Dissertation
Petra Jelenec

seit 1558

Girls and the Leaky Math Pipeline

Implicit Math-Gender Stereotypes and Math Withdrawal in Female Adolescents and Women

Dissertation
zur Erlangung des akademischen Grades
doctor philosophiae (Dr. phil.)

vorgelegt dem Rat der Fakultät für Sozial- und
Verhaltenswissenschaften
der Friedrich-Schiller-Universität Jena

von Dipl.-Psych. Petra Jelenec
geboren am 07.04.1975 in Bergneustadt

Gutachter

1. Prof. Dr. M. C. Steffens

2. Prof. Dr. P. Noack

Tag des Kolloquiums: 26.05.2008

Acknowledgements

I am greatly indebted to Prof. Dr. Melanie C. Steffens for being a most valuable mentor who provided me with an enormous amount of support and encouragement. My special thanks are dedicated to my second supervisor Prof. Dr. Peter Noack.

I want to thank a large number of research students who helped with recruiting subjects and running experiments. Many thanks to Lars Buder, Kristine N. Goergens, Julia Anheuser, Julia Lichau, Yvonne Still, Ulrike Leich, Ulrike Zieger, and also Miriam Edelmann, Jennie Graef, Saskia Lucht, Katrin Meinerling, Sarah Nase, Mirco Peveling, Elena Redwanz, Björke Roos, and Kathrin Utz!

I would like to thank Janette Schult, Irena D. Ebert, and Rul von Stülpnagel for all helpful comments and advice in all stages of this project. Many thanks also to Nilanjana Dasgupta and Jean-Claude Croizet for detailed comments and suggestions on this work.

Many thanks go to all children, teachers and directors of all schools, especially Marlen Meyer, Christa-Maria Goergens, and Walter Ammon.

My very special thanks go to my parents, my aunt Ana Klemencic, and my friends Stefan Huber, Romy Lauche, and Helmut Hahn for their continuous encouragement of my work.

Table of Contents

List of Tables

List of Figures

1 Introduction

During the last decades women have caught up with men on post-secondary education, outnumbering male undergraduate and graduate students in several western countries (e.g., U.S. Department of Education, 2000). However, fewer women than men enter math-intensive fields like engineering or computer science, and this apparently self-imposed gender segregation in course selection can be observed already at high school. Math has been characterized as the “critical filter” in the job market, being a necessary precondition for access to higher paying or prestigious occupations (Sells, 1973). Thus, the gender disparities in math-intensive fields have raised concerns (see Steele, 2003).

Recent research has demonstrated the detrimental influence of stereotypic beliefs that women cannot excel in math on women’s math performance and math interest (e.g., Davies, Spencer, Quinn, & Gerhardstein, 2002; Spencer, Steele, & Quinn, 1999). Further, women can be susceptible to these math-gender stereotypes and reveal math-gender stereotyping independently of their personal endorsement of these stereotypes. This is the case for so-called implicit math-gender stereotypes (e.g., Nosek, Banaji, & Greenwald, 2002b). As the main career decisions are made during school years, the current research examines implicit math-gender stereotypes and their relations with math-related outcomes in children and adolescents.

In the following section, several theoretical approaches that have been developed to explain the gender gap in math-intensive fields will be presented. After describing explanations based on math performance and biological factors, a brief overview will be given of findings regarding math ability self-concepts and math-gender stereotypes. Then, the measurement of implicit stereotypes will be addressed, followed by findings regarding implicit math-gender stereotypes. Finally, an overview of the current research will be given.

1.1 The gender gap in math-intensive fields and its theoretical explanations

Still today, a large disparity between men and women in math or science course enrolments can be observed. While in the U.S. 31% of computer science master's degrees, 26% of physics master's degrees, and 22% of engineering master's degrees were earned by women in 2004, in Germany women received only 15.4% of all computer science diplomas, 16% of physics diplomas, and 15.1% of engineering diplomas (National Science Foundation, 2006; Ramm & Bargel, 2005). Similarly, German girls are underrepresented in advanced physics (less than about 20%) or computer science courses at high school (around 11%, according to personal communications with ministries of education of most German federal states). In the U.S., proportions of female students taking Advanced Placement exams in physics (25% to 35%) and computer science (15%) are only slightly higher, leading to a bad preparation for college majors in science and engineering (National Research Council, 2006). This phenomenon of losing a large percentage of girls and women is sometimes referred to as girls or women leaving the *math and science pipeline* (cf. American Association of University Women, 1999; Ivie & Ray, 2005).

