Zimmerman Temperament Survey, the California Psychological Inventory, and the Minnesota Multiphasic Personality Inventory (see Beer 1990).

3.1.1 Critism

While having been popular for quite some time, these instruments got criticized quite heavily. One of the most notable critiques came from Constantinople in 1973. She stated that research supported neither the assumed bipolarity of masculinity and femi-

ninity as opposites, nor the supposed unidimensionality of the construct (Constantinople 1973). Furthermore, the instruments were a mishmash of items that differentiated between the sexes without any theoretical grounding. One could therefore wonder why certain items were chosen in favour of others that distinguished men from women just as well. Finally, several instruments that were supposedly measuring the same construct barely correlated, demonstrating that the construct was badly defined and/or multidimensional. This was combined with a rising feminist criticism, stating that the construct exaggerated the differences between men and women (Lippa 2001; Stets and Burke 2000). Furthermore, it was posited that the instruments were based on cultural stereotypes and were essentially sexist, since the feminine items often carried negative connotations. However, people usually feel the need to conform to these cultural stereotypes and expectations, leading to socially desirable answers. While this issue is hardly limited to gender research, and rather is a pervasive concern for any survey study (Billiet and Waege 2003), it is hard to ascertain to what degree it distorts the answering patterns to gender identity scales. Or as Constantinople (1973) puts it: *"While it is clear…that item content, sex role stereotypy and social desirability interact in measures of M-F, making it difficult to obtain a relatively pure measure, it is not yet clear how (and how much) to control for their effects."* (p. 403).

3.2 Androgyny in the seventies

In response to the earlier-discussed criticisms on gender identity measures, instruments that adopted a two-dimensional approach to masculinity and femininity were developed. The most famous ones are the Bem Sex Role Inventory (BSRI) and Spence's Personal Attributes Questionnaire (PAQ) (Lippa 2001; Smiler 2004; Stets and Burke 2000). While earlier measures usually took (vocational) interests into account, these new instruments focused resolutely on psychological traits. Masculinity was assumed to cover traits such as independence, assertiveness and dominance, while femininity was thought to be represented by sensitivity, kindness and empathy (Bem 1974; Spence et al. 1975). Respondents indicated to what degree these traits described themselves. Since masculinity and femininity were thought to be independent, orthogonal dimensions, scores on both dimensions were combined. People could consequently score high on masculinity and low on femininity; high on femininity and low on masculinity; low on both; or high on both. With this last category, the concept of androgyny was born. In an attempt to resolve the social desirability issues of gender scales, Bem (1974) also included a social desirability scale consisting of gender neutral items. Her research demonstrated that the correlations with this scale were near zero, indicating that the BSRI did not measure a tendency to answer socially desirable.

While extreme scores in accordance to biological sex were considered to be ideal before the seventies, the present approach took androgyny for the summum (Lippa 2001; Smiler 2004; Stets and Burke 2000). Especially Bem (1974) considered androgyny to be the new "standard of psychological health" (p. 162). She believed that the BSRI measured not only masculinity and femininity, but also the degree to which people were gender schematic. This meant that people with extreme scores on either masculinity or femininity thought and acted in a sex-typed way: they would see the

world through a gendered lens and act accordingly. Androgynous people, on the other hand, who possessed both masculine and feminine traits in equal measure, had a broad range of possible behaviours and could therefore adapt their behaviour to what the situation required. According to Bem (1974), this ability was of the utmost importance in "a society where rigid sex-role differentiation had already outlived its utility" (p. 162). Spence (1993; Spence et al. 1975), however, did not support the gender schema theory, and considered the PAQ to be measuring one of the psychological representations of masculinity and femininity, rather than global masculinity and femininity. Nevertheless, with the rise of feminism and the growing criticism on earlier measures, the concept of androgyny rapidly became popular (Lippa 2001) and several researchers jumped on the bandwagon (e.g. Boldizar 1991; Chusmir and Koberg 1988; Orlofsky 1977; Schiedel and Marcia 1985). Consequently, androgyny and the corresponding instruments were responsible for a rise in gender identity research.

3.2.1 Criticism

The BSRI and PAQ remain popular even to this day (e.g. Brown et al. 2006; Harter et al. 1998; Robison-Awana et al. 2002; Smith et al. 1999). Nevertheless, criticism of these measures and the accompanying ideas was formulated quite quickly. Research showed, for instance, that some of the basic assumptions on which the measures were built were faulty. Firstly, masculinity and femininity, as measured by both PAQ and BSRI, failed to relate to gender-related behaviour, attributes or attitudes (Lippa and Connelly 1990; Signorella 1999; Spence 1993). This spoke against gender schema theory, since sex-typed persons would be acting according to either a 'masculine' or 'feminine' behavioural pattern. Secondly, masculinity and femininity were supposed to be independent dimensions, while investigations indicated they actually correlated mildly with each other (Marsh and Myers 1986; Spence et al. 1975). Thirdly, while androgyny was in fact positively associated with wellbeing and self-esteem (Spence et al. 1975), hereby confirming one of Bem's central tenets, researchers also showed that this was explained solely by the presence of high scores on the masculinity scale (Gill et al. 1987; Lippa 2001; Marsh and Myers 1986). Certain items on the masculinity and wellbeing scales were similar and would thus explain the association. Furthermore, researchers began to question whether the masculinity and femininity scales actually consisted of one dimension. Some research revealed that several underlying factors existed (Bernard 1981; Choi and Fuqua 2003). Most studies found one feminine factor, two 'masculine' dimensions (one could be described as an instrumental orientation, the other as self-sufficient), and one bipolar factor consisting of the adjectives 'feminine' and 'masculine' (Choi and Fuqua 2003).

The operationalization of the masculinity and femininity concept was criticized as well. The traits which were said to describe a 'masculine' personality actually depicted an instrumental-oriented personality, whereas the 'feminine' traits depicted an expressive personality (Gill et al. 1987; Lippa 2001; Marsh and Myers 1986). As such, these measures failed to grasp global masculinity–femininity and could be considered to measure solely a sub-aspect. After all, one would be hard put to defend the view that masculinity and femininity entail only these characteristics.

These remarks on the operationalization of the masculinity–femininity construct was further expanded with new feminist criticism. Feminists stated that the traits chosen to represent masculinity and femininity were once again replications of cultural stereotypes, which were sexist in nature (Lippa 2001). After all, many women nowadays would say that they are in fact assertive, independent, logical or individualistic–characteristics all taken by PAQ and BSRI to describe a 'male' personality. This view was backed by later American research, which showed that women tended to score higher on masculinity over the years and that so-called 'masculine' traits had become more normative and desirable for women (Auster and Ohm 2000; Gill et al. 1987; Palan et al. 1999). These findings displayed a societal change in which it had become more acceptable for women to be independent and work. At the same time, it showed that these measures were firmly linked to a cultural frame of reference and thus failed to grasp the abstract reality of masculinity and femininity.

