Abstract

Several European studies indicate that boys and girls possess different academic stereotypes. More specifically, boys possess negative academic stereotypes, and girls possess positive academic stereotypes. A growing body of research suggest that groups possessing negative stereotypes show decreased performance on various tasks as a result of stereotype threat. As group membership and its associated stereotype becomes salient for the target individual, several automatic cognitive and affective processes are initiated, these processes ”steals” cognitive resources from the task at hand and results in poorer performance. The studies regarding academic stereotypes have all relied on direct measures (i.e., self-report) to indicate stereotypical beliefs. Self-report have on a general level proved to be vulnerable for invalidating issues like limited introspective access and response bias. As stereotyping and their activation are automatic processes, the current project have acted to complement the research field by developing an indirect audio-visual measurement procedure, designed to capture children´s *implicit* gender based academic stereotypes. The measurement procedure is based on the Implicit Association Test; inferring implicit cognitive associations between social constructs based on response latency. The procedure was implemented on a sample of young adults (*N* = 30, 15 females). The results indicate that there indeed exists an implicit pro-academic stereotype favouring females. The project concludes that the developed measurement procedure is valid and reliable enough to be used in further research.

Acknowledgements and Dedication

First, I would like to express my sincere gratitude to my supervisor, Professor Ute Gabriel. Your guidance, contributions and feedback have helped me immensely.

I dedicate this thesis to my family. Your support have been essential in innumerable ways, thank you.

Trondheim, april 2015.

Stian Tangen Hansen

Table of Contents

Abbreviations

IAT	–	Implicit association test
MeCIAS	–	Measure of children´s implicit academic stereotypes
IQ	–	Intelligence quotient
FSIQ	–	Full scale intelligence quotient
CNS	–	Central nervous system
TNGS	–	Theory of neuronal group selection
CAM	–	Cell adhesion molecules
SAM	–	Substrate adhesion molecules
Ch-IAT	–	Child implicit association test

Figures

Tables

Appendix

1. Introduction

Our educational system is fundamental and a cornerstone in our culture. In Norway this institutional process starts when children are 5 or 6 years old, and is compulsory for 10 consecutive years until the child has completed both primary and secondary school. There are well established systems that provides equal educational opportunities for all Norwegian children, independent of parents income (no tuition fees are required in public schools), religious background, or physical disabilities. Nevertheless, data show a gender gap in academic achievement: girls are consistently reported to score higher than boys on all subjects scoring an average of 4 points[1] more than boys (Bakken, Borg, Hegna, & Backe-Hansen, 2008; Statistisk Sentralbyrå, 2014), the only exception being physical education (Tuhus, 2013). Furthermore, boys are over represented in special education needs; 70% of the special education cases are boys (Tuhus, 2013). Given Norway´s fundamental interest in equality and fairness which is reflected in Norway being referred to as one of the most gender equal societies in the world (United Nations Development Programme, 2013), the academic achievement gap between boys and girls is a topic deserving undivided attention. The academic achievement gap between the boys and girls may possibly be explained by the fact that boys exhibit behaviour, or is expected to exhibit kinds of behaviour - which is not congruent with the current school regime. Whereas expectations of girls and their behaviour is more consistent with the preferred behaviours. Characteristics (e.g., expectations of behaviour) given solely on the basis of group membership (in this case, gender) are called stereotypes.

Hartley and Sutton´s (2013) showed in their study that stereotypes regarding girl`s superior academic performance arise as early as 4 and 7 years of age in their sample of girls and boys, respectively. The procedure was as follows: a female experimenter told the child a story depicting a fictitious agent practicing good or poor academic behaviour, from which the child would have to determine the gender of the agent (see Hartley & Sutton, 2010 supplementary document for items). The child then reported their answer through pointing to one of two silhouette pictures of a girl or a

[1] School points are calculated by the average of all the marks. This average, with two decimals, are then multiplied by 10.

boy (i.e., self-report), and a definite answer was required for being included in the analysis (i.e., the children who answered "both", or "none" were excluded). However, the fact that the children had to choose between one or the other, and only used a female experimenter to present the stories, may well have shifted the response from the participants through several confounding effects (e.g., limited response opportunities; limited introspective access; expectancy effects; social desirability; compliance). Furthermore, self-report measures are more likely to predict behaviours that are under people's conscious control, as opposed to more spontaneous, uncontrollable (i.e., automatic) behaviours (Asendorpf, Banse, & Mücke, 2002).

These stereotypical beliefs also have the potential to influence intellectual performance for the group possessing the less fortunate (i.e., negative) stereotype (Ambady, Shih, Kim, & Pittinsky, 2001; Hartley & Sutton, 2013; Schmader, 2002), through a mechanism coined stereotype threat (Steele, 1997). These stereotypes are threatening because performance on tasks that demand deliberate mental activity decrease as a result of the individual using cognitive resources on task-irrelevant operations, namely suppressing negative mental (affective and cognitive) behaviour associated with group affiliation (Schmader, Johns, & Forbes, 2008).

Addressing the gap in research: Stereotypes arise from automatic non-intentional mental processes (Bargh, Chen, & Burrows, 1996; Chen & Bargh, 1997; Devine, 1989). These automatic processes comprise of the spontaneous activation of associations or responses that have been repeatedly reinforced through ontogenetical experience (Edelman & Tononi, 2000). The strength of these implicit associations (both relative and absolute) have been made possible to measure through e.g., the implicit association test (Greenwald, McGhee, & Schwartz, 1998), and the go no-go association test (Nosek & Banaji, 2001), respectively. An implicit measure is the outcome of a measurement procedure that is causally produced by psychological attributes in an automatic manner (De Houwer, Teige-Mocigemba, Spruyt, & Moors, 2009a).

Considering the literature acknowledging stereotyping and its activation as automatic/implicit processes (Bargh et al., 1996; Chen & Bargh, 1997; Devine, 1989), the early time for onset for academic stereotypes (Hartley & Sutton, 2013), and the negative effects of such stereotypes (Ambady et al., 2001; Schmader & Johns, 2003), a definite gap in the research unravels. With these notions readily available, the idea of measuring implicit academic stereotypes indirectly is ought to transpire. And it comes most naturally to measure the result of these processes (i.e., associations) through procedures derived from the field of implicit social cognition. First and foremost to complement the already existing measurement procedures which depends on self-report, and to acknowledge and utilize the aspired technology in the measurement of [implicit] mental constructs.

This thesis carries two principal aims: The first is to develop a measurement procedure capable of detecting implicit stereotypes in children. More specifically, it will aim for the successful development of a nomothetic measurement procedure which successfully manages to capture implicit mental associations between psychological constructs (gender/sex and characteristics). This measure is abbreviated MeCIAS in this thesis (**Me**asure of **C**hildren´s **I**mplicit **A**cademic **S**tereotypes). For this, all materials and stimuli needed to be collected, coded, selected and implemented in a test procedure.

The initial second aim was to utilize this developed measure cross-sectionally on elementary school children, first throughout seventh grade. However, this was not possible due to sudden and unexpected withdrawal of the target sample. Instead, the measurement procedure will be used on a sample of university students mainly for checking test validity and reliability, and to test whether the gendered academic stereotypes exists in a this sample.

First, the background and motive for this research topic will be presented, firstly by exploring earlier research on gender differences in academic behaviour and achievement. Then, bringing up more specifically what academic behaviour is, and if there is pointed out any gender differences in their propensities in the scientific literature. Next up, a chapter focusing on stereotypes and their function. Including the potential impact stereotypes has on performance for individuals possessing them and how this is related to academic achievement. Before entering the domain of implicit social cognition and an explanation the global theoretical framework from which we intend to base our assumptions for measurement. Then, the methodological aspects of psychometrics will be described, including the distinction between direct and indirect measurement procedures and their limitations. Following with a deeper look into the implicit association test. Then, presenting the present research, the development of MeCIAS, including selection of materials, and production of the measurement procedure and piloting. Following with the method section, providing information about the participants, materials and procedure. Then, presenting the results, through data preparation, and hypothesis testing. Ending in a discussion and conclusion.

2. Gender Differences in Academic Behaviour and Achievement

In a longitudinal study of boys´ development through elementary school Nielsen (2009) followed a Norwegian school class from the first- throughout ninth grade, with a follow up after high school. The focus of the study was to observe the development of gender in an academic setting, and how the middle-class masculinity role has been influenced by the ”Nordic gender equality regime”. Nielsen (2009) found indications of boys´ academic inferiority to girls in a number of areas, namely cognitive, social, and inhibitory competence. As a group, girls are better at self-regulation and better socially adjusted than boys as a group (Bakken et al., 2008). Björklund and Kipp (1996) supports the claim that girls have superior inhibitory competence, explaining this difference mainly through differences in selection pressure on men and women throughout our species phylogenetical development. Following Nielsens (2009) study, observations from first grade when school started - indicated that the girls were more prepared for the academic demands than boys were. Several of the girls knew the letters and even how to read when starting at school, none of the boys in the study could. The boys also showed a lesser developed social awareness – through a demonstration where the teacher proposed a situation that depends on a somewhat developed theory of mind (i.e., see themselves from another person´s perspective), all the girls passed this test. The boys also showed more difficulty following instructions, and generally acted more disinhibited than girls during situations requiring obedience (Nielsen, 2009). From this study we can conclude that there are certain behaviours – such as social awareness, self-regulation and social adjustment - that´s more acknowledged in an academic setting than other behaviours, and that the girls in this sample demonstrate these [academic] behaviours more frequent than boys as a group. This difference might explain the academic achievement gap, but what exactly is academic behaviour and how it is linked to academic achievement?

Several other studies have linked certain behaviours to academic achievement, ranging from intellectual capabilities, such as reading and mathematics (Latsch & Hannover, 2014), academic involvement (Rispoli et al., 2011), to social behaviours such as prosocial behaviour (Green, Forehand, Beck, & Vosk, 1980; Wentzel, 1993), and compliance (Cobb, 1972).

The next section will explain concepts that are shown to be related to academic behaviour: the concept of prosocial and antiscocial behaviour (section 2.1), obedience and disobedience (2.2), and finally, intellect (2.3) in more detail and whether the literature have shown any gender differences in their propensities.

2.1 Prosocial and Antisocial Behaviour

Prosocial behaviour is defined as any behaviour that is positive and calculated to promote the interest and well being of of society (Colman, 2009, p. 610).

Men and women are both equally prosocial, but differ in their expression of prosocial behaviour (Eagly, 2009): men show more agentic, strength intensive, and collectively oriented prosocial behaviour (e.g., helping others, taking initiative), whereas women show more communal and relational prosocial behaviour (e.g., friendly, unselfish, empathic and emotionally expressive).

Antisocial behaviour is characterized by a behavioural pattern of disregard for and violation of the rights of others, failure to conform to social norms, unlawful behaviour, deceitfulness, lying and impulsivity (Colman, 2009, p. 46).

Antisocial behaviour includes two subcategories (Crick & Grotpeter, 1995, p. 713):

Overt aggression: e.g., ”Hits, pushes others”, ”yells and calls others mean names”, starts fights”.

Relational aggression: e.g., ”When mad, gets even by keeping the person from being in their group of friends”, ”tells friends they will stop liking them unless friends do what they say”, when mad at a person, ignores them or stops talking to them”, ”tries to keep certain people from being in their group during activity or play time”.

Boys and girls were shown (Crick & Grotpeter, 1995) to differ in their expression of aggression, whereas 15% of the boys in the study showed overt aggression, under 1% of the girls did. However, 17% of the girls in the study showed relational aggression compared to 2 % of the boys. Suggesting that both genders have relatively equal degree of antisocial propensities: boys show more overt aggression; girls show more relational aggression.