There has been much debate on the reasons for this gender gap in participation in math-intensive fields. For instance, it has been supposed that women avoid math-intensive fields because of lower performance in math tasks. In a large meta-analysis including participants at various educational levels, Hyde, Fennema, and Lamon (1990a) reported a small overall male superiority ($d = .15$) in solving standardized math tests. However, as pro-male gender differences in standardized math tests tend to be at most moderately large and do not emerge consistently until the 9th or 10th grade, Hyde et al. (1990a) concluded that these gender differences are unlikely to account for the large gender gap in career choices. Interestingly, if math achievement is measured by math grades, gender differences predominantly favour female students. This is the case for junior high school through college math courses, contradicting results in standardized math tests (Kimball, 1989).

A biological explanation for the gender gap in outstanding math achievement proposed by Benbow (1988) has been severely criticized (e.g., Bleier, 1988; Eysenck, 1988; Friedman, 1989). Benbow observed a large male-to-female ratio of 12:1 among extremely gifted 7th graders taking the SAT-M (e.g., Benbow & Stanley, 1983). Among these students, high rates of left-handedness and auto-immune disorders were found. As more males than females are left-handed or have auto-immune disorders, she interpreted these coincidences as an evidence for a biological basis of the gender differences. However, only correlational instead of causal relationships were provided, and the male-to-female ratio in the sample of high achievers scoring higher than 700 on the SAT-M has decreased from 12:1 in 1983 to 3:1 in 2005 (Brody & Mills, 2005). Further, Hyde et al. (1990a) demonstrated in their meta-analysis that the pro-male gender differences in standardized math tests were larger in older studies published in 1973 and earlier (d = .31) than in newer studies published in 1974 and later (d = .14). As biologically based abilities cannot change in such a short period, these gender differences seem to be based largely on socializational factors (see also Bussey & Bandura, 1999).

As a crucial socialization model for the explanation of gender differences in achievement-related choices, Eccles and colleagues (e.g., Eccles, 1994; Parsons, Adler, & Meece, 1984) developed an extensive expectancy x value model. According to this model, task choices are influenced (i) by the subjective value of the task (i.e., deeming a task interesting, personally important or useful) and (ii) by the expectation of success. These two components are influenced by students' ability self-concepts and also by gender roles or gender stereotypes. In the following section, the role of math ability self-concepts and math-gender stereotypes in math-related outcomes will be described.

Beliefs about one's ability in a certain field are strongly linked to participation and performance in that area. This stresses the importance of ability self-concepts. According to a meta-analysis by Hansford & Hattie (1982), (math and verbal) ability self-concepts and performance were correlated, on average, with r = .42. In several longitudinal studies, a causal effect of math ability self-concept on subsequent math achievement could be detected (Marsh, 1990; Marsh & Yeung, 1997). Additionally, the math ability self-concept was a stronger predictor for choosing a specialization in math at senior high school in the following year than actual math grades (Köller, Daniels, Schnabel, & Baumert, 2000).

Given the relevance of ability self-concepts, gender differences in perceived math ability raise concern. Boys report consistently higher math ability beliefs than girls, and these gender differences are often larger than the gender gap in actual math performance, even occurring in the absence of any performance differences (Marsh, 1989; Rustemeyer & Jubel, 1996; Stipek & Gralinski, 1991; Tiedemann & Faber, 1995). Though gender differences in math self-concept seem to be particularly pronounced in adolescence (Hyde, Fennema, Ryan, & Frost, 1990b), they can already be observed in grade 3 or 4 at primary school. Marsh (1989) found higher math self-concepts in boys than in girls from grade 4 and higher verbal self-concepts in girls than in boys from grade 2. He concluded that "these effects are relatively stable from preadolescence to early adulthood" (p. 425).