Some researchers concluded that, although the PAQ and BSRI could still be used to measure an instrumental versus expressive personality, they were by no means able to measure the complexity of global masculinity or femininity (Choi and Fuqua 2003; Marsh and Myers 1986; Spence 1993).

3.3 The trough of the wave: crisis in gender identity theory

When doubt was cast on the validity of the BSRI and PAQ, it was clear that gender identity theory needed a new impulse. It was, however, unclear what this new impulse should be. From the eighties well up into the nineties we can speak of a trough in the wavelike motion of gender identity theory. Discussion and disagreement was widespread. Some authors tried to advance theory (Marsh and Myers 1986; Spence 1993), others developed new instruments. Since a lot of research happened without considering previous results by peers, there was a great deal of confusion, overlap and influence between authors, theories and methodology.

These discussions were not limited to social psychologists, since economists theorized about gender identity as well. Akerlof and Kranton (2000, 2002), for instance, adapted the economic utility function to include the identity processes. One of the possible identities they theorized about, next to ethnic identity, is gender. However, their conceptualization of gender identity is quite different from the one used here. They do not consider a continuum from masculine to feminine identity, but rather how central an identity as a man or a woman is for decisions and behaviours.

When considering gender identity as a continuum, Bernard uttered as early as 1981 that a multidimensional approach might be favourable and suggested, by analysing previous measures, as many as six categories: aesthetic interests, manual and physical interests, timidity and sentimentality, temerity, power, and empathy. Lippa and Connelly (1990) also supported multidimensionality, but used a different approach that was based on factual differences between men and women. Initially, they proposed the empiric-based gender diagnosticity approach in which a probability is computed for every person. This probability reflects the degree to which a respondent is male-like or female-like in a certain field compared to a local peer group. After years of research, Lippa (2001) concluded that it would suffice to consider three dimensions:

vocational interests, instrumentality and expressiveness. Wood and Eagly (2009) also proposed a three-dimensional approach to cover the multidimensionality of gender identity, even though their categories were quite different from those of Lippa. They suggested considering traits and interests, whether a person had independent versus collective-oriented relationships, and to what degree a person felt they belonged to the social category of man or woman. Wood and Eagly (2009) also included a short overview of research, including both direct and indirect measures. Indirect measures usually employ implicit attitude tests or priming tasks, which are not subject to conscious control. Interestingly, the majority of the discussed findings seem to accord, regardless of research method. So, even though effect sizes may certainly vary, this overview seems to suggest that the problems with social desirable answers are not as large as to distort general directions and tendencies in the field of gender identity.

Storms (1979), on the other hand, did not develop a multidimensional method. Instead, he developed a new gender identity measure in which people had to answer three face-valid questions about their masculinity and femininity. Storms' data even suggested a return to the original bipolar continuum with masculinity and femininity as opposites. He stated that despite all scientists' discussions, people continued to see masculinity–femininity as diametric extremes.

3.3.1 Questions and criticism

Notwithstanding all these new ideas and research, several researchers started to doubt the actual existence of such a thing as masculinity and femininity other than as cultural fictions (Lippa 2001; Perry and Pauletti 2011; Spence 1984). Their questions included: Is masculinity and femininity just a personal trait? What is masculinity or femininity for that matter? Is it unchanging through time and space or is it firmly linked to a specific expression of culture? And if the latter is so, what is the use of trying to investigate this ever-changing concept?

Next to this very fundamental criticism, more specific ones were uttered as well. One of these is what we call the identity versus identification discussion. Past measures had employed an identity approach, in which researchers use disguised measures to infer someone's masculinity or femininity (Stern et al. 1987; Tobin et al. 2010). This meant that respondents were asked to answer certain questions or indicate to what degree certain items described themselves. The respondents were, however, unaware that what was actually being scrutinized was their masculinity or femininity. The researchers themselves inferred gender identity by comparing the respondents' score to a self-formulated standard or normative score. By using this approach, several respondents might consider the result they ended up with as not reflecting their *felt* gender identity. After all, most people consider several aspects of their personality, behaviour, interests or appearance as influential for their overall felt masculinity or femininity. The catch is that what is most central for felt gender identity tends to differ from person to person (Perry and Pauletti 2011; Spence 1993; Tobin et al. 2010) and would rarely coincide with what the researcher places most importance on (Egan and Perry 2001). For instance, a woman may score high on the instrumentality of the BSRI, but may still feel very feminine because she is interested in fashion and dance. While another might have no interest at all in a so-called 'feminine' appearance, but may consider herself to be

very feminine because she is empathic and relationship-oriented. With this kind of criticism, researchers realized that masculinity and femininity are part of several domains in life, and are not just psychological traits. They therefore doubted that masculinity–femininity could be inferred from any single instrument, whether it measured interests, psychological traits or appearance. They concluded consequently that no existing instrument was able to measure such a broad concept as masculinity–femininity (Egan and Perry 2001; Perry and Pauletti 2011; Spence 1984, 1993; Stern et al. 1987).

This train of thought fitted nicely into the commencing understanding of the multifactorial nature of the masculinity–femininity construct. While such thoughts had been uttered early on by, for instance, Constantinople (1973) and Spence (1984), it had never been a popular notion as it was outshone by the rise of androgyny. However, once this concept and its measures did not prove to be up to par, attention started to shift towards the idea of a multifactorial masculinity–femininity construct.

3.4 New wave or the rise of multifactorial theory

3.4.1 The theory

By the end of the nineties, a new consensus started to form on multifactorial theory. This theory states that not only is masculinity–femininity multidimensional, it is multifactorial as well. This means that masculinity–femininity is evident in several fields of life, such as interests, behaviour, appearance, traits, attitudes, and so forth. However, the scores in each of these fields do not necessarily correlate much with a score in another field. Furthermore, the correlations between these factors are not fixed, but may vary from individual to individual (Spence 1993, 1999; Perry and Pauletti 2011). Spence stated concerning this relation that *"categories of gender-related attributes, beliefs and behaviors typically contribute to separate factors whose relationships to other factors are variable in magnitude, even though often close to zero, and are often complex*" (Spence and Hall 1996, 686). It was therefore not surprising that the PAQ and BSRI failed to have clear associations with other gender-related measures such as gender attitudes (Spence 1993, 1999).