2.2 Obedience and Disobedience

Obedience (and its opposite, disobedience) refers to an individual's propensity to (not) follow explicit instructions or orders from an authoritative figure (Colman, 2009, p. 518). Milgram (1963) demonstrated the extent and impact of authority in his highly controversial experiment where participants were ordered to administer increasingly amounts of [fictitious] electric current to a confederate under the belief that it was part of a learning experiment. The real agenda with the experiment was to see how far the participants were willing to go in obeying orders. Over 83% of the subjects went past delivering 150 volts, and 65% of the participants - both in the experiment containing only men, and when only women were present - showed the propensity to obey until the end, believing they delivered a lethal 450 volts to the confederate. A recent replication (Burger, 2009) of Milgram´s experiment, showed that obedience in the sample from 2009 was slightly lower (see Table 1) than in the original study done in 1963 (65%) This study (Burger, 2009) had an extra condition in addition to the original study: In the other condition the participants watched a (confederate) model refuse to obey the authority in the experiment prior to their own participation. This study (Burger, 2009) saw small gender (although not statistical significant) differences in obedience: A larger percent of the men in the sample stopped before reaching 150 volts, than women. Both in the baseline condition, and in the condition where they saw a confederate refuse orders before entering the experiment themselves. Similar studies using [Jordanian] children (age range: 6 – 16 years) as subjects (Shanab & Yahya, 1977), showed that 73% of the children obeyed until the end of the shock scale. There were no differences in obedience based on sex or age, however, significantly more of the girls than the boys attributed their punishing behaviour to be a function of obeying orders.

Table 1

Results from replication of Milgram´s obedience study. Participants (percentages). (Burger, 2009).

Condition	Men	Women
Baseline		
< 150v	6 (33.3)	6 (27.3)
> 150v	12 (66.7)	16 (72.7)
Model refuse		
< 150v	5 (45.5)	6 (31.6)
> 150v	6 (54.5)	13 (68.4)

These studies provide limited generalizability to an academic setting: being obedient in an academic setting rarely includes hurting others, but it often includes doing tasks (e.g., sitting still, doing homework) that is of little initial interest. The conclusion from this quick review is that the sex differences in obedience from the literature aren´t big enough to generate any hypothesis of which of the genders that is stereotypically more linked with obedience.

2.3 Intellect

Intelligence encompasses a multitude of definitions and operationalizations. Neisser et al., (1996, p. 77) defines intelligence (modestly) as people´s ability to understand complex ideas, adapting effectively to the environment, to learn from experience, engage in various forms of reasoning, and to overcome challenges by taking control over thought.

Intelligence is closely related to the intelligence quotient (IQ), which have shown to be positively correlated with cortical grey[2] volume, especially in the prefrontal regions of the cerebrum (Reiss, Abrams, Singer, Ross, & Denckla, 1996). However, Narr et al., (2007) found that women show stronger correlations between prefrontal and temporal association cortices and full-scale intelligence quotient (FSIQ), whereas men exhibit correlations with FSIQ primarily in temporal–occipital association cortices. This may explain the differences in the expression of intelligence between the genders, men have consistently shown to be superior to women on

[2] Grey matter refers to unmyeliated nerve cell bodies and neuropil in the central nervous system (CNS) (Wickens, 2009).

mental rotation tasks (Masters & Sanders, 1993), whereas women show superior inhibition compared to men (Bjorklund & Kipp, 1996).

Cognitive operations is rooted in neuroanatomical structures (i.e., the central nervous system, CNS), so studies considering these structures is of interest for investigating intelligence (Sternberg, Mio, & Sternberg, 2012). Magnetic resonance imaging study considering sex differences (N = 85, 64 females; M_{age} = 10.55, SD = 2.75, age range = 5-17 years) in neuroanatomical structures and (cross-sectional) development revealed that boys in the sample had significantly higher total grey volume (d = 1.02), and cortical grey volume (d = 1.05) compared to girls total grey volume (see table 2, and figure 1). Cohen (1977) defines effect sizes (*d)* as *small* if they are ≤.20, *medium* if they are ≤.50, and *large* if they are ≥.80.

Table 2.

Gender differences in volumetric displacement in neuroanatomical regions, units in cm^3 *(standard deviation)* (Reiss et al., 1996).

Region	Male	Female	p	Effect size (*d*)
Cerebrum	1290.6 (147.4)	1182.5 (104.7)	<.0005	.85
Right Hemisphere	643.9 (72.6)	591.1 (52.5)	<.0005	.84
Left Hemisphere	646.7 (75.0)	591.4 (52.8)	<.0005	.86
Total grey matter	707.4 (71.4)	637.2 (66.1)	<.0001	1.02
Cortical grey matter	674.8 (67.8)	605.6 (64.0)	<.0001	1.05
Caudate nucleus	11.1 (1.9)	10.7 (1.6)	n.s.	.23
Lenticular nucleus	9.9 (1.6)	9.5 (1.6)	n.s.	.25
Thalamus	11.6 (2.2)	11.4 (1.8)	n.s.	.10
White matter	488.3 (92.0)	443.8 (61.7)	<.01	.58
Total CSF	94.9 (28.8)	101.5 (25.1)	n.s.	-.24
Extraventricular CSF	79.4 (23.8)	89.6 (23.7)	n.s.	-.42
Lateral ventricles	15.5 (9.5)	11.9 (5.7)	<.05	.47

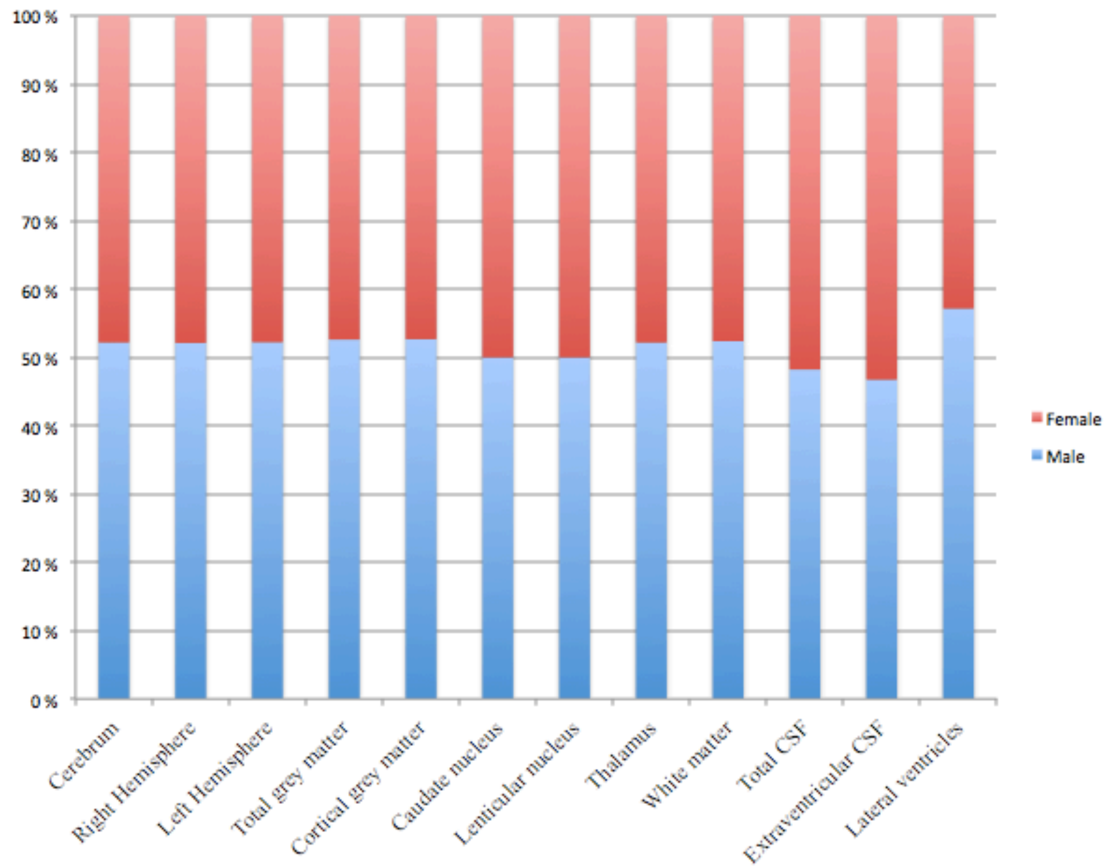

Figure 1. Gender relative displacement in neuroanatomical regions (adapted from Reiss et al., 1996).

Neuroanatomical markers for intelligence are also shown to be linked to the dynamics in cortical grey development, and not simply the static state of the cortex obtained from one measurement (Shaw et al., 2006). Following this notion, results from Reiss et al., (1996) indicate that the cross-section of girls in the sample show a somewhat stronger positive correlation between age and white matter[3] than the cross-section of the boys in the sample (albeit not significant, see figure 2). An increased quantity of white matter can interpreted as a result of increased connectivity between neurons. The decline in grey matter can be attributed to neuronal selection and refinement (further explanation of this in section 4.1 Theory of neuronal group selection)

[3] Outbound communication from nerve cells is transmitted via axons springing out of nerve cells, these axons are coated with a sheath of myelin to increase the velocity of the electric signals interconnecting nerve cells. Myelin is produced by oligodendrocytes (supporting glia cell in CNS) which appear white in color because of its fatty composition, hence the term "white matter" (Wickens, 2009).

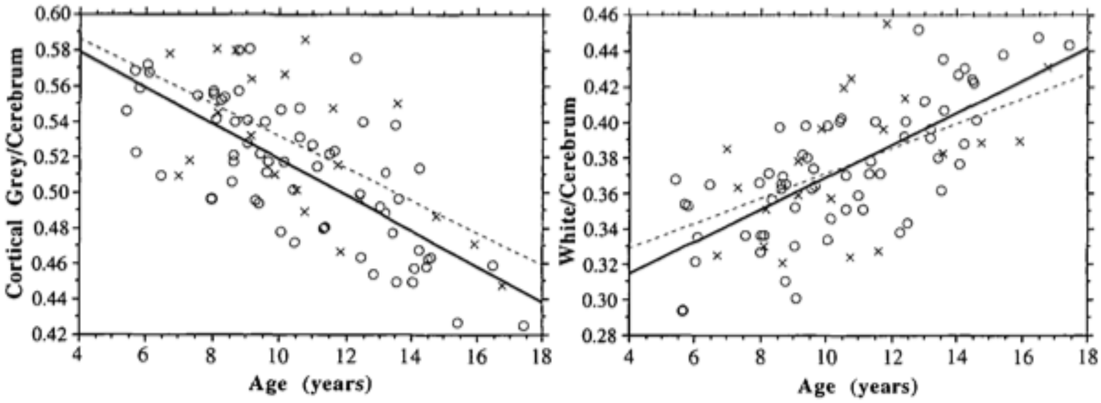

Figure 2**.** Development of the proportion of grey (nerve cells), and white (connectivity) matter in the cerebrum (Reiss et al., 1996).

Note. Crosses and dashed lines = male. Circles and continuous lines = female

The notions from the reviewed literature indicate that men and women express intelligence in different ways. Males show superior spatial intelligence, females transcend men in executive function (i.e., inhibitory competence). In a classroom setting, inhibitory competence (pay attention and inhibit impulses) may be more desired trait than spatial intelligence, compared to an arena where spatial intelligence is more important (e.g., manoeuvring in terrain, playing soccer, building artefacts).