Together with ability self-concepts, students' gender stereotypes and gender roles can influence achievement-related outcomes, and these gender stereotypes are conveyed to children and adolescents by their socializers. Teachers, for instance, often hold gender-stereotyped views of children's abilities (e.g., Keller, 2001; Rustemeyer, 1999). Investigating beliefs of university students pursuing teacher certification, participants believed that boys use a more autonomous and creative approach to math problems than girls (Rustemeyer, 1999). Keller (2001) showed that teachers' math-gender stereotypes were related to their high school students' math-gender stereotypes even after controlling for achievement-related variables. Parents are sources of probably even stronger stereotypic expectancy effects. Eccles and colleagues (e.g., Frome & Eccles, 1998; Jacobs & Eccles, 1992; Yee & Eccles, 1988) reported that (especially) mothers tended to underestimate girls' math and boys' English abilities and to overestimate girls' English and boys' math abilities even after accounting for actual performance. Further, mothers endorsing stronger gender stereotypes revealed stronger biases in the ability estimations of their children (Jacobs & Eccles, 1992). Importantly, children's ability self-concepts in math and English are related more strongly to their parents' beliefs than to their actual school grades (e.g., Eccles, Freedman-Doan, Frome, Jacobs, & Yoon, 2000). Altogether, teachers' and parents' math-gender stereotypes are expected to affect students more indirectly (e.g., mediated by parental beliefs about their child's ability).

Recent research has identified another mechanism through which math-gender stereotypes can influence math-related outcomes in women. Activating math-gender stereotypes by stereotype-related cues during a math test may impair women’s math performance and undermine their interest in math (e.g., Davies et al., 2002; Shih, Pittinsky, & Ambady, 1999; Spencer et al., 1999). Possible cues are, for example, describing a math test as usually producing gender differences or as diagnostic of math ability, or a minority status of women together with a male majority when taking the test (see Maass & Cadinu, 2003, for an overview). This decrease in performance can be explained by women’s concerns of confirming negative stereotypes about their group; this phenomenon is called *stereotype threat* (Steele & Aronson, 1995). Conversely, women’s math performance increases when math-gender stereotypes are described as being irrelevant for a particular testing situation, for example by characterizing a math test as gender-fair or as non-diagnostic of math-ability (Maass & Cadinu, 2003). Girls’ math performance is affected by stereotype threat already at high school (e.g., grade 10; see Keller & Dauenheimer, 2003), and even 5-7-year-old girls showed an impaired math performance after their gender identity was subtly activated (Ambady, Shih, Kim, & Pittinsky, 2001). In addition to stereotype activation effects in math tests, girls may refrain from math-intensive fields in order to be congruent with a female gender role orientation (see Zemore, Fiske, & Kim, 2000).

Although math-gender stereotypes can affect women’s and girls’ math performance and interest, students’ endorsement of these stereotypes is rather low. In a meta-analysis, Hyde et al. (1990b) reviewed self-report gender-stereotypes of male and female students aged 11-25 years. According to Hyde et al., mean ratings of both women and men “fall on the portion of the scale indicating a rejection of stereotypes” (p. 310). Regarding math-gender stereotypes in elementary school children, girls aged 6-10 years rated women as being less interested and less capable of doing math than men (Steele, 2003). However, elementary school children often denied math-gender stereotypes regarding boys and girls of their own age group (e.g., Ambady et al., 2001; Steele, 2003), or rated their own gender as being superior in math (Heyman & Legare, 2004). Thus, students of various age groups often reject math-gender stereotypes when being asked directly.

1.2 Implicit stereotypes

The majority of research on math-gender stereotypes has been conducted with self-report (or direct) measures capturing conscious (or deliberate or explicit) stereotyping. These measures depend on a person's willingness and ability to report the own beliefs accurately. However, people might distort their answers regarding math-gender stereotypes due to social desirability concerns or personal egalitarian standards. Further, people may not have a full introspective access to their stereotypes and attitudes so that direct measures cannot capture them (see Nisbett & Wilson, 1977).

However, deliberate stereotypes are not the only form of stereotyping. In the last decades, research in social cognition has demonstrated numerous instances of so-called implicit stereotypes (e.g., Blair, 2001; Greenwald & Banaji, 1995; Rudman, Greenwald, & McGhee, 2001). These stereotypes can be activated automatically when encountering stereotypic cues, and stereotype activations can occur without intention and control (Banaji & Hardin, 1996). Further, implicit stereotypes can influence behaviour without a person's awareness for that influence (Rudman & Borgida, 1995), or a person might be even unaware of holding the stereotype itself (see Wilson, Lindsey, & Schooler, 2000). Such implicit stereotypes can be viewed as associations of (social) groups with stereotypic attributes, for example *math-male*, and associations can differ in strength.