Spence, however, did suggest that there was an overarching masculinity–femininity construct that was the summary of the scores on all of these fields, which could be called gender identity (Spence 1984). This gender identity-score would be calculated according to the importance each individual placed on every one of these gender-related fields for feeling masculine/feminine. How much each of the fields contributes to this overall felt masculinity–femininity would differ for each person (Spence 1984, 1993; Tobin et al. 2010). For instance, one person would place more importance on interests, while another would find personality characteristics more decisive.

This theory explains why people who are so very different in characteristics, behaviour, interests or appearance might all still feel equally masculine or feminine. This system functions as a sort of defence-mechanism that allows people to feel secure in their gender identity, no matter how different they are to the reigning gender-stereotypes. That is, people tend to place importance on their gender-congruent characteristics, while discounting a lack of certain gender-congruent traits or possession of gender-

incongruent characteristics (Perry and Pauletti 2011; Spence 1993; Tobin et al. 2010). What is interesting is that this overarching femininity-masculinity construct would, in fact, be bipolar; thereby once again returning to the earlier models where masculinity and femininity form the endpoints of one continuum. The rationale here is that regardless of the academic discussions about one-dimensionality, two-dimensionality or multidimensionality, in people's heads masculinity and femininity have been, are, and will continue to be each other's opposites. This has been backed by several investigations which show that people, in fact, see overall masculinity and femininity as each other's opposites, that both constructs are defined as what the other is not and that which is one can therefore not be the other (Lippa 2001; Spence 1984, 1993; Spence et al. 1975).

This does not have to be a problem or a step backwards in scientific research, since multifactorial theory does break with earlier constructions of masculinity and femininity. Where earlier models assumed that overall masculinity–femininity could be inferred from scores on one sex-differentiating quality, multifactorial theory clearly separates gender identity from any single domain-score. The overarching masculinity–femininity is people's felt gender identity, a continuum where every person can pinpoint his/her own place on the line as to how masculine or feminine he/she feels him/herself to be. This gender identity is a phenomenological sense of self and, since most persons feel secure in their gender identity, people tend to situate themselves towards the endpoints of the continuum. Clearly separated from this identity concept are the gender-related characteristics and behaviours that people possess. Multifactorial theory states that all of these features contribute to people's overall felt masculinity and femininity through a subconscious calculus, whose idiosyncratic result is gender identity. The specific constellations of gender-related features tend to vary widely from one person to another. This does not pose a problem to a safe sense of gender identity, since most people assign extra importance to those gender-congruent qualities they do possess, while regarding those they do not possess as inconsequential. Acknowledging that every gender-related characteristic contributes in a different way to people's felt gender identity, recognizes that what defines somebody as masculine or feminine is different for each person. Hence, it is impossible for a researcher to infer someone's overall gender identity based on a score in a single field (Spence 1984, 1993; Tobin et al. 2010). This way, the proposed gender identity theory succeeds in explaining two core characteristics of gender-related behaviour that have been puzzling researchers for years: (1) how people within each gender category are vastly different in behaviour, attitudes and interests (Tobin et al. 2010)—so much so that the vast majority of behaviours differs more between people of one gender category than between the sexes (Perry and Pauletti 2011; Stets and Burke 2000). (2) At the same time, despite this huge variance within the sexes, most people feel very confident in their own masculine or feminine identity.

3.4.2 The accompanying methodology

Ever since multifactorial theory has become accepted, research into gender identity is starting to rise. However, research questions cannot be answered as long as there is no valid and reliable instrument that measures gender identity correctly. There have been

a few instruments that are in accordance with the basic principles of multifactorial theory. For instance, the Sexual Identity Scale (SIS) is a multidimensional, bipolar and undisguised measurement of people's self-assessed gender identity (Palan et al. 1999; Stern et al. 1987). Respondents are asked to rate how feminine or masculine they feel on a 5-point scale concerning their looks, interests, behaviour and feelings. This way, the SIS taps the physical, cognitive, societal, and emotional dimensions of gender identity. While the SIS is elegantly simple and provides face-validity, it appears that scores tend to be quite extreme. That is, the instrument correlates very highly with biological sex (Stern et al. 1987) and sex accounts for 87 % of the variance in responses to the SIS (Palan et al. 1999). This means that the SIS cannot provide a surplus value to gender identity research, simply because most people look, act and feel like their own sex. It should be mentioned, however, that these investigations have been with an American-based representative sample. We speculate that the instrument might provide more variance with certain populations, such as with people who identify as transgender.

For research with representative samples, and more specifically with children, the Self-Perception Profile of Egan and Perry has proved to be valuable. According to Egan and Perry, gender identity is a construct that should be measured multidimensionally (Egan and Perry 2001) 2001). They differentiate several aspects of gender identity, such as membership knowledge, gender compatibility, pressure for gender-conforming behaviour and attitudes toward gender groups (Egan and Perry 2001; Tobin et al. 2010). Membership knowledge, more commonly referred to as gender constancy, indicates to what degree children realize they are of a certain sex and that this will not change of its own accord. Since most children reach gender constancy around age six or seven, the researchers agreed that it was not opportune to include this in the measure. Gender compatibility, which refers to a sense of belonging and contentment with one's gender category, is divided into two dimensions: gender typicality and gender contentedness. Gender contentedness originally referred to how happy one is with one's sex. Nowadays, researchers tend to agree that the scale actually measures how happy one is with one's gender-role (Leaper et al. 2012). Secondly, gender typicality refers to the degree to which one feels similar to others in the gender category (Egan and Perry 2001; Tobin et al. 2010). In accordance with multifactorial theory, it is supposed that different children will feel gender typical for different reasons. Pressure for gender-conformity refers to the degree to which one experiences pressure from parents, peers and self to not exhibit cross-gender behaviour. Finally, 'attitudes toward gender groups' refers to the degree of in-group bias.

This instrument has proved to be both reliable and valid with American adolescents (Egan and Perry 2001; Tobin et al. 2010). Consequently, it has become more widely used throughout the last decade.

While most research in gender identity in the past decades has been about validating or refuting the reigning theory of the time, there is at last a dawning consensus on both theory and instrument. This means that the time has come for gender identity to step back into the world of applied research.

4 (Re-)introducing gender identity theory to gender gap research

In the following, we will give a brief overview of research that connects gender identity theory with research into the educational gender gap. One should keep in mind that the first part focuses on the use of PAQ or BSRI, which should not be considered a correct measure of gender identity. We include this, however, for a sense of completeness. Moreover, androgyny-research has been such an important part of gender identity research that it cannot be ignored. Furthermore, this research can provide valuable inspiration for research questions or explanatory mechanisms. In the second part, we give an overview of the recent research that has employed the multifactorial theory and its instruments. While these investigations are still scarce, they seem a promising new avenue for research into the educational gender gap.