3. Stereotypes and Academic Achievement

Stereotypic expectations have the propensity to impact academic performance (Aronson et al., 1999; Hartley & Sutton, 2013; Latsch & Hannover, 2014a, 2014b; Spencer, Steele, & Quinn, 1999; Steele & Aronson, 1995). The purpose of this chapter is to denote and describe the origin and phenomenon of stereotyping (3.1), including a functional explanation of stereotyping, and ultimately draw the link between stereotypes and academic performance (3.2).

3.1 Stereotypes

The word *stereotype* comes from the Greek words στερεός (stereos), meaning "firm, solid" and τύπος (typos), "impression" - hence, "solid impression" (Liddell, Scott, Jones, & McKenzie, 1996). A stereotype is [in psychology] defined as "psychological representations of the characteristics of people that belong to particular groups" (McGarty, Yzerbyt, & Spears, 2002).

Stereotyping is a fundamentally automatic process (Devine, 1989), and although the etymological aspect of stereotyping implies rigidity, multiple studies have demonstrated the malleability of these automatic stereotypes through several interventions (Blair, 2002): e.g., diversity education (Rudman, Ashmore, & Gary, 2001), self-image motives (Sinclair & Kunda, 1999), social relations (Lowery, Hardin, & Sinclair, 2001) and promoting counter-stereotypes (Blair & Banaji, 1996; Blair, Ma, & Lenton, 2001).

A functional explanation of stereotypes is provided by McGarty et al., (2002), and encompasses three principles: (a) stereotypes are aids for explanation, they serve as the foundation of a social categorization process, forming the initial impression of a group. Without this categorization, we would not be able to tell the difference between one group and another. Categorization is the cognitive process by which we detect those differences and similarities between groups. This need for categorization can be explained by our fundamental need for control over the environment (White, 1959), as information about the environment is imperative for the instantiation of perceived control.

(b) Stereotypes are energy saving devices (Macrae, Milne, & Bodenhausen, 1994; McGarty et al., 2002): perceiving individuals as group members rather than individuals saves limited cognitive energy. Because we can ignore the majority of diverse and detailed information associated with an individual and still maintain a certain degree of overview over expected behaviour if take group membership and associated behaviour for granted. This ”energy saving” often results in misunderstanding, and as Asch (1952) remarked over 60 years ago: ”Stereotypes has come to symbolise nearly all that is deficient in popular thinking” (p. 232).

Finally (c), stereotypes are shared group beliefs, drawn from a common cultural pool of social knowledge. Shared stereotypes are useful for understanding and predicting behaviour of members of one group from another. The process by which this occurs is well understood: intra-group members engage in processes of differentiation to make their own group distinctive from other groups (McGarty et al., 2002). Intra-group consistency in social information (i.e., stereotypical belief) also fuels humans fundamental [social] need to belong (Baumeister & Leary, 1995).

3.2 Stereotype Threat

Stereotype threat has refers to the phenomenon whereby individuals which identify with a particular stigmatized group show reduced performance on tasks requiring cognitive resources (Steele, 1997). A focused aim of social psychologists investigating stereotype threat has been to discover which psychological variables affect individuals´ receptivity to this threat.

For a stereotype to be threatening, it must be negative, and self-relevant. However, one need not believe the stereotype nor even be worried that it is true of oneself, it´s enough to believe that it is expected behaviour from one´s social group. This phenomenon is not tied to the psychology of the particular stigmatized groups, rather it affects the member of any group whom there exist a known negative stereotype (Aronson et al., 1999; Steele, 1997). Stereotype threat can be conceptualized as a situational predicament, felt in situations where one can be treated in the terms of, judged by, or self-fulfil negative stereotypes about one´s perceived social group (Spencer et al., 1999). This threat is triggered by situations where group identity associated with a negative stereotype is made salient, e.g., asked to report your gender before a math test (Aronson et al., 1999), or your ethnicity before a verbal test (Steele & Aronson, 1995). When such a setting integrates both stereotyped and non-stereotyped individuals, it may make the stereotype more salient and thus more strongly felt (Schmader, 2002). This mechanism also explains the variability of stereotype threat; different group experience different forms and degrees of stereotype threat because the stereotypes about them differ in content, in scope, and in the situations to which they apply. For the individual residing in the stereotyped group, this can result in cognitive imbalance in which ones expectations of success conflicts with the social stereotype suggesting poor performance. This state of imbalance functions as an acute stressor that triggers physiological manifestations of stress, cognitive monitoring, and affective responses. These processes demands cognitive resources, resulting in lesser resources available for solving the real task at hand, resulting in poorer performance (Schmader et al., 2008; Spencer et al., 1999). Additionally, if the threat is experienced in the midst of domain performance (e.g., classroom presentation, test-taking), the emotional reaction it causes could directly interfere with performance (Steele, 1997).

Stereotype threat disrupts performance via three distinct, yet interrelated, mechanisms: (a) a physiological stress response that directly impairs prefrontal processing, (b) a tendency to actively monitor performance, and (c) efforts to suppress negative thoughts and emotions in the service of self-regulation (Schmader et al., 2008).

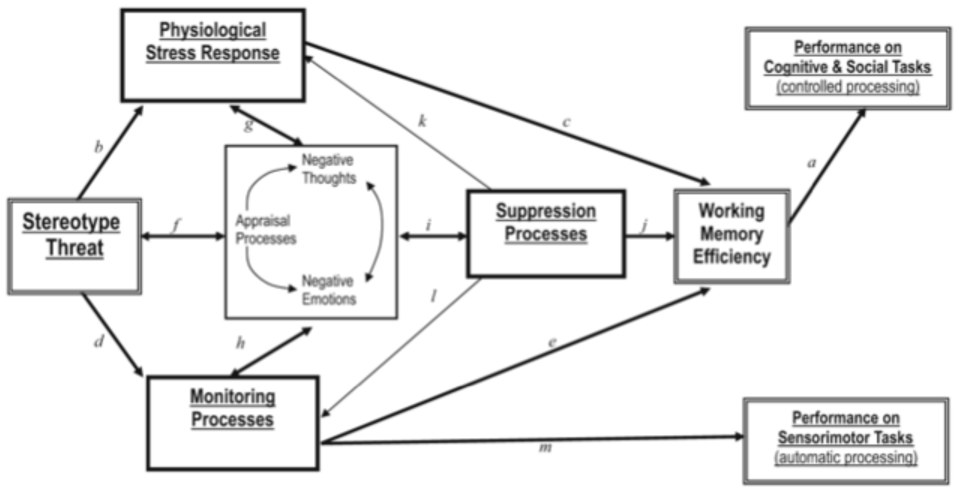

Figure 3**.** Mechanisms of stereotype threat (Schmader, et al., 2008).

The topic will now return to Hartley and Sutton´s (2013) study to make the connection between stereotype threat and academic performance in children. In their second study Hartley and Sutton (2013) measured children´s performance in the domains of math, reading, and writing exercises with two experimental conditions. The sample consisted of a total of 162 British schoolchildren (80 boys and 82 girls) aged 7–8 years (M_{age} = 7.40), and divided into two groups containing an equal amount of boys and girls. The first group performed math, reading and writing tests under the saliency of the stereotype that girls perform better: “We’ve looked at how well children do on this test and we have found that girls do better than boys. Boys don’t do as well”. The second group functioned as a control where no stereotypes where made salient: “We’ve looked at how well children do on this test and we just want to see how you do”. The results show that boys perform significantly worse under the stereotype threat condition, and that this effect was absent in the control group.

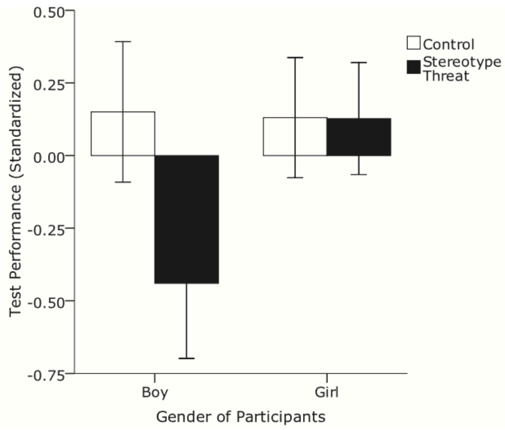

Figure 4. Test performance as a function of stereotype threat condition. Error bars represent 95% confidence interval (Hartley & Sutton, 2013).

Summarized, the main motive behind this chapter was to describe the phenomenon of stereotyping, and give a functional explanation of stereotyping. Following up with the potential implications of possessing negative stereotypes, conceptualized by stereotype threat. Further, explaining the mechanisms behind stereotype threat. Ending with the link between stereotypes and academic performance

4. Implicit Social Cognition

There is a distinction between the processes that are automatic, rapid, and implicit/unconscious (system 1), and those that are deliberate, slow and explicit/conscious (system 2) (Evans, 2008). The former processes are phylogenetically younger than the latter processes (Evans & Over, 1996; Stanovich, 1999), and signifies a more basic and more effortless level of processing. The occurrences of system 1 processes depends only on the presence of a stimulus and those basic conditions that ensure that the stimulus is physically registered (Moors, Spruyt, & De Houwer, 2010). System 2 is a reflective process (Strack & Deutsch, 2004) – considering multiple assemblies of neuronal maps before arriving at the most appropriate option, and is dependent upon selective attention and taking control over the mind. Implicit social cognition is concerned with the automatic processes (i.e., system 1 processes) that govern social behaviour (e.g., social judgements, attitudes, stereotypes). Virtually every intellectual inquiry in the field of social psychology has in the last two decades been shaped to some extent by the theories and methods of implicit social cognition (Gawronski & Payne, 2010). Greenwald and Banaji (1995) defines implicit social cognition as "introspectively unidentified (or inaccurately identified) traces if past experiences that mediate favourable or unfavourable feeling, thought, or action towards social objects" (p. 5). These traces of experience is subject for explanation in the next section (4.1). Following up with implicit modalities (4.2), which explaining how implicit associations shape automatic responses in the visual and auditive modalities.

4.1 Theory of Neuronal Group Selection

The theory of neuronal group selection (TNGS) (Edelman, 1992) will be used as a theoretical framework for explaining how mental representations (e.g., concepts, associations) manifests themselves physically in the mind, and consequently in external observable behaviour (e.g., response latency). Mental representations are the residue of a lifetime of observation, thought and experience, from these events, neural connections in dynamic harmony appear (Carlston, 2010). Edelman´s (1992) theory explains this through three tenets, whereas the second and third is of greatest

relevance for the current thesis: The first tenet proposes the (1) development of a primary repertoire of neuronal connections during the prenatal stage, and consist basically of neuronal- proliferation, migration, and apoptosis guided by cell adhesion molecules (CAM) and substrate adhesion molecules (SAM) with probabilistic and reciprocal properties (i.e., CAM and SAM guides morphology, and morphology also guides the function of the CAM and SAM). After this stage is set, an experiential stage (2) is initiated. Key events in the experiential stage is the development and refinement of neuronal maps (i.e., groups of neurons connected synaptically) through experience - this is the stage where mental concepts arise. The retrieval process of these maps is reconstructive, this proposes a potential integration of several different concepts (i.e., association) during retrieval, which also is a part of the next stage. The third and last stage is called re-entry, and describes the integration of neuronal maps, which in turn gives rise to a spatiotemporal conscious experience.
The ideas presented by TNGS supports a response latency-based measure as an appropriate paradigm for measuring stereotypes: as the retrieval of related concepts and categories will be more effective (i.e., shorter response latency) than non-related concepts, as a result of neuronal integration.