1.2.1 Measuring implicit stereotypes

During the last few decades, computerized techniques for measuring implicit stereotypes (or attitudes) have been developed. Most of them rely on response latencies, and as these tasks require fast reacting, they cannot be distorted easily even if participants guess the purpose of the task. Most prominent measures are priming tasks and Implicit Association Tests (IATs) (Fazio & Olson, 2003; Greenwald, McGhee, & Schwartz, 1998).

IATs have been developed to capture the strength of associations between two pairs of concepts, for instance, *math* vs. *language* with *male* vs. *female*, and they are based on the principle that it is easier to react with the same response to concepts that are strongly related than to concepts that are not related. Stimuli belonging to four concepts are usually presented in a randomized order, and participants have to classify them with two response options as fast as possible. The math-gender stereotype IAT comprises two tasks. Participants associating *math* with *male* and *language* with *female* should be faster in the task requiring one response for stimuli belonging to *math* or *male* and the other response for stimuli belonging to *female* or *language* than in the task where stimuli for *male* or *language* and stimuli for *female* or *math* should be classified together. The difference in average response latencies between these two tasks is called the IAT effect; larger latency differences with faster responses in the *math-male/language-female* task are supposed to indicate stronger stereotypic associations. IATs are assumed to capture automatically activated cognitions because due to time pressure, the response speed cannot be controlled as easily as responses to questionnaires (see Steffens, 2004).

When evaluating IATs as measurement tasks, their validity and reliability has to be taken into account. Already in the very first publication introducing the IAT, Greenwald et al. (1998) demonstrated the known-groups validity of IATs (see also Banse, Seise, & Zerbes, 2001; Kühnen et al., 2001; Teachman, Gregg, & Woody, 2001). IAT effects were also related to behavioural measures and showed incremental validity, especially with respect to hardly controllable or spontaneous aspects of behaviour (e.g., Hugenberg & Bodenhausen, 2003; McConnell & Leibold, 2001; Steffens & Schulze-Koenig, 2006). However, when interpreting IAT effects, one has to keep in mind that IATs deliver only relative stereotype or attitude measures. For example, the math-gender stereotype IAT reveals the combined strength of *math-male/language-female* associations compared to the strength of *math-female/language-male* associations. Separate associations of the two academic domains with gender (*math* with *male* vs. *female*; *language* with *male* vs. *female*) cannot be investigated with IATs, and large IAT effects may result due to strong associations of *math* with male and/or *language* with *female*. This need not be a serious flaw of IATs as a variety of judgements or decisions are made in the context of dichotomous alternatives, for example, gender or in- vs. out-group evaluations (see Nosek et al., 2002b). For disentangling stereotypic or evaluative associations, Go/ No-go Association Tasks (GNAT) can be applied (Nosek & Banaji, 2001).

As far as the reliability of IATs is concerned, internal consistencies are often high, exceeding Cronbach's α >.80, but test-retest correlations reach lower values, for example, an average correlation of $r = .56$ as reviewed by Schmukle & Egloff (2004) (see also Lane, Banaji, Nosek, & Greenwald, 2007; Steffens & Buchner, 2003). The large amount of systematic variance in IAT effects as it is expressed in internal consistencies does not depend solely on semantic or evaluative associations, but also on method-specific variance produced by cognitive processes required in IATs (e.g., task-switching, see Mierke & Klauer, 2003).

Material properties can influence IAT effects and threaten the usefulness of IATs as measures of concept associations by diminishing their internal validity. Steffens & Plewe (2001) demonstrated that IAT effects depended on associations of both concepts and individual stimuli. As a consequence, large IAT effects indicating, for example, strong associations *old-negative/young-positive* might be based not on a negative evaluation of the concept *old*, but simply on a preference for modern (Julia) over old-fashioned (Gertrude) names that are often used as stimuli for the concepts *old* vs. *young*. One way to circumvent this problem is using concept labels and their synonyms instead of category exemplars as stimuli. Such a Concept Association Task (CAT) revealed similar effect sizes as traditional IATs, and CAT effects showed somewhat higher correlations with other implicit and self-report measures than traditional IATs (Steffens, Kirschbaum, & Glados, in press). However, other material features like salience asymmetries in the concepts and the stimuli cannot be controlled easily and may limit the internal validity of IAT applications (Rothermund & Wentura, 2004). Taken together, IAT effects capture – although not purely – implicit evaluative or semantic associations. However, IAT effects of individuals should not be interpreted or used for diagnostic purposes, but IATs can serve as valuable research tools investigating groups of subjects. Further, one has to consider that IATs are merely a class of techniques, and every IAT application in the context of its sample has to show its reliability and validity.