4.1 The first investigations using androgyny

When everybody jumped on the androgyny-bandwagon, the measures were applied in several research areas. One of these areas was the explanation of differential academic achievement by boys and girls. Boldizar (1991) found for instance that 'masculinity', as measured by the Children's Sex Role Inventory (CSRI), connected to self-perceived scholastic competence in American 9 to 13-year olds of both sexes and boys' mathematic achievement scores. Femininity, on the other hand, negatively related to girls' mathematic achievement scores. Robison-Awana et al. (2002) found, in a similar vein, that American 10 to 12 year-old girls with high academic competence had high masculinity scores, as high as those of the average boy. By considering attribution style, self-esteem and masculinity–femininity scores, the researchers were able to correctly classify 83.3 % of the above-average achieving girls. This led them to conclude that certain 'masculine' characteristics were beneficial for girls' achievement. They were, however, unable to satisfactorily classify the boys according to these variables. This led them to conclude that what determined girls' academic achievement was not necessarily similar for boys.

While the above-mentioned investigations found that masculinity–femininity related to academic achievement, other authors, however, did not. For instance, Brown et al. (2006) did not find a link between the PAQ and academic achievement among American computer science students. Similarly, Zand and Thomson (2005) found no direct link between gender identity and academic achievement among Afro-American 11 to 14-year olds. Nevertheless, they did find an indirect connection. Gender identity explained self-worth, which was associated with school bonding, which in turn related to school grades. Such an indirect connection was found as well by Eisele et al. (2009) in a similar sample. Gender identity explained differences in acceptance by peers, which related to behavioural conduct, which in turn was associated with school bonding, which once again related to academic achievement. In both investigations, it was the independence and leadership aspects, the so-called 'masculine' orientation of gender identity, which explained both self-worth and acceptance by peers. And in both investigations, once controlled for gender identity, sex itself was no longer relevant.

We can conclude this section by stating that research into academic achievement using the BSRI and PAQ has provided some inconclusive results. While some researchers found direct connections (Boldizar 1991; Robison-Awana et al. 2002), others only found indirect links (Eisele et al. 2009; Zand and Thomson 2005) and still others found no connections at all (Brown et al. 2006). Furthermore, when considering those who do find a connection between gender orientation and academic achievement, it is quite notable that only the 'masculinity' scale provides positive associations. This runs counter to the current findings in gender gap research, where girls are outperforming boys. However, we should keep in mind that the measures used here should not be considered correct interpretations of an overarching gender identity construct. As we have stated before, the PAQ and BSRI only measure instrumental versus expressive traits. It is therefore wrong for researchers to portray their results as the influence of 'masculinity' or 'femininity' on academic achievement. Only a few researchers, such as Zand and Thomson (2005), explicitly state they measure leadership and independence qualities and not gender identity or masculinity. Once we keep this in mind, however, certain puzzling or astounding results are quite easily explained. For instance, it has been demonstrated that instrumental-oriented people have higher self-esteem (Marsh and Myers 1986; Orlofsky 1977; Spence et al. 1975). It follows quite easily that these confident people who feel they are independent, logical and intelligent, have higher self-perceived scholastic competence and tend to score high in mathematics (since this is a school subject in which logical thinking is key).

4.2 Riding the wave: using the self-perception profile

Even though the Self-Perception Profile has started to become more accepted and widespread in the academic world, its use has been limited to mostly psychological adjustment research with adolescents and children. Consequently, research that utilizes the Self-Perception Profile for the explanation of educational achievement is still in its infancy. Therefore, the research overview we can present here is limited.

Research by Leaper et al. (2012), showed that felt pressure for gender-role conforming behaviour was related to American adolescent girls' motivation in English, mathematics and science. Felt pressure for gender-conforming behaviour from parents, for instance, was positively related to motivation for English and negatively related to science and mathematics motivation. Gender typicality and contentedness, on the other hand, were not related to girls' motivation. The researchers explained this unforeseen finding by suggesting that being interested in mathematics or science is in our present day and age no longer incompatible with feeling like a typical girl. Nonetheless, an investigation by Leaper and Van (2008) showed that gender typicality was related with American college men's self-efficacy and interest in traditional versus non-traditional majors. Men who were high in masculinity ideology and gender typicality were most likely to hold traditional interests (such as engineering, physics, economy and mathematics), while men who were low in gender typicality and covert sexism had the highest self-efficacy in non-traditional fields (such as social sciences and linguistics). Consequently, men's scores on covert sexism and gender typicality fully explained the effective choice for traditional versus non-traditional majors. Thus, this research

showed that men who felt to be gender typical or endorsed traditional gender attitudes were less interested in 'feminine' majors, and felt traditional majors to be more compatible with their self-concept. As a consequence, the likelihood that these men took up these non-traditional majors was lower in favour of pursuing traditional majors.

Patterson and Pahlke (2011) investigated American 10 to 13 year-old girls' achievement predictors in a single sex school and took gender typicality into account. It should be noted here that the measure they used was not the measure developed by Egan and Perry. However, for a sense of completeness of the research overview, we include this investigation. They found no connection between gender typicality of these girls and their academic achievement 1 year later. The authors did remark that this might have been due to the fact that gender typicality was measured before the girls entered the single-sex school. The authors hint that a student's felt gender typicality may have changed as a consequence of the single sex environment and that their—lack of—results may be due to this.

After this short review, we can conclude that the application of the new gender identity measures on gender gap research is still minimal. However, the cited investigations here do offer some first valuable insights and, even more so, raise questions by pointing out gaps in knowledge. For instance, it is an interesting finding that while gender typicality is related to college men's interest, self-efficacy and choice for academic fields, it is not related to girls' motivation in English, mathematics or science. We could wonder whether this is an age-effect since Leaper and Van's study (2008) concerned college-aged persons, while Leaper et al. research (2012) concerned 13 to 18-year olds. However, it might be more plausible that this is a gender effect, since the first study concerns men and the second concerns girls. It might be possible that different variables explain girls' versus boys' motivation for gender-appropriate versus gender-atypical study subjects. That girls' motivation was not related to gender typicality while men's was, might be a sign that gender-atypical behaviour for women has become much more accepted than for men in the current Western society. This is a suggestion that has been put forward by several other investigations in different fields (Egan and Perry 2001; Kimmel 2007; Palan et al. 1999; Stern et al. 1987). It is stated that in our western society, with the influence of feminism and emancipation, women have gained the freedom to behave more like men. However, men have not gained the freedom to behave more like women without suffering the ridicule of peers and society. We suggest such statements to be the focus of new research into the relation between gender typicality and academic motivation.