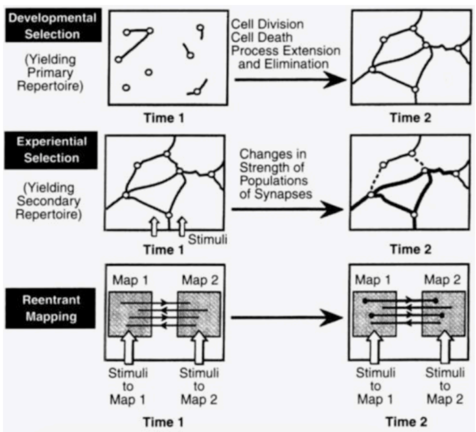

Figure 5. Visual presentation of the theory of neuronal group selection (Edelman, 1992).

4.2 Implicit Modalities

With the theory of neuronal group selection still cognitively salient. I will now explain how this relates to implicit associations and how different modalities (i.e., implicit perception) is affected by implicit associations. Implicit associations are a set of strongly interconnected synaptical networks, with such connective strength that it guides system 1 (automatic) behaviour, and ultimately forms the basis for system 2 (conscious) behaviour (Evans, 2008). One of the earliest measurements that revealed the interfering effect of system 1 was done by Stroop (1935). In his famous experiment, Stroop (1935) demonstrated that an interference effect occurs when participants are asked to report the inherent aspect of words (e.g., the colour they are written in), rather than their semantic meaning. This effect can be explained as follows: when we read words, the most frequent and natural behaviour is to read the semantic meaning rather than focusing on inherent characteristics. When we are asked to report other characteristics than what is ”normal”, we have to implement system 2 to override the natural response of just reading and reporting semantic meaning, this takes more time than automatic system 1, and results in a slower response time.

Stroop (1935) addressed the visual modality, however, later studies have addressed the automatic effects in the auditory modality as well. In Most, Sorber and Cunninghams´s (2007) auditory stroop experiment, participants categorized the sex of voices saying names and words stereotypically associated with male (e.g., baseball, captain, gun, pirate), female (e.g., bracelet, cheerleader, doll, lipstick) or neutral (e.g., apple, door, draw, paper) words. Both adults (N = 42, 21 females) and children (N = 48, 23 females) were slower when the voice's sex was stereotypically incongruent with the spoken word or name (see table 3).

Table 3

Auditory stroop effects in adults and children as a function of stereotypical congruent, incongruent and neutral mode of delivery. Mean response time in milliseconds with standard deviations in paranthesis. (Most et al., 2007).

	Congruent	Incongruent	Neutral	Difference
	Adults			
Words	780 (190)	802 (195)	787 (198)	23 (29)
Names	782 (187)	829 (214)		47 (39)
	Children			
Words	841 (159)	907 (196)	846 (162)	66 (90)
Names	873 (165)	902 (193)		28 (77)

The main points from this chapter was to introduce the implicit level of processing. and how mental associations manifest themselves in the mind. To argue that automatic mental processes precedes and shapes our conscious behaviour, with the ultimate intention to motivate the development of an indirect measure that takes automatic visual and auditive perception into account.

5. Measuring Psychological Constructs.

Psychometrics is a general discipline in psychology that involves the statistical models and measurement procedures that have been developed to infer, describe, and summarize psychological constructs from empirical data collected in psychological research (Hayes, 2000). Jones and Thissen (2007) divides psychometrics into three divisions: (1) tests, a set of procedures and methods to assign quantitative values to objects or events using data from participant assessments; (2) methods and procedures derived from factor analysis, which try to explain the observed correlation between a set of variables by looking at variations in latent (i.e. unobserved, constructed) variables in order to capture most of the variation in the observed variables (i.e., data reduction); and (3) measurement theory, which forms a global theoretical foundation for psychological research methods, where topics such as validity and reliability are addressed.

This thesis will ultimately spring from the first and third division - test development and validity. More specifically, it will aim for the successful development of a nomothetic measurement procedure which successfully manages to capture implicit mental associations between psychological constructs (gender/sex and characteristics). The term psychological construct will in this thesis encompass the *internal* information that a person possess about themselves and the external world – which in turn impact intra- and interpersonal behaviour (Ajzen, 1991). However, this internal information does not reflect actual conditions, nor do they manifest themselves in the physical world (Bunge, 1974) - they are *constructed* (Edelman & Tononi, 2000), hence the term mental construct. Furthermore, this information is generally only accessible from the individual who possesses the construct, through valid psychological measurement. A measurement is valid if, and only if (1) the construct exist, and (2) inter-individual variations in the construct causally leads to inter-individual variations in the measurement outcome (Borsboom, Mellenbergh, & van Heerden, 2004).

This chapter will describe direct and indirect psychological measurement procedures (5.1), following with a thorough explanation of the implicit association test (5.2).

5.1 Direct and Indirect Psychological Measurement Procedures

On a general level, psychological measurement procedures can be distinguished into two groups: *direct* and *indirect*. De Houwer et al., (2009a) defines *direct* and *indirect* as descriptions of the measurement *procedure*; and *implicit* and *explicit* as descriptions of the *psychological attributes* that are evaluated by the different measurement procedures.

Direct measurement procedures are characterized by two properties: (1) the measurement outcome is the result of a self-assessment by the participant, and (2) the target of the self-assessment is the attribute that the measurement outcome is assumed to capture. Furthermore, the outcome of a direct measurement is based solely on the responses from the participant. Conversely, indirect measurement procedures are characterized by the fact that they do not require self-assessment from the participant, and that the outcome of the measure need an additional step to make sense - the measurement outcome is not based on the responses per se, but rather the interpretation of these responses (De Houwer & Moors, 2010).

Several limitations exist for both direct and indirect measurement procedures, the former demonstrates issues related to restraints in introspective access (Greenwald & Banaji, 1995; Wilson & Dunn, 2004) and response bias (Fazio & Olson, 2003). The latter is sensitive to context effects (i.e., temporary shifts in implicit cognition as a result of a particular context) (Gawronski & Sritharan, 2010).

Response bias is a generic term describing motivational factors which bias the response from a correct, honest, and accurate response in self-report measures (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). This includes social desirability which results in faking ”good” responses, as well as its opposite - faking ”bad” responses (e.g., as a result of psychological reactance). Social desirability is defined by Crowne and Marlowe (1964) as ”the need for social approval and acceptance and the belief that it can be attained by means of culturally acceptable and appropriate behaviours” (p. 109). Psychological reactance (Brehm, 1966) on the other hand is when a person feels that someone or something is limiting his or her choices or limiting the range of personal alternatives, potentially resulting in malignant behaviour (e.g., faking response).

Other relevant response biases is extremity response set, where the respondent always chooses extreme opposites, and mid-point response set, where the respondent deliberately chooses a moderate response (Furnham, 1986). These kinds of response sets is claimed to stem from the nature of the question (i.e., how it is presented), as well as the motives of the respondents (Kalton & Schuman, 1982).

Context effects appears when situational cues activate associative neuronal structures which interferes with the measurement. These can be avoided by keeping the environment neutral - in sense to attitudes related to the topic of measurement - before the measurement procedure is initiated.

An indirect measure should conform to three normative criteria: (1) It should be clear which attributes the measure reflects (what-criterion); (2) the nature of the processes by which the attributes cause variations in the measure should be known (the how-criterion); and (3) it should be clear that the underlying processes are automatic in a certain manner (the implicit-criterion) (De Houwer, Teige-Mocigemba, Spruyt, & Moors, 2009b).

This project conforms to these three criteria:

(1) The what: Stereotypes, associations between groups of people (male and female) and certain characteristics (propensities for academic behaviour).
(2) The how: Variations in the measure depends on neural integration (Edelman, 1992).
(3) The implicit: Stereotyping is an automatic process (Devine, 1989).

A multitude of indirect measurement procedures have been developed in the last two decades (see (De Houwer & Moors, 2010) p.176) However, the most profound and imperative contribution in the indirect measurement paradigm regarding implicit social cognition has been Greenwald, McGhee and Schartz´ (1998) Implicit Association Test (IAT). The IAT has undergone thorough psychometric validation and has been by far the most significant test for measuring implicit associations (Gawronski & Payne, 2010)

5.2 Implicit Association Test.

The Implicit Association Test (IAT) is a double-categorization task that measures relative association strength between two pairs of contrasted concepts (e.g., black and white names, pleasant and unpleasant words) (Greenwald et al., 1998). This is achieved through comparing response latencies between two differently combined categorization tasks. The basic idea behind the IAT is that if two concepts are highly associated, categorization will be easier when they share the same response key (compatible block), than when they require the opposite response (incompatible block). The IAT procedure consists of seven blocks presented in Table 4.

The application details of the IAT encompass: (1) the instruction to respond as quickly and accurately as possible, (2) display of category labels assigned to left key response and right key response in the corresponding upper screen corners, (3) two to eight stimuli words per category, (4) random sequential presentation of words corresponding to the categories, (4) a fixed inter-trial presentation interval of between 150msec-750msec, (5) correction of erroneous response by presentation of an error cue (red X) below the stimulus word, (6) reverse configuration of category labels from the fifth block while the other categories remain in their original placement.

Table 4

Implicit association test procedure. (Greenwald et al., 1998).

Block	Trials	Function	Items assigned to left-key response	Items assigned to right-key response
1	20	Practice	Black names	White names
2	20	Practice	Pleasant words	Unpleasant words
3	20	Practice	Black + Pleasant	White + Unpleasant
4	40	Test	Black + Pleasant	White + Unpleasant
5	20	Practice	White names	Black names
6	20	Practice	White + Pleasant	Black + Unpleasant
7	40	Test	White + Pleasant	Black + Unpleasant

The original IAT (Greenwald et al., 1998) inferred the strength of implicit association between concepts by using latency data from the fourth and seventh block, and log-transforming the latencies before averaging them. The logarithmic transformation was used because it is a statistic offering the required stability of variance for analysis. Some further refinement to the measurement was done by recoding outlying latencies (<300 ms and = >3000 ms) to the nearer boundary value

(300ms and 3000ms, respectively). These outliers typically represent a response initiated before stimuli onset, or a lack of task attention and distorts the mean and variance. It is important to note that the IAT effect need to be interpreted in a relative aspect. E.g., an IAT- effect does not permit any conclusions about a respondent`s evaluation of a group in general, but provides only information about a respondent`s preference for that particular group compared to the contrasting group.

The conventional scoring algorithm was found obsolete by Greenwald, Nosek, & Banaji (2003) a couple of years later in a meta-study which assessed the strength of five different scoring algorithms: *Median; Mean; Logarithmic; Reciprocal*; and *D´*. Utilizing an extensive data-set (*N* = 8218), Greenwald et al., (2003) examined the alternate scoring algorithms in terms of their (a) correlations with parallel self-report measures, (b) resistance to the artefact associated with response speed (i.e., the effect of age of the participant), (c) internal consistency, (d) sensitivity to known influences of IAT measures (e.g., pervasive implicit attitudes and stereotypes), and (e) resistance to known procedural influences (e.g., order-effects). The *D*` algorithm was announced as the best scoring algorithm for IAT, showing stronger correlations with explicit measures, resistance to inter-subject speed variation, higher internal consistency, low order effects, and reasonable effect sizes.

D`-score is computed by this sequence (adopted from Greenwald et. al., 2003, p. 214): (1) Compute the mean of correct latencies for each block. (2) Compute one pooled standard deviation for all trials in block 3 and 6; another for block 4 and 7. (3) Replace each error latency with block mean + 600 ms. (4) Compute the differences: block 3 – block 6; and block 4 – block 7. (5) Divide each difference by its associated pooled standard deviation from step 2. (6) Compute the weighted average of the two differences.

The focus of this chapter have been to highlight the methodological aspects of psychological measurement, and how this current project relates to these notions. The difference between direct and indirect measurement procedures has been discussed, with their respective limitations. Finally, a thorough explanation of the implicit association test has been provided to make the foundation for the present research.