1.2.2 Implicit math-gender stereotypes

Using IATs, implicit associations *math-male* and *liberal arts* (or *humanities*)-*female* could be detected both in college samples (Kiefer & Sekaquaptewa, 2007a, 2007b; Nosek et al., 2002b) and in a large internet sample with over 60,000 adults (Nosek, Banaji, &

Greenwald, 2002a). Women revealed similar degrees of implicit math-gender stereotypes as men (Nosek et al., 2002a; Nosek et al., 2002b), and these implicit math-gender stereotypes were related to math identification, math attitudes and math performance for both female and male college students (Nosek et al., 2002b). Implicit – but not explicit – math gender stereotypes were linked to less favorable math-related cognitions and performance for women, but not for men. Kiefer and Sekaquaptewa (2007b) extended this research line with a prospective study investigating the impact of women's gender identification and implicit math-gender stereotypes measured with IATs on math performance and career goals. Female university students who scored low in both gender identification and implicit math-gender stereotypes performed best in the final exam of a calculus course. Self-reported interest in pursuing math-related careers was higher for women with either low gender identification or low implicit math-gender stereotypes. Further, implicit math-gender stereotypes measured with IATs seem to moderate stereotype threat effects (Kiefer & Sekaquaptewa, 2007a). Describing a math test as non-diagnostic of math ability as opposed to diagnostic improved performance only for women with low implicit stereotypes, but not for women with strong (and therefore probably chronically accessible) stereotypes. In sum, implicit math-gender stereotypes could be demonstrated in men and women, and these stereotypes have revealed unique predictive power regarding math-related outcomes. Up to now, research on implicit math-gender stereotypes has been carried out with adult participants only. However, as main career decisions are made during school years, implicit math-related cognitions should be investigated in children and adolescents, as well.

1.3 Outline of the dissertation

In Chapter 2 (Study 1 and 2), the onset of implicit math-gender stereotypes in elementary school children and relations between implicit math-gender stereotypes and math-related outcomes in adolescents were investigated. Study 2 served as conceptual replication of the results obtained in Study 1, using paper-and-pencil IATs instead of computerized IATs as in Study 1.

In Chapter 3 (Study 3 and 4), implicit gender stereotypes regarding math and language were investigated separately with GNATs. Results obtained in a sample of adolescents (Study 3) should be replicated in an adult sample with university students (Study 4).

In Chapter 4 (Study 5), factors activating implicit *math-male* stereotyping in women were examined. A stereotypic vs. non-stereotypic math test description and the exposure to that test were the independent variables. Please note that Chapters 2-4 are set up as separate journal articles.

In Chapter 5, the present findings are summarized and discussed. In addition, remarks about practical implications and possible future research directions are made. Finally, some concluding remarks are presented in Chapter 6.

2 What's on a Girl's Mind? Implicit Math-Gender Stereotypes and Math Withdrawal in Female Adolescents

When Ruth Lawrence graduated in math at Oxford University in 1985 at the age of 13, her success received much attention in terms of media coverage (e.g., BBC News Archive, 1985). Still today, even less spectacular female math or science role models cannot be taken for granted. The present research investigated stereotypes as factors contributing to this gender gap in math-related careers. Math-gender stereotypes stressing the incompetence of women in math have a great impact on women by lowering their performance and interest in math (e.g., Davies et al., 2002; Spencer et al., 1999). As explicit measures often show a rejection of math-gender stereotypes (e.g., Hyde et al., 1990b), we applied implicit measures. Further, as crucial career decisions are made during school years, we examined implicit math-related cognitions in adolescents.

During the last decades women have caught up with men on post-secondary education, outnumbering male students (e.g., U.S. Department of Education, 2000). However, fewer women than men enter math-intensive fields like engineering or computer science, with percentages lower than one third in the U.S. and one sixth in Germany (National Science Foundation, 2006; Ramm & Bargel, 2005). Ability self-concepts in math offer a promising approach for understanding the math gender gap (e.g., Eccles, 1994). These self-concepts exert a causal influence on math achievement (Marsh & Yeung, 1997), and they can have a greater impact on subsequent course selections than math grades (Köller et al., 2000). Further, though boys' higher math self-concepts relative to girls' are particularly pronounced in adolescence, and they are much larger than actual performance differences (Hyde et al., 1990a; Hyde et al., 1990b), they can already be observed in grade 3 or 4 at elementary school (e.g., Marsh, 1989).