Furthermore, it is notable that investigators cited here? focused their research on either males or females. So far, we are unaware of any investigation that includes both sexes simultaneously in its sample. However, this is necessary in order to make legitimate claims as to differing effects of gender typicality on men or women. Nevertheless, we suggest at the same time to break through the sex dichotomy that has been pervading gender identity research since its beginnings. Too much research has focused on differences between men and women. With the multidimensional measure of Egan and Perry, it would be most interesting to contrast the experiences of gender typical with atypical persons, gender-role content with gender-role discontent people and individuals who experience pressure for gender conformity with those who do not. We propose that, rather than blindly focusing on sex *differences*, there might be

mechanisms that unify people across the sexes according to their sense of typicality or felt pressure. It is therefore important for researchers to pay attention to effects of gender identity, while keeping the possibility of interactions between sex and gender identity in mind.

4.3 How gender identity theory can contribute to gender gap research

To summarize, we believe that gender identity theory could provide a valuable contribution to the already-existing gender gap research in two ways. Firstly, gender identity theory can be a direct factor in the explanation of the gender gap in education. Gender identity could, for instance, explain achievement, motivation and interest differences between boys and girls by considering how gender typicality connects to these outcomes. As became clear in the previously discussed investigations, it could be that high gender-typicality prevents boys, but not girls, from being successful in gender-atypical study fields, thus creating a gender gap in, for instance, language achievement. We could further speculate that high pressure for gender-conformity associates with lower motivation and interest in gender-atypical courses for both boys and girls, while at the same time positively relating to achievement in gender-appropriate studies. We could also wonder whether overall felt 'masculinity' or 'femininity' is connected to a certain study method or classroom behaviour, which could be either characterized by high self-discipline, conscientious and calm demeanour versus a more rebellious, active, or fun-oriented pragmatism.

Secondly, gender identity theory can contribute to gender gap research in an indirect fashion by ameliorating the explanatory power of masculinity theories. We have mentioned before that masculinity theories offer post-hoc, group-based and time and place-specific explanations which remain firmly linked to sex. Gender identity theory's strength, on the other hand, is that it deals with an individual-based characteristic, which can be considered independently from specific peer cultures or biological sex. We suggest that, by connecting the insights of gender identity theory to masculinity theory research, more profound explanations of the impact of masculinity culture on the gender gap in education can be achieved. For instance, research that uses masculinity theories to explain academic achievement tends to divide the school population in several 'study or peer cultures'. However, the explanation of why any person is a member of this or that study culture tends to be characterized by a circular argument. A certain boy is part of a certain group because of his gender ideology, study attitudes, and so forth. But, at the same time, it is assumed that boys acquire these characteristics *because* they are part of these groups. We propose that gender identity can provide a possible explanation for group-membership. For instance, it is plausible that children who feel extremely gender typical or experience strong pressure for gender-conforming behaviour have a higher likelihood of becoming members of 'macho-groups'. Whereas children who feel atypical and do not experience much pressure might feel free to become part of less rebellious, less 'cool' or even pro-school groups. In this way, by connecting the findings of gender identity research with those of masculinity theory, we can better understand the influences on both the individual and the social level. As a result, we might develop theoretical models that are better able

to take into account the complexity and nuances of social realities. Hence, this might enable us to go from explanations that necessarily remain post-hoc and categorical to explanations that can be nuanced and predictive.

5 Conclusion

In this article, we discussed the gender gap in education. Even though this phenomenon occurs in many western countries, we have focused on Anglo-Saxon literature. The reader should thus be careful in extrapolating the presented findings, effects and suggestions beyond this geographical scope. Several scientific explanations for why either girls or boys underachieve have been examined. There is, for instance, attention for traits and behaviour, such as self-discipline or lack thereof, school attitudes, and classroom behaviour. Other theories go beyond these rather descriptive explanations and aim to explain *why* such behaviour occurs. One of the most notable theories herein is the masculinity theory, which discusses how the culture surrounding masculinity and "macho" behaviour influences boys' school behaviour and achievement.

While this theory has been very illuminating, we formulated several criticisms. Firstly, masculinity theory tends to conflate gender with sex, since people cannot be part of a gender culture that contrasts with their biological sex (Francis 2000, 2010). Secondly, femininity remains under-theorized and under-investigated when compared to its counterpart masculinity (Connell and Messerschmidt 2005; Schippers 2007). Thirdly, nuance and contradictions go unrecognized because of masculinity theory's group-based approach. Several people get lumped into a single masculinity-category, which leads to overlooking people's individual behaviour and the contradictions and differences between people in this category (Francis 2000, 2010; Swain 2006). Finally, masculinity theory is almost exclusively used in qualitative investigations, thereby limiting its findings to post-hoc explanations and specific places and time.

We would like to suggest that gender identity theory could provide an answer to most of these criticisms, and hence help to further gender gap research. It is remarkable, however, that gender identity theory has scarcely been used in applied educational research up to this day. We have discussed the history of gender identity theory and have shown how both theoretical and methodological ambiguity has impeded its applied research. Nevertheless, in the recent decade important advances have been made in gender identity theory, providing us with a theoretical framework and an accompanying instrument that has received more and more support in scientific circles.

Hence, we would like to propose that the time has come to re-introduce gender identity theory to gender gap research. While the first investigations combining both fields have been rather scarce, they do provide some promising new pathways and food for thought. For instance, it would be interesting to investigate how gender identity differently relates to boys' and girls' motivation and interest for gender appropriate versus atypical school subjects. Next to these possible sex differences, there might be mechanisms that unify people across the sexes according to their sense of typicality or felt pressure. For instance, gender typical boys and girls may feel and behave similarly at school, while their experiences may contrast with those of gender atypical boys and

girls. In this way, gender identity might be linked to certain classroom behaviours or study methods.

All in all, there is still much to discover regarding the impact of gender identity on school achievement. We posit therefore that this is a promising new area of interest for both gender researchers and educational researchers alike.

Acknowledgments We would like to acknowledge the project "Teaching in the bed of Procrustes", financed by the Institute for Science and Technology (Project Number: SBO 110020), which made this research possible.

References

Akerlof, G. A., & Kranton, R. E. (2000). Economics and identity. *The Quarterly Journal of Economics*, *115*(3), 715–753.

Akerlof, G. A., & Kranton, R. E. (2002). Identity and schooling: Some lessons for the economics of education. *Journal of Economic Literature*, *40*(4), 1167–1201.

Anyon, J. (1983). Intersections of gender and class: Accommodation and resistance by working-class and affluent females to contradictory sex-role ideologies. In S. Walker & L. Barton (Eds.), *Gender, Class and Education* (pp. 19–37). Sussex: The Falmer Press.

Auster, C. J., & Ohm, S. C. (2000). Masculinity and femininity in contemporary American society: A reevaluation using the Bem Sex-Role Inventory. *Sex Roles*, *43*(7), 499–528. doi:10.1023/A:1007119516728.