6. The Present Research

The principal aim for this research is to design an audio-visual indirect measurement procedure based on the IAT to detect and describe the direction, magnitude, and development of implicit academic gender stereotypes in elementary school children. To the best of my knowledge, no IAT on academic gender stereotypes, suitable for adults or children, has been published. Thus, all stimuli materials had to be collected, selected, prepared and piloted, and finally incorporated into the measurement procedure from scratch. This was done through several interviews with the target sample (children attending elementary school), spanning over several days and including plenary interviews of all grades.

A second step for this research is to use the IAT on a target sample. If similar gendered academic stereotypes exist in Norway as in other European countries (e.g., Hartley & Sutton, 2013; Latsch & Hannover, 2014), and if the experimental (IAT-) procedure is sufficiently sensitive to capture these stereotypes, then this would be reflected in facilitated responses to female/pro-academic behaviour pairs compared to male/pro-academic behaviour pairs.

Since the initial sample withdrew unexpectedly from the study, and no other equivalent sample was available for measurement, the measure will be employed in a sample of young adults (age 20-25).

I expect that young adults show an IAT effect (Hypothesis 1). As the gendered academic stereotype is in favour of women, automatic in-group biases may amplify or attenuate such an IAT effect. Hence, I expect that women will show a stronger IAT-effect than men (Hypothesis 2).

The following sections include the development of the measurement procedure 6.1, with subsections describing the stimuli material (6.1.1, 6.1.2). Following with stimuli production and implementation details (6.2), and a description of the pilot (6.3).

6.1 Development of MeCIAS

The measure of children`s implicit academic stereotypes (MeCIAS) is procedurally based on Greenwald et al., (1998) implicit association test, measuring the relative association between gender and academic behaviour inferred from response latency. The content of MeCIAS is developed by the author based on several interviews with elementary school children. Additionally, since MeCIAS is intended to use on elementary school children who have not yet automated their reading skills - it includes an auditive stimulus corresponding to the visual stimulus word . The trials in MeCIAS present the corresponding visual and auditive stimulus simultaneously. The measurement script is written in Inquisit (Version: 4.0.5.0 64bit, build 2444). Banajii and Baron`s (2006) Child-Implicit Association Test (Ch-IAT) was used as a starting point. The Ch-IAT (Baron & Banaji, 2006) measures the association between good and bad words, and insects and flowers. The words are presented auditory, while the insects and flower are presented visually. The differences in the procedure between the Ch-IAT and MeCIAS are that MeCIAS presents visual and auditory stimuli in parallel (i.e., simultaneously), and that Ch-IAT presents words auditory, and pictures visually. Additionally, MeCIAS uses the voice of a young child rather than the voice of an adult women to bypass auditory stroop effects.

Table 5

Experimental setup in MeCIAS

Gender	Academic behaviour
Boy names	Pro academic characteristics
Girl names	Counter academic characteristics

6.1.1 Selection of names. Names from Statistics Norway was used, based on the 30 most popular names in Norwegian statistical history a sample of 16 names was extracted, eight male and eight female. This sample of names was satisfactory on a range of surface cues: the names where statistical equal in length between the genders; and they started and ended with a consonant and vowel an equal amount of times.

Table 6

Names used in MeCIAS, sorted alphabetic and for gender.

Stimulus item	Gender	Audio signal length in ms	Visual length i letters
Anders	M	660	6
Arne	M	710	4
Espen	M	670	5
Henrik	M	670	6
Ole	M	510	3
Rune	M	620	4
Stian	M	740	5
Terje	M	610	5
Anita	F	720	5
Berit	F	760	5
Elin	F	560	4
Hilde	F	620	5
Ida	F	580	3
Ingrid	F	620	6
Maria	F	680	5
Marit	F	800	5

Note. Average audiosignal length for boy names = 648.8ms, average length in letters = 4.75 letters. Average audiosignal length for girl names = 667.5ms, average length in letters = 4.75 letters

6.1.2 Selection of academic characteristics. Characteristics describing pro- and counter academic behaviour were gathered from primary school children attending elementary school in Trondheim. Grades through 1st to 7th was interviewed in plenary and asked to report good academic behaviour (pro-academic) and bad academic behaviour (counter-academic) verbally in their respective classrooms. A total of 118 descriptions were gathered (64 good; 54 bad; see appendix for an exhaustive list). The list was reduced via synonym extraction, and coded into sub-facets. Five facets emerged from this process: Prosocial, antisocial, obedience, disobedience, and intellect. The descriptions that were reported multiple times and the ones that converged in the synonym extractions were weighted as more important in the selection for the IAT procedure.

The selection of stimulus words was based on three criteria: (1) how often the behaviour was the reported in interviews, (2) convergence with existing theory, and (3) free for surface effects (e.g., confounding response latency because of stimuli length).

Table 7

Academic characteristics in MeCIAS (english translation in parenthesis)

Stimulus item	Category	Audio length in ms	Visual length in letters
Smart (Smart)	Intellect	760	5
Flink (Clever)	Intellect	650	5
Snill (Nice)	Prosocial	630	5
Vennlig (Friendly)	Prosocial	760	7
Lydig (Obedient)	Obedience	580	5
Rolig (Calm)	Obedience	550	5
Slem (Mean)	Antisocial	590	4
Utestenge (Exclude)	Antisocial	990	9
Slåss (Fight)	Antisocial	750	4
Bråke (Brawl)	Disobedience	670	5
Hærverk (Vandalism)	Disobedience	860	7
Urolig (Restless)	Disobedience	620	6

Note. Average pro academic stimulus length = 655ms, 5.33 letters; Average counter academic stimulus length = 626.67, 5.17 letters

.

6.2 Stimuli Production and Implementation

All the descriptions were recorded in a sound studio, the voice stimuli were provided by a 9-year old girl. This was to make an effort to avoid auditory stroop-effects (discussed in implicit modalities). The voice of children is more ambiguous than adult voices in regards to which gender it belongs to, since the vocal chords which distinguish men from female voices are developed in a later stage of life.

The measurement procedure script was developed using Inquisit (Version: 4.0.5.0 64 bit, build: 2444). The script is based on Banajii and Baron`s (2006) Child-Implicit Association Test (Ch-IAT), and further developed to administer visual and auditory stimuli in parallel: The visual (text) and auditory (voice) information is presented simultaneously, with the possibility to respond before the auditory stimulus ended. The inter-trial interval was set to 400 milliseconds. A correct response is necessary to advance to the next trial. MeCIAS is coded such as a negative score (below 0) reflects a faster response when boy names is paired with pro-academic behaviour, while a positive score (over 0) means a faster response when girl names is paired with pro-academic behaviour. The script contains two different block order conditions: Block condition number 1 combines boy names with pro-academic adjectives first; Block order condition number 2 combines girl names with pro-academic adjectives first (see Table 8). The block order condition is randomized.

Table 8

MeCIAS procedure.

Block	No. Of Trials	Function	Items assigned to left-key response	Items assigned to right-key response
1	20	Practice	Boys names	Girls names
2	20	Practice	Pro-academic	Counter-academic
3	20	Practice	Boys + Pro-academic	Girls + Counter-academic
4	40	Test	Boys + Pro-academic	Girls + Counter-academic
5	20	Practice	Girls names	Boys names
6	20	Practice	Girls + Pro-academic	Boys + Counter-academic
7	40	Test	Girls + Pro-academic	Boys + Counter-academic

Note. Block order in condition 2 = 5,2,6,7,1,3,4.

6.3 Pilot

Firstly, the selected academic items were tested for face validity: four women and three men (M_{age} = 22 years) sorted the randomized academic items (without indications of which group they belonged to) into either pro-academic or counter academic, all items were sorted correctly. The pilot (N = 10, M_{age} = 23.8 years, 5 females, M_{time} = 5 min 40 sec), revealed no significant outliers on item response time on the practice trials. *D*`-score is computed by this sequence (adopted from Greenwald et. al., 2003, p. 214): (1) Compute the mean of correct latencies for each block. (2) Compute one pooled standard deviation for all trials in block 3 and 6; another for block 4 and 7. (3) Replace each error latency with block mean + 600 ms. (4) Compute the differences: block 3 – block 6; and block 4 – block 7. (5) Divide each difference by its associated pooled standard deviation from step 2. (6) Compute the weighted average of the two differences.

Average *D*`-score from pilot = .37 (*SD* = .27), producing an effect size of *d* = 1.37. Men scored on average .26 (*SD* = .15), and women scored .49 (*SD* = .33), producing an effect size of *d* = .90. Cohen (1977) defines effect sizes (*d)* as *small* if they are ≤.20, *medium* if they are ≤.50, and *large* if they are ≥.80.

7. Method

The method section will encompass the implementation of MeCIAS on the target sample. G*Power (Version 3.1) is used to calculate sample size. Data from MeCIAS (i.e., response latency and correct response) is analysed using SPSS (version 20).

7.1 Participants

For hypothesis 1 (young adults will show an IAT effect), an a priori power analysis (t-statistics one sample) using G*power based on the effect size from the second pilot (d = 1.37) setting alpha error probability to .05 and the power to .80 indicated a required sample size of 5.

For hypothesis 2 (females will show a stronger effect than males), an a priori power analysis (t-statistics independent samples) using G*Power based on the effect size from the second pilot (d = 0.90), setting alpha error probability level to .05, and power level to .80, indicated a required sample size of 34.

Thirty (15 male, 15 female) Norwegian young adults participated in the study (M_{age} = 22.8 years, SD = 1.77, age range: 20-25 years). Participants were recruited via flyers and snowball sampling. No incentives were given for participation. Informed verbal consent was given from participants after basic information about the study was provided to them. The basic information consisted of telling the participants that they would be performing the test on a computer, and that the objective of the test was to sort single words to the correct side based on different conditions.

7.2 Apparatus and Materials

The measure was administered through a laptop using Inquisit (Version: 4.0.5.0 64bit, build 2444). Laptop specs: Intel Core i5-2520M CPU @ 2.50 GhZ 64-bit, running Windows 7 Enterprise 64-bit (Service pack 1) with a 13.3” screen, using generic headphones for administering sound stimuli.

7.3 Procedure

Participants were tested individually. The two block order conditions (Boys pro-academic first; and girls pro-academic first) were randomized by coin toss prior to the procedure (N^{c1} = 16, 6 males; N^{c2} = 14, 9 males). The participants were seated at a table in front of the laptop, in a well lit and quiet room (eye-screen distance approximately 55 cm), and given the following instructions verbally by the experimenter:

English translation from Norwegian (Norwegian instructions in appendix):

> ”Welcome, and thanks for participating in this study. Your task is to sort the words appearing on the screen and in the headphones as fast and accurately as possible with by using the E and I keys on the keyboard. First, names for boys and girls will appear. Secondly, descriptions of good and bad academic behaviour will appear. In the third part, your task is to sort both names and descriptions to the correct side. If anything is unclear, just ask me”.

The experimenter then started the program and entered the participant gender and test condition. Next, the experimenter instructed the participant to put on the headphones, then, he started the implicit association test and left the room. Instructions were displayed on the screen, instructing the participant to start by pressing the spacebar. All trials demanded a correct response to advance to the next trial. The average duration of the test procedure was 5 minutes 12 seconds (*SD* = 51 seconds). After completing the test, the participants were asked orally by the male experimenter about their explicit stereotype regarding which gender they associated with pro-academic behaviour, and were given three options: Male pro academic stereotype, female pro academic stereotype, or no academic stereotype (scored as -1, 1, and 0, respectively). After this, the participants were explained that the intention of the test was to measure which gender they associated automatically with good academic behaviour, and that their gendered academic stereotype would be inferred based how fast and slow they correctly respond to the combined tasks. Finally, the participants were thanked for their participation.