Contributing to ability self-concepts, students' gender stereotypes have been identified to affect achievement-related behaviour. Students are confronted with math-gender stereotypes at various occasions, for example stereotypic beliefs expressed by teachers or parents (e.g., Bhanot & Jovanovic, 2005; Jacobs & Eccles, 1992). Recent stereotype threat research has demonstrated the detrimental effect of math-gender stereotypes on women. Activating math-gender stereotypes by stereotype-related cues during a math test can impair women's math performance and undermine their interest in math (e.g., Davies et al., 2002; Shih et al., 1999; Spencer et al., 1999). Girls' math performance has been shown to be affected by stereotype threat already at high school (Keller & Dauenheimer, 2003) and even in 5-7-year-olds (Ambady et al., 2001).

In order to analyze the impact of math-gender stereotypes, it should be investigated to what extent students have internalized these stereotypes. When being asked directly, children and adolescents often disavow math-gender stereotypes (Ambady et al., 2001; Hyde et al., 1990b), and young children may reveal ingroup bias viewing their own gender as being more successful in math (Heyman & Legare, 2004). Though students often reject math-gender stereotypes, these stereotypes may affect them. Asked about math-gender stereotypes, even young students might distort their answers in order to hide their views on these socially sensitive topics. Despite counter-stereotypic self-reports, students might possess stereotypes of women being incompetent in mathematical fields. These stereotypes can be viewed as associations between gender and stereotypic attributes, for example, *math-male* or *language-female*, and associations may differ in strength (see also Dasgupta & Asgari, 2004). Stereotypic associations can be activated automatically without intention or control, and they may influence behaviour without the person's awareness of that specific impact—the person might be even unaware of holding the stereotype itself (see Greenwald & Banaji, 1995). These so-called implicit stereotypes are typically measured by computerized techniques like Implicit Association Tests (IATs) (Greenwald et al., 1998). IATs have shown good measurement properties in a large amount of studies. Internal consistencies of IATs are often high, exceeding Cronbach's α >.80 (see Lane et al., 2007; Steffens & Buchner, 2003). Further, IATs are often related to explicit measures and can predict behaviour, showing incremental validity particularly with respect to hardly controllable or spontaneous aspects of behaviour (e.g., Hugenberg & Bodenhausen, 2003; Steffens & Schulze-Koenig, 2006).

Nosek, Banaji, & Greenwald (2002b) assessed implicit math-related cognitions with IATs in college students. Whereas both men and women showed strong math-gender stereotypes (i.e., associations *math-male* and *arts-female*), women revealed more negative attitudes (i.e., associations *math-unpleasant* and *arts-pleasant*) towards math than men. Regarding the math (vs. arts) identity, women identified themselves with *arts (*i.e., associations *self-arts, other-math)*; men, on average, did not show any implicit identification with *math* or *arts*. Implicit—but not explicit—math-gender stereotypes were related to implicit and explicit math attitudes, math identity, and performance. Men with stronger implicit math-gender stereotypes showed more positive math attitudes, higher math identification, and performance; stronger math-gender stereotypes in women were related to lower math preferences, identity, and performance. Implicit math-gender stereotypes measured with IATs were also demonstrated in an internet sample comprising over 60,000 adults, with women and men showing similar levels of stereotyping (Nosek et al., 2002a). In a prospective study, stronger implicit math-gender stereotypes predicted worse math performance and lower interests in math-related careers in female college students (Kiefer & Sekaquaptewa, 2007b). Further, implicit math-gender stereotypes appear to moderate stereotype threat (Kiefer & Sekaquaptewa, 2007a).

Up to now, implicit math-gender stereotypes have been investigated in adults only. However, main career decisions are made during school years. In two studies, we assessed implicit math-related cognitions using IATs in children and adolescents. The youngest participants were 4th graders aged about 9 years; two older groups consisted of 7th and 9th graders aged about 13 and 15 years. First, we investigated whether implicit math-gender stereotypes can be shown already in elementary-school children. Second, it seemed important to test whether girls have acquired stronger stereotypes than boys. This might be the case because girls, but not boys, experience stereotype threat regarding math. Third, it should be assessed at what age children show an implicit identification with the verbal or the math domain (i.e., implicit math identity). In general, girls are expected to show a stronger implicit identification with language (or the respective school subject, in our case, German) vs. math than boys, with larger gender differences in adolescents than in younger children.