Beer, C. A. (1990). *Gender roles: A handbook of tests and measures*. New York: Greenwood Publishing Group.

Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology*, *42*(2), 155–162. doi:10.1037/h0036215.

Benjamin, S. (2003). Gender and special educational needs. In C. Skelton & B. Francis (Eds.), *Boys and girls in the primary classroom* (pp. 98–112). Berkshire: Open University Press.

Bernard, L. C. (1981). The multidimensional aspects of masculinity–femininity. *Journal of Personality and Social Psychology*, *41*(4), 797–802. doi:10.1037/0022-3514.41.4.797.

Billiet, J., & Waege, H. (2003). *Een samenleving onderzocht: methoden van sociaal-wetenschappelijk onderzoek*. Antwerpen: De Boeck Hoger.

Boldizar, J. P. (1991). Assessing sex typing and androgyny in children—The Children's Sex-Role Inventory. *Developmental Psychology*, *27*(3), 505–515. doi:10.1037//0012-1649.27.3.505.

Brooks, C., & Bolzendahl, C. (2004). The transformation of US gender role attitudes: Cohort replacement, social-structural change, and ideological learning. *Social Science Research*, *33*(1), 106–133.

Brown, C., Garavalia, L. S., Fritts, M. L. H., & Olson, E. A. (2006). Computer science majors: Sex role orientation, academic achievement, and social cognitive factors. *Career Development Quarterly*, *54*(4), 331–345. doi:10.1002/j.2161-0045.2006.tb00198.x.

Brutsaert, H. (2006). Gender-role identity and perceived peer group acceptance among early adolescents in Belgian mixed and single-sex schools. *Gender and Education*, *18*(6), 635–649. doi:10.1080/09540250600980204.

Buchmann, C., & DiPrete, T. A. (2006). The growing female advantage in college completion: The role of family background and academic achievement. *American Sociological Review*, *71*(4), 515–541.

Buchmann, C., DiPrete, T. A., & McDaniel, A. (2008). Gender inequalities in education. *Annual Review of Sociology*, *34*, 319–337. doi:10.1146/annurev.soc.34.040507.134719.

Byrne, E. M. (1978). *Women and education*. London: Tavistock Publications.

Choi, N., & Fuqua, D. R. (2003). The structure of the Bem Sex Role Inventory: A summary report of 23 validation studies. *Educational and Psychological Measurement*, *63*(5), 872–887. doi:10.1177/0013164403258235.

Chusmir, L. H., & Koberg, C. S. (1988). Gender identity and sex role conflict among working women and men. *Journal of Psychology*, *122*(6), 567–575. doi:10.1080/00223980.1988.9915531.

Cole, D. A., Martin, J. M., Peeke, L. A., Seroczynski, A. D., & Fier, J. (1999). Children's over- and underestimation of academic competence: A longitudinal study of gender differences, depression, and anxiety. *Child Development*, *70*(2), 459–473. doi:10.1111/1467-8624.00033.

Connell, R. W. (1989). Cool guys, swots and wimps—The interplay of masculinity and education. *Oxford Review of Education*, *15*(3), 291–303. doi:10.1080/0305498890150309.
Connell, R. W. (1996). Teaching the boys: New research on masculinity, and gender strategies for schools. *Teachers College Record*, *98*(2), 206–235.
Connell, R. W., & Messerschmidt, J. W. (2005). Hegemonic masculinity—Rethinking the concept. *Gender & Society*, *19*(6), 829–859. doi:10.1177/0891243205278639.
Constantinople, A. (1973). Masculinity–femininity: An exception to a famous dictum? *Psychological Bulletin*, *80*(5), 389–407. doi:10.1037/h0035334.
Derks, A., & Vermeersch, H. (2001). *Gender en schools presteren: Een multilevel-analyse naar de oorzaken van de grotere schoolachterstand van jongens in het Vlaams secundair onderwijs*. Brussels: Ministry of the Flemish Community, Education Department (vol. 26): TOR.
Duckworth, A. L., & Seligman, M. E. P. (2006). Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores. *Journal of Educational Psychology*, *98*(1), 198–208. doi:10.1037/0022-0663.98.1.198.
Egan, S. K., & Perry, D. G. (2001). Gender identity: A multidimensional analysis with implications for psychosocial adjustment. *Developmental Psychology*, *37*(4), 451–463. doi:10.1037/0012-1649.37.4.451.
Eisele, H., Zand, D. H., & Thomson, N. R. (2009). The role of sex, self-perception, and school bonding in predicting academic achievement among middle class African American early adolescents. *Adolescence*, *44*(176), 773–796.
Epstein, D. (1997). Boyz'Own Stories: Masculinities and sexualities in schools. *Gender and Education*, *9*(1), 105–116. doi:10.1080/09540259721484.
Epstein, D. (1998). Real boys don't work: 'underachievement', masculinity and the harassment of'sissies'. In D. Epstein, J. Elwood, V. Hey, & J. Maw (Eds.), *Failing boys? Issues in gender and achievement* (pp. 96–108). Buckingham: Open University Press.
Epstein, D., Elwood, J., Hey, V., & Maw, J. (1998). Schoolboy frictions: feminism and 'failing' boys. In D. Epstein, J. Elwood, V. Hey, & J. Maw (Eds.), *Failing boys? Issues in gender and achievement* (pp. 3–18). Buckingham: Open University Press.
Fergusson, D. M., & Horwood, L. J. (1997). Gender differences in educational achievement in a New Zealand birth cohort. *New Zealand Journal of Educational Studies*, *32*(1), 83–96.
Foster, V., Kimmel, M., & Skelton, C. (2001). What about the boys? An overview of the debates. In W. Martino & B. Meyenn (Eds.), *What about the boys? Issues of masculinity in schools* (pp. 1–23). Buckingham: Open University Press.
Francis, B. (2000). *Boys, girls and achievement: Addressing the classroom issues*. London: Routledge-Falmer.
Francis, B. (2010). Re/theorising gender: Female masculinity and male femininity in the classroom? *Gender and Education*, *22*(5), 477–490. doi:10.1080/09540250903341146.
Gill, S., Stockard, J., Johnson, M., & Williams, S. (1987). Measuring gender differences—The expressive dimension and critique of androgyny scales. *Sex Roles*, *17*(7–8), 375–400. doi:10.1007/bf00288142.
Gunderson, E. A., Ramirez, G., Levine, S. C., & Beilock, S. L. (2012). The role of parents and teachers in the development of gender-related math attitudes. *Sex Roles*, *66*(3–4), 153–166. doi:10.1007/s11199-011-9996-2.
Harter, S., Waters, P. L., Whitesell, N. R., & Kastelic, D. (1998). Level of voice among female and male high school students: Relational context, support, and gender orientation. *Developmental Psychology*, *34*(5), 892–901. doi:10.1037//0012-1649.34.5.892.
Herek, G. M. (1987). On heterosexual masculinity: Some psychical consequences of the social construction of gender and sexuality. In M. Kimmel (Ed.), *Changing men: New directions in research on men and masculinity* (pp. 68–82). London: Sage.
Hodgetts, K. (2008). Underperformance or 'getting it right'? Constructions of gender and achievement in the Australian inquiry into boys' education. *British Journal of Sociology of Education*, *29*(5), 465–477. doi:10.1080/01425690802326887.
Jackson, C. (2003). Motives for 'laddishness' at school: Fear of failure and fear of the 'feminine'. *British Educational Research Journal*, *29*(4), 583–598. doi:10.1080/0141192032000099388.
Jackson, D. (1998). Breaking out of the binary trap: Boys' underachievement, schooling and gender relations. In D. Epstein, J. Elwood, V. Hey, & J. Maw (Eds.), *Failing boys? Issues in gender and achievement* (pp. 77–95). Buckingham: Open University Press.