8. Results

8.1 Data Preparation

Data were prepared in line with Greenwald, Nosek and Banajii´s (2003) improved scoring algorithm, by eliminating trials over 10000ms (no cases), and replacing erroneous response trials with block mean + 600ms. Two hundred and fifty one responses needed recoding (less than 1 % of total responses), see Table 9.

Table 9.

Frequency of corrected responses by block.

Block	Description	Corrected Trials
1	Names	25
2	Characteristics	30
3	Boy/Pro-Academic	66
4	Boy/Pro-Academic	118
5	Names	31
6	Girl/Pro-Academic	42
7	Girl/Pro-Academic	89
	Total	251

Average *D*`-score was .59 (*SD* = 55) 95% CI [0.38, 0.79], generating an effect size of *d* = 1.07. Cohen (1977) defines effect sizes (*d)* as *small* if they are ≤.20, *medium* if they are ≤.50, and *large* if they are ≥.80.

Internal consistency was examined by comparing subjects *D*`-scores from block 3 and 6, with those from block 4 and 7. Cronbach alpha for the two *D*`-scores was reported to .79, indicating good reliability (Cronbach, 1951).

8.2 Preliminary Analysis: The Influence of Block Order.

To test whether manipulating the block order (block condition: boy names & pro-academic adjectives first vs. girl names & pro-academic adjectives first) had an impact on participants' correct responses: chi-square tests was performed to examine the effect of block order condition on initial correct response per block. Following up with an examination of block condition effects on average response latency in separate blocks. Finally, examining block condition effects on *D*`-scores.

A chi-square test of independence was performed to examine the relation between block order and initial correct answer per block. The relation approached statistical significance, X^2 (1, N = 5400) = 3.69, p = .055. An equal test block-wise, revealed that it was block 6 (combined task, female and pro academic) generated the largest difference between block condition 1 and 2 (see Table 10).

Table 10

Correct response in separate blocks as a function of block order.

Block	Description	Boy/Pro academic first Correct	Girl/Pro academic first Correct	Difference	*p*
	Overall	92 %	93 %	-1	.055
1	Names	95 %	97 %	-2	.308
2	Characteristics	95 %	95 %	0	.708
3	Boy/pro	89 %	89 %	0	.795
4	Boy/pro	89 %	91 %	-2	.245
5	Names	95 %	95 %	0	.854
6	Girl/pro	91 %	95 %	-4	.055
7	Girl/pro	92 %	93 %	-1	.441

The effect of block order on response latency overall and in separate blocks are presented in Table 11.

Table 11

Average response latencies as a function of block order.

Block	Description	Boy-pro first Average (*SD*)	Girl-pro first Average (*SD*)	*d*	*p*
	Overall	818.62 (411.62)	854.81 (400.18)	-0,09	.001
1	Names	852.35 (558.47)	785.19 (283.71)	0,16	.064
2	Characteristics	747.52 (739.13)	708.70 (311.52)	0,07	.402
3	Boy/pro	1038.04 (537.67)	1069.40 (608.53)	-0,05	.504
4	Boy/pro	894.24 (394.23)	958.55 (470.53)	-0,15	.010
5	Names	661.19 (243.53)	739.07 (303.36)	-0,28	.001
6	Girl/pro	788.97 (374.38)	928.58 (459.43)	-0,33	<.001
7	Girl/pro	748.01 (330.91)	794.18 (364.19)	-0,13	.022

The average difference in D`-scores between boy/pro first (*M* = .69, *SD* = .55) and girl/pro-academic first (*M* = .47, *SD* = .55) was 0.22, generating an effect size of *d* = .40, $t(28) = 1.09$, $p = .283$.

In sum, the preliminary results indicate that if blocks were presented such that girl names and pro-academic characteristics had to be combined first, participants showed a higher correct response rate than if male names and pro-academic characteristics had to be combined first. If blocks were presented such as boy names and pro-academic characteristics had to be combined first, participants showed a generally faster response rate than if girl names and pro-academic characteristics were to be combined first. Finally, if boy and pro-academic was combined first, it produced a stronger *D*`-score than if girl and pro-academic was combined first.

However, sex of participant was not equally distributed across the two block order conditions. Block order 1 (boys and pro-academic pairs first) consisted of 6 men and 10 women; block order 2 (girls and pro-academic pairs first) consisted of 9 men and 5 women (i.e., twice as many women in block order 1 as in block order 2). This may result in illusory block order effects that actually reflect sex of participant effects, especially when testing for block order effects on response latency and *D*`-scores (bearing in mind that the effect size of sex of participant on *D*`-scores was reported to *d* = .90 in the second pilot).

8.3 Hypotheses Testing

Hypothesis 1 stated that "Academic stereotypes is in favour of females, and that these gendered academic stereotypes will be expressed through facilitated responses to female and pro-academic behaviour pairs compared to male and pro-academic behaviour pairs."

The average sample *D*`-score was .59 (*SD* = .55). A one-sample t-test revealed that this *D*`-score was significantly different from 0, $t(29) = 5.89$, $p < .001$, 95% CI [0.38, 0.79], generating an effect size of *d* = 1.07. This effect size is considered large (Cohen, 1977), and is in the direction predicted by the first hypothesis.

As the gendered academic stereotype is in favour of females (hypothesis 2), it was hypothesized that automatic in-group biases (in-group favouritism) would amplify or attenuate the *D*`-effect, resulting in a higher *D*`-scores from women than from men.

An independent T-test with sex of participant as grouping variable and the *D*′-score as test variable revealed an effect of sex of participant with females (M = .98, SD = .36) receiving higher scores than males (M = .20, SD = .41) $t(28) = -5.53$, p <.001. The sex difference generated a very large effect size of 2.03 (Cohen, 1977) and is in direction predicted by the second hypothesis.

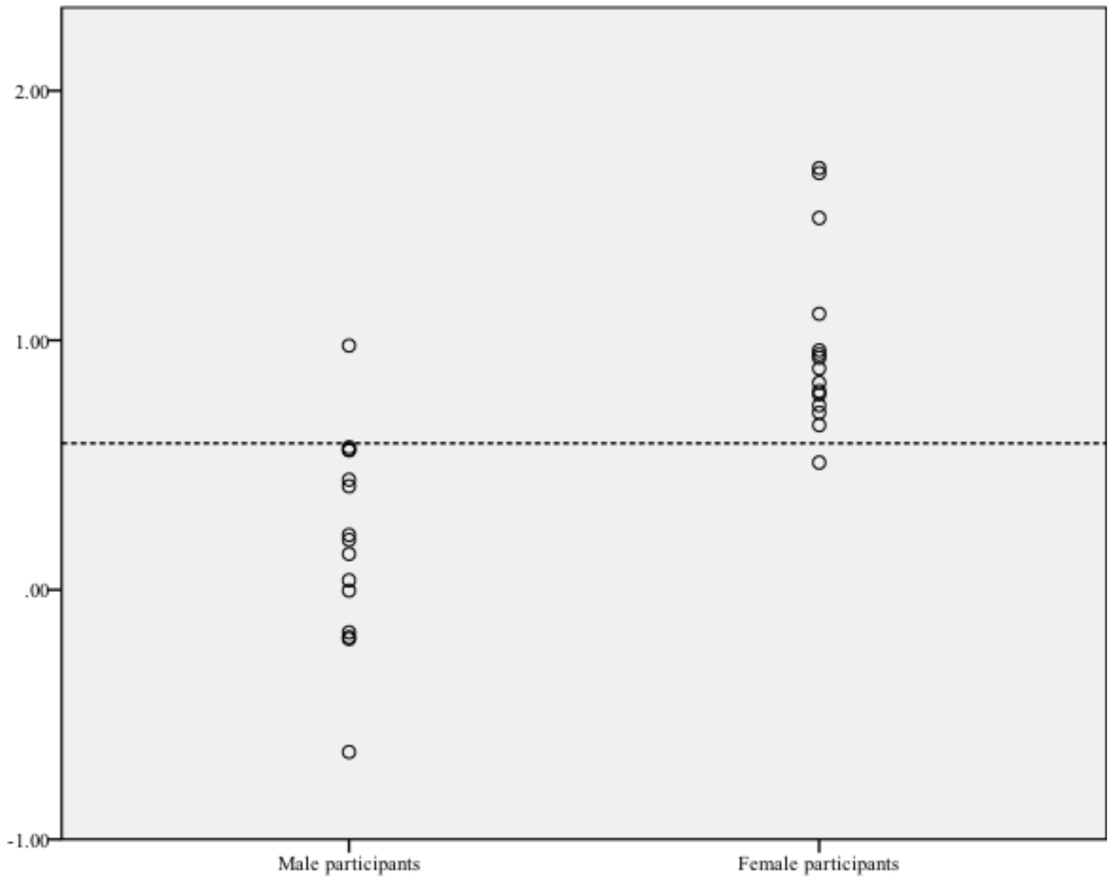

Figure 6. Scatterplot of *D*`-scores as a function of sex of participant. Dashed line indicates average *D*`-effect

Note. Positive *D*`-score indicate a pro-academic female implicit stereotype.

Given the strong sex of participant effect and apparent block order effects on *D*`-score, a 2x2 ANOVA with sex of participant and block order as between subjects factor was performed to examine if these effects are qualified by a block order by sex of participant interaction. The analysis of variance revealed a main effect of sex of participant $F(1,26) = 26.13$, p = <.001, $\eta^2 = .501$. However, no main effect was found for block order *F*(1,26) <1, and no significant interaction effect between sex of participants and block order *F*(1,26) <1.

Suggesting that the block order effect on response latency and *D*`-effect that appeared in the preliminary analysis was illusory and caused by the randomization error (the sex of participant was not equally distributed across the two block order conditions).

Furthermore, the strong sex of participant effect facilitated the motive to check whether both sexes differed from 0. A one-sampled t-test split for sex of participant revealed that the men in the sample had an average difference from 0 of .19 (*SD* = .41), *t(*14) = 1.84, *p* = .087, and the women in the sample had an average difference from 0 of .98 (*SD* = .36), *t*(14) = 10.54, *p* <.001.

As a note, the exclusion of *one* particular male participant would have reduced the standard deviation and amplified the average male *D*`-score (*M* = .25, *SD* = .35) and pushed the male group below the alpha limit, *t*(13) = 2.73, p = .017.

8.4 Exploratory Research – Implicit vs. Explicit Academic Stereotypes

After the initial measure, participants were asked to report their explicit academic gender stereotype. They were given three options: pro-male academic stereotype (= -1); neutral/no stereotype (= 0); or pro-female academic stereotype (= 1). None of the participants reported an explicit pro-male academic stereotype. Surprisingly, there were absolutely no differences in explicit stereotypes between men and women: both men and women scored on average 0.66 (*SD* = .49) in explicit academic stereotype, producing an effect size of *d* = 1.35. The effect size from the implicit measure was *d* = 1.07, producing a difference in effect size between implicit and explicit measure of *d* = .28. Indicating a small to medium effect size (Cohen, 1977)

9. Discussion and Limitations

The initial motive for this study was to develop a measure for examining elementary school children´s implicit academic stereotypes, and to investigate whether elementary school children possess gendered implicit academic stereotypes. For this, an indirect measurement procedure was developed, named Measure of Children`s Implicit Academic Stereotypes (MeCIAS). The procedure was based the implicit association test (IAT), which is rooted in well established theory and research (e.g., Gawronski & Payne, 2010; Greenwald, et al, 1998; Greenwald et al., 2002).

The stimuli items in MeCIAS consists of boy and girl names, and pro- and counter academic characteristics. MeCIAS was designed to take account of children who not yet had automated their reading skills. This was approached through presenting the stimuli material both visually and auditory simultaneously in each trial. The names used in MeCIAS was based on the most common names in Norway, picked from Statistics Norway. Stimuli content which described pro- and counter-academic behaviour in MeCIAS was extracted via several interviews from the target population (elementary school children). These characteristics formed a total of five sub-facets: characteristics of pro-social, obedience, and intellect formed the stimuli in the pro-academic category; characteristics of anti-social and disobedience in the counter-academic category. The measurement procedure was initially meant to be implemented on elementary school children, first through seventh grade, from a school in Trondheim. Despite showing initial interest in participation, the responses from the principal of the school ceased suddenly and without apparent cause.

Due to withdrawal of the initial target sample, the measurement procedure was implemented on thirty Norwegian university students (age range of 20 – 25 years), with both sexes equally represented. Since this sample still is attending an educational institution, they are still relevant for the research topic regarding implicit academic stereotypes.

Regarding the test procedure and data preparation. Since a correct response was required for each trial of the procedure, and several different options regarding how to deal with erroneous responses was available, the author assessed three different data preparation procedures: Not recoding them, to recoding the responses transcending 3000ms to the 3000ms regardless of correct or incorrect response (following the data handling paradigm from the original IAT (1998), and recoding them according to the improved scoring algorithm (recoding erroneous responses to block average + 600 ms). No particular differences were observed on the final results as a function of data handling procedure, both the first and the second hypothesis was supported regardless of how the data was prepared. The author therefore conformed to the improved scoring algorithm since this data handling procedure has most empirical and theoretical relevance for contemporary use of the implicit association task paradigm

The main test results obtained using MeCIAS on this sample clearly states that there is an implicit academic stereotype favouring girls, supporting the first hypothesis of this study. Furthermore, a very strong sex of participant effect indicate that women holds a stronger implicit academic stereotype favouring girls than men, which supports the second hypothesis of this study. The author will also like to mention that male one subject (which transcended over 2 standard deviations from the average) weakened the average *D*`-scores for the males to such an extent that his exclusion would have amplified the D`-scores and reduced the standard deviation for the male group to such an extent that the male group would have attained statistical significance on the .05 level.

Regarding the test validity. Since the stimuli items were gathered from elementary school children, and the test was implemented on university students, one may question the validity of the measurement procedure and the results. However, the face validity check performed in the pilot study, convergence with the predicted hypothesis´, and with the explicit measure suggests that the validity of MeCIAS is good. The reliability (internal consistency) of MeCIAS is satisfactory in terms of established measurement standards, providing a sufficient Cronbach alpha to claim good internal consistency. Moreover, MeCIAS was easy to administer; no difficulties were reported from the participants. However, the author acknowledges that the MeCIAS-procedure may pose different challenges when it is implemented on elementary school children.

Considering the block order effect, MeCIAS has two different block order conditions: Block order condition one pairs boy names and pro-academic characteristics first, and girls and pro-academic after. Whereas block order condition two pairs girl names with pro-academic first, and boys and pro-academic after. As an attempt to control for order effects, the block order was randomized via coin toss prior to the procedure. Unfortunately, this randomization failed in distribute an equal amount of men and women in the two block order conditions. Combined with the strong sex of participant effect, this created an illusory block order effect in the preliminary analysis on response latency and *D`*-score. This was because there was a majority of men (with a weaker *D`*-score) in the second block order condition (girl names paired with pro-academic first), and a majority of women (with a stronger *D`*-score) in the first block order condition (boy names with pro-academic first). Acknowledging the strong sex of participant effect makes it salient how important it is to control for equal groups in block order conditions. This knowledge is advantageous for further research using MeCIAS: The author suggest pseudo-randomization to ensure that both sexes is represented equally in each block order condition.

Regarding the explicit measure. Both men and women report to an equal degree that they hold the explicit stereotype that girls are more associated with pro-academic behaviour. However, on the implicit level, the men revolt: men in the sample show to a lesser degree an implicit pro-academic stereotype for females than the women in the sample, suggesting that automatic biases (in-group favouritism) also attenuates men´s automatic response. Limitations of explicit measure: A majority of the subjects reported that they based their explicit stereotype solely on their self-monitored performance on the MeCIAS. This could have been avoided by asking the participants about their explicit stereotype prior to the indirect measurement. Additionally, further research should consider expanding the explicit measure to include more than one variable.

So, what should be kept in mind when using this measure in future research? The strong sex of participant effect makes it particularly important to ensure an equal distribution of the sexes in the two block order conditions. If this requirement is not met, illusory block order effects will most probably arise.

10. Conclusion

The MeCIAS procedure has shown to be a valid and reliable procedure, and future research should consider using it to for studies interested in children´s – as well as adults´ - implicit academic stereotypes. A natural next step for this measure could be to utilize it either cross sectionally - or preferably longitudinally - on a sample of elementary school children, to determine *when* these stereotypes are established in Norway. This paves the way for unravelling exactly *what* causes these stereotypes. With the ultimate intention to facilitate further development of the school regime to adjust to the knowledge acquired from this research.

11. References

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179-211. doi: http://dx.doi.org/10.1016/0749-5978(91)90020-T

Ambady, N., Shih, M., Kim, A., &, & Pittinsky, T. L. (2001). Stereotype susceptibility in children: Effects of identity activation on quantitative performance. *Psychological Science, 12*(5), 385-390.

Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., & Brown, J. (1999). When white men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology, 35*(1), 29-46.

Asch, S. E. (1952). *Social Psychology*. New Jersey: Prentice Hall.

Babad, E. Y., Inbar, J., & Rosenthal, R. (1982). Pygmalion, Galatea, and the Golem: Investigations of biased and unbiased teachers. *Journal of Educational Psychology, 74*(4), 459.

Bakken, A., Borg, E., Hegna, K., & Backe-Hansen, E. (2008). *Er det skolens skyld? En kunnskapsoversikt om skolens bidrag til kjønnsforskjeller i skoleprestasjoner*. Oslo: Norwegian social research.

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of personality and social psychology, 71*(2), 230-244.

Baron, A. S., & Banaji, M. R. (2006). The development of implicit attitudes evidence of race evaluations from ages 6 and 10 and adulthood. *Psychological Science, 17*(1), 53-58.

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychological bulletin, 117*(3), 497-529.

Bjorklund, D. F., & Kipp, K. (1996). Parental investment theory and gender differences in the evolution of inhibition mechanisms. *Psychological Bulletin, 120*(2), 163-188.

Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review, 6*(3), 242-261.

Blair, I. V., & Banaji, M. R. (1996). Automatic and controlled processes in stereotype priming. *Journal of personality and social psychology, 70*(6), 1142-1163.

Blair, I. V., Ma, J. E., & Lenton, A. P. (2001). Imagining stereotypes away: the moderation of implicit stereotypes through mental imagery. *Journal of personality and social psychology, 81*(5), 828-841.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The Concept of Validity. *Psychological Review, 111*(4), 1061-1071.

Brehm, J. W. (1966). *A theory of psychological reactance*. New York: Academic Press.

Bunge, M. (1974). *Treatise on Basic Philosophy, Vol. I Semantics I: Sense and Reference.* . Dordrecth-Boston: Reidel Publishing Co.

Burger, J. M. (2009). Replicating Milgram: Would people still obey today? *American Psychologist, 64*(1), 1-11.

Carlston, D. (2010). Models of implicit and explicit mental representation. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition*. Guilford Press.

Chen, M., & Bargh, J. A. (1997). Nonconscious behavioral confirmation processes: The self-fulfilling consequences of automatic stereotype activation. *Journal of Experimental Social Psychology, 33*(5), 541-560.

Cobb, J. A. (1972). Relationship of discrete classroom behaviors to fourth-grade academic achievement. *Journal of Educational Psychology, 63*(1), 74-80.

Cohen, J. (1977). *Statistical Power Analysis for the Behavioral Sciences* (2 ed.). New York: Academic Press.

Colman, A. M. (2009). *A Dictionary of Psychology*. Oxford University Press.

Crick, N. R., & Grotpeter, J. K. (1995). Relational aggression, gender, and social ‐ psychological adjustment. *Child development, 66*(3), 710-722.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334.

Crowne, D., & Marlowe, D. (1964). *The Approval Motive: Studies in Evaluative Dependence*. New York: Wiley.

De Houwer, J., & Moors, A. (2010). Implicit measures: Similarities and Differences. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications*. Guilford Press.

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009a). Implicit measures: A normative analysis and review. *Psychological Bulletin, 135*(3), 347-368.

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009b). Theoretical claims necessitate basic research: Reply to Gawronski, Lebel, Peters, and Banse (2009) and Nosek and Greenwald (2009). *Psychological Bulletin, 135*(3), 377-379.

Devine, P. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*(1), 5-18.

Eagly, A. H. (2009). The his and hers of prosocial behavior: an examination of the social psychology of gender. *American Psychologist, 64*(8), 644-658.

Edelman, G. M. (1992). *Bright air, brilliant fire: on the matter of the mind*. New York: Basic Books.

Edelman, G. M., & Tononi, G. (2000). *A universe of consciousness: How matter becomes imagination*. New York: Basic Books.

Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59*, 255-278.

Evans, J. S. B., & Over, D. E. (1996). *Rationality and reasoning*. Hove: Psychology Press.

Fazio, R. H., & Olson, M. A. (2003). Implicit Measures in Social Cognition Research: Their Meaning and Use. *Annual Review of Psychology, 54*(1), 297-327.

Furnham, A. (1986). Response bias, social desirability and dissimulation. *Personality and individual differences, 7*(3), 385-400.

Gawronski, B., & Payne, B. K. (2010). *Handbook of implicit social cognition: Measurement, theory, and applications*. Guilford Press.

Gawronski, B., & Sritharan, R. (2010). Formation, Change, and Contextualization of Mental Associations: Determinants and Principles of Variations in Implicit Measures. In B. Gawronski & B. K. Payne (Eds.), *Handbook of Implicit Social Cognition: Measurement, theory, and applications*. Guilford Press.