Jones, S., & Myhill, D. (2004). 'Troublesome boys' and 'compliant girls': Gender identity and perceptions of achievement and underachievement. *British Journal of Sociology of Education*, *25*(5), 547–561. doi:10.1080/0142569042000252044.

Kimmel, M. (2007). Masculinity as homophobia: Fear, shame and silence in the construction of gender identity. In N. Cook (Ed.), *Gender relations in global perspective: Essential readings* (pp. 73–83). Toronto: Canadian Scholar's Press.

Kleinfeld, J. (1999). Student performance: Males versus females. *Public Interest*, *134*, 3–20.

Leaper, C., Farkas, T., & Brown, C. S. (2012). Adolescent girls' experiences and gender-related beliefs in relation to their motivation in Math/Science and English. *Journal of Youth and Adolescence*, *41*(3), 268–282. doi:10.1007/s10964-011-9693-z.

Leaper, C., & Van, S. R. (2008). Masculinity ideology, covert sexism, and perceived gender typicality in relation to young men's academic motivation and choices in college. *Psychology of Men & Masculinity*, *9*(3), 139–153. doi:10.1037/1524-9220.9.3.139.

Lippa, R., & Connelly, S. (1990). Gender diagnosticity—A new Bayesian-approach to gender-related individual-differences. *Journal of Personality and Social Psychology*, *59*(5), 1051–1065. doi:10.1037//0022-3514.59.5.1051.

Lippa, R. A. (2001). On deconstructing and reconstructing masculinity–femininity. *Journal of Research in Personality*, *35*(2), 168–207. doi:10.1006/jrpe. 2000.2307.

Lyng, S. T. (2009). Is there more to "Antischoolishness" than masculinity? On multiple student styles, gender, and educational self-exclusion in secondary school. *Men and Masculinities*, *11*(4), 462–487. doi:10.1177/1097184x06298780.

Mac an Ghaill, M. (1994). *The making of men: Masculinities, sexualities and schooling*. Buckingham: Open University Press.

Marrs, H., & Sigler, E. A. (2012). Male academic performance in college: The possible role of study strategies. *Psychology of Men & Masculinity*, *13*(2), 227. doi:10.1037/a0022247.

Marsh, H. W., & Myers, M. (1986). Masculinity, femininity and androgyny: A methodological and theoretical critique. *Sex Roles*, *14*(7–8), 397–430. doi:10.1007/bf00288424.

Martínez, A., Julia A., Mari-Klose M. & Mari-Klose, P. (2012). *Looking inside the gender gap in school achievement: A comparative analysis of its individual, family and contextual causes*. Paper presented at the ESA Midterm, Ghent, 14/09/2012.

Martino, W. (1996). Gendered learning practices: Exploring the costs of hegemonic masculinity for girls and boys in schools. In E. Ministerial Council on Education, Training and Youth Affairs (Ed.), *Gender equity: A framework for action in Australia Schools*, Canberra, pp. 124–144.

Martino, W. (1999). 'Cool boys', 'party animals', 'squids' and 'poofters': Interrogating the dynamics and politics of adolescent masculinities in school. *British Journal of Sociology of Education*, *20*(2), 239–263. doi:10.1080/01425699995434.

McHugh, M. C., & Frieze, I. H. (1997). The measurement of gender-role attitudes: A review and commentary. *Psychology of Women Quarterly*, *21*(1), 1–16.

Meece, J. L., Glienke, B. B., & Burg, S. (2006). Gender and motivation. *Journal of School Psychology*, *44*(5), 351–373. doi:10.1016/j.jsp.2006.04.004.

Meyer, J. W., Ramirez, F. O., Rubinson, R., & Boli-Bennett, J. (1977). The world educational revolution, 1950–1970. *Sociology of Education*, 242–258.

Orlofsky, J. L. (1977). Sex role orientation, identity formation, and self-esteem in college men and women. *Sex Roles*, *3*(6), 561–575. doi:10.1007/bf00287839.

Palan, K. M., Areni, C. S., & Kiecker, P. (1999). Reexamining masculinity, femininity, and gender identity scales. *Marketing Letters*, *10*(4), 357–371. doi:10.1023/A:1008110204546.

Patterson, M. M., & Pahlke, E. (2011). Student characteristics associated with girls' success in a single-sex school. *Sex Roles*, *65*(9–10), 737–750. doi:10.1007/s11199-010-9904-1.

Perry, D. G., & Pauletti, R. E. (2011). Gender and adolescent development. *Journal of Research on Adolescence*, *21*(1), 61–74. doi:10.1111/j.1532-7795.2010.00715.x.

Robison-Awana, P., Kehle, T. J., Bray, M. A., Jenson, W. R., Clark, E., & Lawless, K. A. (2002). Self-esteem, gender-role perception, gender-role orientation and attributional style as a function of academic competence: Smart girls are different, but a boy is a boy is a boy. *Canadian Journal of School Psychology*, *17*(1), 47–64. doi:10.1177/082957350201700105.

Schiedel, D. G., & Marcia, J. E. (1985). Ego identity, intimacy, sex role orientation, and gender. *Developmental Psychology*, *21*(1), 149–160. doi:10.1037/0012-1649.21.1.149.

Schippers, M. (2007). Recovering the feminine other: Masculinity, femininity, and gender hegemony. *Theory and Society*, *36*(1), 85–102. doi:10.1007/s11186-007-9022-4.
Signorella, M. L. (1999). Multidimensionality of gender schemas: Implications for the development of gender-related characteristics. In W. Swann, L. Langlois, & L. Gilbert (Eds.), *Sexism and stereotypes in modern society. The gender science of Janet Taylor Spence* (pp. 107–125). Washington, DC: American Psychological Association.
Skelton, C., & Francis, B. (2011). Successful boys and literacy: Are "Literate Boys" challenging or repackaging hegemonic masculinity? *Curriculum Inquiry*, *41*(4), 456–479. doi:10.1111/j.1467-873X.2011.00559.x.
Smiler, A. P. (2004). Thirty years after the discovery of gender: Psychological concepts and measures of masculinity. *Sex Roles*, *50*(1–2), 15–26. doi:10.1023/B:SERS.0000011069.02279.4c.
Smith, C. J., Noll, J. A., & Bryant, J. B. (1999). The effect of social context on gender self-concept. *Sex Roles*, *40*(5–6), 499–512. doi:10.1023/a:1018879811991.
Spence, J. T. (1984). Gender identity and its implications for the concepts of masculinity and femininity. In T. Sonderegger (Ed.), *Nebraska symposium on motivation*, vol. 32, pp. 59–95.
Spence, J. T. (1993). Gender-related traits and gender ideology: Evidence for a multifactorial theory. *Journal of Personality and Social Psychology*, *64*(4), 624–635. doi:10.1037/0022-3514.64.4.624.
Spence, J. T. (1999). Thirty years of gender research: A personal chronicle. In W. Swann, L. Langlois, & L. Gilbert (Eds.), *Sexism and stereotypes in modern society. The gender science of Janet Taylor Spence* (pp. 255–289). Washington, DC: American Psychological Association.
Spence, J. T., & Hall, S. K. (1996). Children's gender-related self-perceptions, activity preferences, and occupational stereotypes: A test of three models of gender constructs. *Sex Roles*, *35*(11–12), 659–691. doi:10.1007/bf01544086.
Spence, J. T., Helmreich, R., & Stapp, J. (1975). Ratings of self and peers on sex role attributes and their relation to self-esteem and conceptions of masculinity and femininity. *Journal of Personality and Social Psychology*, *32*(1), 29. doi:10.1037/h0076857.
Stern, B. B., Barak, B., & Gould, S. J. (1987). Sexual identity scale—A new self-assessment measure. *Sex Roles*, *17*(9–10), 503–519. doi:10.1007/bf00287732.
Stets, J. E., & Burke, P. J. (2000). Femininity/masculinity. In E. Borgatta & R. Montgomery (Eds.), *Encyclopedia of Sociology* (Revised ed., pp. 997–1005). New York: MacMillan.
Storms, M. D. (1979). Sex-role identity and its relationships to sex-role attributes and sex-role stereotypes. *Journal of Personality and Social Psychology*, *37*(10), 1779–1789. doi:10.1037//0022-3514.37.10.1779.
Stoudt, B. G. (2006). "You're Either In or You're Out" school violence, peer discipline, and the (re)production of hegemonic masculinity. *Men and Masculinities*, *8*(3), 273–287.
Swain, J. (2005). Masculinities in education. In M. Kimmel, J. Hearn, & R. W. Connell (Eds.), *Handbook of studies on men and masculinities* (pp. 213–229). London: Sage.
Swain, J. (2006). Reflections on patterns of masculinity in school settings. *Men and Masculinities*, *8*(3), 331–349.
Tobin, D. D., Menon, M., Menon, M., Spatta, B. C., Hodges, E. V. E., & Perry, D. G. (2010). The intrapsychics of gender: A model of self-socialization. *Psychological Review*, *117*(2), 601–622. doi:10.1037/a0018936.
UNDP. (2011). Gender inequality index and related indicators. In Human Development (Ed.), *Report* (pp. 139–142). New York: Palgrave Macmillan.
Van de Gaer, E., Pustjens, H., Van Damme, J., & De Munter, A. (2006). The gender gap in language achievement: The role of school-related attitudes of class groups. *Sex Roles*, *55*(5–6), 397–408. doi:10.1007/s11199-006-9092-1.
Van Houtte, M. (2004). Why boys achieve less at school than girls: The difference between boys' and girls' academic culture. *Educational Studies*, *30*(2), 159–173.
Van Landeghem, G., Goos, M., & Van Damme, J. (2010). Vroege schoolverlaters in Vlaanderen. Evolutie van de ongekwalificeerde uitstroom tot 2007. Steunpunt SSL-rapport SSL/OD1/2009.25.
Van Woensel, A. (2007). *Genderjaarboek*. Leuven: Departement Werk en Sociale Economie.
Warrington, M., Younger, M., & Williams, J. (2000). Student attitudes, image and the gender gap. *British Educational Research Journal*, *26*(3), 393–407. doi:10.1080/01411920050030914.
West, C., & Fenstermaker, S. (1995). Doing difference. *Gender & Society*, *9*(1), 8–37.
West, C., & Zimmerman, D. H. (1987). Doing gender. *Gender & Society*, *1*(2), 125–151.
Willis, P. E. (1977). *Learning to labor: How working class kids get working class jobs*. New York: Columbia University Press.

Wood, W., & Eagly, A. H. (2009). Gender identity. In R. H. M. Leary (Ed.), *Handbook of individual differences in social behavior* (pp. 109–125). New York: Guilford Press.

Younger, M., & Warrington, M. (1996). Differential achievement of girls and boys at GCSE: Some observations from the perspective of one school. *British Journal of Sociology of Education*, *17*(3), 299–313. doi:10.1080/0142569960170304.

Younger, M., Warrington, M., & Williams, J. (1999). The gender gap and classroom interactions: Reality and rhetoric. *British Journal of Sociology of Education*, *20*(3), 325–341. doi:10.1080/01425699995290.

Zand, D. H., & Thomson, N. R. (2005). Academic achievement among African American adolescents: Direct and indirect effects of demographic, individual, and contextual variables. *Journal of Black Psychology*, *31*(4), 352–368. doi:10.1177/0095798405278198.

Wendelien Vantieghem is currently working as a Ph.D. student at the department of Sociology, Ghent University. Her research focuses on gender-differential achievement in education, gender identity, pressure for gender conformity and well-being.

Hans Vermeersch is a doctor in Sociology, connected to the department of Sociology at Ghent University, and occupies a teaching position at KATHO. His research focuses on the interaction between social influences and biology on risk-taking among adolescents.

Mieke Van Houtte, PhD., is a senior lecturer in the Department of Sociology at Ghent University (Belgium), and head of the research team CuDOS. Her research interests cover diverse topics within the sociology of education, particularly the effects of structural and compositional school features on several outcomes for students and teachers.