Green, K. D., Forehand, R., Beck, S. J., & Vosk, B. (1980). An assessment of the relationship among measures of children's social competence and children's academic achievement. *Child Development*, 1149-1156.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102*(1), 4-27.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring Individual Differences in Implicit Cognition: The Implicit Association Test. *Journal of Personality and Social Psychology, 74*(6), 1464-1480.
Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*(2), 197-216.
Hartley, B., & Sutton, R. (2013). A Stereotype Threat Account of Boys´ Academic Underachievement. *Child Development, 84*(5), 1716-1733.
Hayes, N. (2000). *Doing psychological research : gathering and analyzing data*. Buckingham: Open University Press.
Jones, L. V., & Thissen, D. (2007). A History and Overview of Psychometrics. *Handbook of Statistics, 26*, 1-27.
Kalton, G., & Schuman, H. (1982). The effect of the question on survey responses: A review. *Journal of the Royal Statistical Society. Series A (General), 145*(1), 42-73.
Latsch, M., & Hannover, B. (2014). Smart girls, dumb boys!? How the discourse on “failing boys” impacts performances and motivational goal orientation in German school students. *Social Psychology, 45*(2), 112-126.
Liddell, H. G., Scott, R., Jones, H. S., & McKenzie, R. (1996). *A Greek-English Lexicon*. Clarendon Press
Lowery, B. S., Hardin, C. D., & Sinclair, S. (2001). Social influence effects on automatic racial prejudice. *Journal of personality and social psychology, 81*(5), 842-855.
Macrae, C. N., Milne, A. B., & Bodenhausen, G. V. (1994). Stereotypes as energy-saving devices: a peek inside the cognitive toolbox. *Journal of Personality and Social Psychology, 66*(1), 37-47.
Masters, M. S., & Sanders, B. (1993). Is the gender difference in mental rotation disappearing? *Behavior genetics, 23*(4), 337-341.
McGarty, C., Yzerbyt, V. Y., & Spears, R. (2002). *Stereotypes as explanations: The formation of meaningful beliefs about social groups.* Cambrigde University Press.
Milgram, S. (1963). Behavioral Study of Obedience. *Journal of Abnormal and Social Psychology, 67*, 371–378.
Moors, A., Spruyt, A., & De Houwer, J. (2010). In search of a measure that qualifies as implicit. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications.* Guilford Press.
Most, S. B., Sorber, A. V., & Cunningham, J. G. (2007). Auditory Stroop reveals implicit gender associations in adults and children. *Journal of Experimental Social Psychology, 43*(2), 287-294.
Nielsen, H. B. (2009). New boys? A Nordic Perspective. In B. Budde & I. Mammes (Eds.), *Jungenforschung empirisch*. VS Verlag für Sozialwissenschaften.
Nosek, B. A., & Banaji, M. R. (2001). The go/no-go association task. *Social Cognition, 19*(6), 625-666.
Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of applied psychology, 88*(5), 879-903.
Reiss, A. L., Abrams, M. T., Singer, H. S., Ross, J. L., & Denckla, M. B. (1996). Brain development, gender and IQ in children: A volumetric imaging study. *Brain, 119*(5), 1763-1774.

Rispoli, M., O'Reilly, M., Lang, R., Machalicek, W., Davis, T., Lancioni, G., & Sigafoos, J. (2011). Effects of motivating operations on problem and academic behavior in classrooms. *Journal of applied behavior analysis, 44*(1), 187-192.

Rudman, L. A., Ashmore, R. D., & Gary, M. L. (2001). "Unlearning" automatic biases: the malleability of implicit prejudice and stereotypes. *Journal of personality and social psychology, 81*(5), 856-868.

Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology, 38*(2), 194-201.

Schmader, T., & Johns, M. (2003). Converging Evidence That Stereotype Threat Reduces Working Memory Capacity. *Journal of Personality and Social Psychology, 85*(3), 440-452.

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review, 115*(2), 336-356.

Shanab, M. E., & Yahya, K. A. (1977). A behavioral study of obedience in children. *Journal of Personality and Social Psychology, 35*(7), 530-536.

Shaw, P., Greenstein, D., Lerch, J., Clasen, L., Lenroot, R., Gogtay, N., . . . Giedd, J. (2006). Intellectual ability and cortical development in children and adolescents. *Nature, 440*(7084), 676-679.

Sinclair, L., & Kunda, Z. (1999). Reactions to a black professional: motivated inhibition and activation of conflicting stereotypes. *Journal of personality and social psychology, 77*(5), 885-904.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*(1), 4-28.

Stanovich, K. E. (1999). *Who is rational?: Studies of individual differences in reasoning.* New Jersey: Elrbaum.

Statistisk Sentralbyrå. (2014). Karakterer ved avsluttet grunnskole, 2014. Retrieved 31. october, 2014, from http://ssb.no/utdanning/statistikker/kargrs/aar/2014-10-30-content

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist, 52*(6), 613-629.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69*(5), 797-811.

Sternberg, R. J., Mio, J., & Sternberg, K. (2012). *Cognition* (6th ed.). Australia: Wadsworth Cengage Learning.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and social psychology review, 8*(3), 220-247.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of experimental psychology, 18*(6), 643-662.

United Nations Development Programme. (2013). Gender Inequality Index. Retrieved 25. october, 2014, from http://hdr.undp.org/en/content/table-4-gender-inequality-index.

Wentzel, K. R. (1993). Does being good make the grade? Social behavior and academic competence in middle school. *Journal of Educational Psychology, 85*(2), 357.

White, R. W. (1959). Motivation reconsidered: the concept of competence. *Psychological review, 66*(5).

Wickens, A. (2009). *Introduction to biopsychology* (3rd ed.). Harlow: Pearson/Prentice Hall.

Wilson, T. D., & Dunn, E. W. (2004). Self-knowledge: Its limits, value, and potential for improvement. *Annu. Rev. Psychol 55*, 493–518.

Appendix A:

All academic characteristics gathered from interviews

	Pro Academic	Counter Academic
1.Klasse	Lære	Slå
	Snill x 3	Si stygge ting
	Gjøre jobben sin	Slem
	Vennlig	Geipe
	Tegne	Lyve
	Aktiv	Hærverk
		Stjele
	N=8	N=8
2.Klasse	Snill	Slå
	Sosial	Klype
	Stille	Sparke
	Ansvarlig	Geipe
		Krangle
		Utestengning
	N=4	N=6
3.Klasse	Lydig	Frekk x 2
	Stå riktig i brannrekka	Lat
	Leke	Kaste søppel
	Vennlig	Slåss x 3
	Snill x 3	Hærverk
	Rolig	Slem
	Jobbe	
	Flink	
	Gjøre så godt man kan	
	Respektere andre	
	N=12	N=9
4.Klasse	Rekke opp hånda	Løpe rundt
	Snill	Bråke
	Sitte på plass	Ikke komme tilbake fra friminutt
	Lydig x 3	Bli stående når brannalarmen går
	Ryddig	
	N=7	N=4
5.Klasse	Humor	Mobbing
	Stille	Slåssing
	Arbeidsro	Utestenging
	Uteaktivitet	Krangling
	Være hyggelig	Baksnakking
	Lære	
	Morsom	

	N=7	N=5
6.Klasse	Snill x 2	Slem
	Smart x 3	Mobbing
	Glad x 2	Prating i timen
	Lydig	Bråkete x 2
	Rolig	Negativ
	Positiv	Ikke høre etter x 2
	Vennlig	Utestenging
	Inkluderende	Slåssing
	Ryddig	Kommentere x 2
	Omsorgsfull	Forskjellsbehandle
	Respektfull	(At lærer favoriserer)
		Himle med øynene
		Kødding
	15	15
7.klasse	Stille	Løpe rundt
	Lydig	Bråke x 3
	Flink	Banne
	Sitte rolig	Ikke jobbe
	Faglig aktiv	Uønsket aktivitet
	Respekt	
	Følge med x 2	
	Lese tydelig	
	Rekke opp hånda x 2	
	11	7

SUM	64	54

Appendix B:
Facets from the coding of academic characteristics

Pro-Social	*N*	**Anti-social**	*N*
Snill	12	Slåss	5
Vennlig	3	Utestenging	3
Respektere andre	3	Slem	3
Glad	2	Slå	2
Humor	1	Geipe	2
Være hyggelig	1	Krangle/Krangling	2
Morsom	1	Frekk	2
Positiv	1	Mobbing	2
Inkluderende	1	Forskjellsbehandle	1
Omsorgsfull	1	Himle med øynene	1
Sosial	1	Kødding	1
Leke	1	Baksnakking	1
Obedience	*N*	Negativ	1
Lydig	6	Stjele	1
Stille	3	Klype	1
Rekke opp hånda	3	Sparke	1
Følge med	2	Si stygge ting	1
Rolig	2	Lyve	1
Arbeidsro	1		
Sitte på plass	1	**Disobedience**	*N*
Sitte rolig	1	Bråke(te)	6
Gjøre så godt man kan	1	Løpe rundt	2
Gjøre jobben sin	1	Ikke høre etter	2
Stå riktig i brannrekka	1	Kommentere	2
Intellect	*N*	Hærverk	2
Smart	3	Banne	1
Lære	2	Ikke jobbe	1
Flink	2	Uønsket aktivitet	1
Ryddig	2	Ikke komme tilbake fra friminutt	1
Lese tydelig	1	Bli stående når brannalarmen ringer	1
Jobbe	1	Prate i timen	1
Faglig aktiv	1	Lat	1
Tegne	1	Kaste søppel rundt	1

Appendix C: Selected stimuli material marked in green. Yellow signifies reformulation to better fit in with the procedure.

Pro-Social
Snill
Vennlig
Respektere andre
Glad
Humor
Være hyggelig
Morsom
Positiv
Inkluderende
Omsorgsfull
Sosial
Leke

Obedience
Lydig
Stille
Rekke opp hånda
Følge med
Rolig
Arbeidsro
Sitte på plass
Sitte rolig
Gjøre så godt man kan
Gjøre jobben sin
Stå riktig i brannrekka

Intellect
Smart
Lære
Flink
Ryddig
Lese tydelig
Jobbe
Faglig aktiv
Tegne

Disobedience
Bråke(te)
Løpe rundt -> Urolig
Ikke høre etter
Kommentere
Hærverk
Banne
Ikke jobbe
Uønsket aktivitet
Ikke komme tilbake fra friminutt
Bli stående når brannalarmen ringer
Prate i timen
Lat
Kaste søppel rundt

Anti-social
Slåss
Utestenge
Slem
Slå
Geipe
Krangle/Krangling
Frekk
Mobbing
Forskjellsbehandle
Himle med øynene
Kødding
Baksnakking
Negativ
Stjele
Klype
Sparke
Si stygge ting
Lyve

Appendix D: Stimulus latency in milliseconds from practice trials as a function of block order

	Pro-male first		Pro-female first		
	Mean	SD	Mean	SD	Difference
Anders	634.72	247.72	703.38	296.44	-68.66
Anita	850.03	1371.72	738.11	290.23	111.92
Arne	745.62	290.62	705.33	292.42	40.29
Berit	651.74	145.34	843.89	476.35	-192.15
Bråke	821.72	413.74	825.78	265.97	-4.06
Elin	693.10	377.27	715.10	233.32	-22.00
Espen	704.33	248.66	625.54	173.50	78.80
Flink	729.04	232.59	724.80	216.18	4.24
Henrik	668.60	236.31	752.80	378.38	-84.20
Hilde	654.86	200.10	735.24	241.97	-80.38
Hærverk	832.50	383.04	740.26	185.45	92.24
Ida	685.19	269.05	709.55	367.91	-24.36
Ingrid	734.38	433.26	702.55	226.10	31.83
Lydig	846.80	366.38	790.96	255.20	55.84
Maria	878.29	1477.67	672.33	222.84	205.95
Marit	710.46	273.66	688.52	243.19	21.95
Ole	620.78	196.66	844.13	498.81	-223.35
Rolig	1078.39	1403.90	736.85	215.93	341.54
Rune	670.00	322.81	688.65	246.82	-18.65
Slem	813.23	380.62	779.33	252.00	33.90
Slåss	792.25	348.33	751.41	276.17	40.84
Smart	907.08	523.46	773.81	233.57	133.28
Snill	771.92	263.02	654.91	140.44	117.01
Stian	742.75	343.87	728.18	273.32	14.57
Terje	659.48	233.80	742.48	247.40	-83.01
Urolig	886.15	322.79	936.46	471.00	-50.31
Utestenge	988.80	510.31	980.27	371.08	8.53
Vennlig	722.56	275.64	720.27	247.88	2.29

Appendix E: Information given to the participants prior to the measurement procedure.

"Velkommen, takk for at du vil delta i dette studiet. Oppgaven deres er å sortere ordene som kommer på skjermen og i hodetelefonene til riktig side ved hjelp av E og I på tastaturet. I første del vil det komme navn på gutter og jenter. I andre delen vil det komme ord som beskriver god og dårlig oppførsel i skolen. I den tredje delen skal dere både sortere navn og ord til riktig side. Om noe er uklart under selve testen, så er det bare å spørre meg".