Relations of implicit math-gender stereotypes with other math-related cognitions and outcomes were investigated in a joint analysis of Studies 1 and 2. Stronger implicit math-gender stereotypes should be related to (i) a stronger identification with language relative to math in girls; (ii) a stronger enrolment preference for language compared to math; and (iii) better German as compared to math grades for girls (cf. Nosek et al., 2002b).

2.1 Study 1

Implicit gender identity, math-gender stereotypes, and math identity were investigated with IATs in a cross-sectional sample of 4th, 7th, and 9th graders. First, the gender identity IAT served as an indicator whether already 4th graders were able to complete simple IATs.[1] This IAT should differentiate clearly between boys and girls in all three age groups because understanding of the concept "gender" is fully established in middle or late elementary school (cf. Bussey & Bandura, 1999). Second, the onset of implicit math-gender stereotyping and implicit identification with either academic domain together with possible gender differences were examined.

Further, we assessed explicit gender stereotypes to compare them with their implicit counterparts. Explicit self-concepts and school grades in math and German were investigated to test whether our sample shows typical gender differences. Gender differences in math self-concepts and enrolment intentions were expected to favour boys whereas the opposite should be the case for German. Girls should outperform boys in German grades, but math grades should not favour boys (see Hannover, 1991; Kimball, 1989; Marsh, 1989).

---

[1] After Study 1 had been conducted, Baron and Banaji (2006) introduced their child IAT even suited for 6-year-olds. However, we used IATs with (simple) words as stimuli in order to keep our IATs constant across age groups.

2.1.1 Method

2.1.1.1 Participants

The initial sample comprised *N* = 147 participants attending various Western German elementary schools, secondary schools (intermediate school track, Realschule), and grammar schools (highest school track, Gymnasium). Permissions to conduct the study were granted by school directors and parents. Children and adolescents participated in the study voluntarily during regular school hours. Seven participants (4th graders: 3; 7th graders: 1; 9th graders: 3) with error rates exceeding 30% in one or more combined IAT tasks were removed from analysis. Altogether, 59 4th graders (mean age = 9 years 5 months; 32 girls, 27 boys), 39 7th graders (mean age = 12 years 10 months; 22 girls, 17 boys) and 42 9th graders (mean age = 15 years 0 months; 21 girls, 21 boys) were included in the analysis. Of the 7th- and 9th graders, 35 attended a secondary school, 46 a grammar school.

2.1.1.2 Materials

Implicit measures. The gender identity IAT, the math identity IAT, and the math-gender stereotype IAT were selected to be simple enough even for 4th graders and appropriate also for adolescents. Only two stimuli were used per concept (cf. McFarland & Crouch, 2002; Nosek, Greenwald, & Banaji, 2005) that were denotative rather than connotative (cf. Steffens et al., in press). Concept labels and stimuli can be obtained from Table A1 in Appendix 1. *German* was chosen as a concept label because students use *German* as a common term for their school subject (Study 2 used *language* instead).

Explicit ability self-concepts in math and German. To measure ability self-concept in math (German), participants rated their agreement to the statements “I like math (German)”, “I am good at math (German)”, and “I learn things quickly in math (German)” (cf. Marsh, 1989). All explicit ratings were made on 5-point scales, with higher values indicating a stronger agreement.

Enrolment intentions. In the 4th grade, enrolment intentions were measured by the students' agreement to "In high school I am going to choose many math (German) classes". Participants attending the 7th or 9th grade were asked to indicate their consent to "I would like to drop my math (German) classes". Additionally, 7th- and 9th graders from grammar schools had to indicate their consent to "I can imagine taking advanced math (German) classes for A-levels".

School grades. Children in the 4th grade were asked to indicate their latest class test and report grades in math, dictation, and composition and further their latest report grade in reading. Adolescents in the 7th or 9th grade were asked for their math and German grades in their latest class tests and their latest report.

Explicit gender stereotypes. First, participants were asked about their agreement to four statements referring to the giftedness of boys and girls in math or German, for example, "Boys are often talented for doing German". Two further items captured comparative gender stereotypes about math and German, using *girls* and *boys* as anchor points (cf. Nosek et al., 2002b). Translations of the explicit measures can be found in Appendix 1.

2.1.1.3 Procedure

After giving their informed consent, participants were tested either individually or in groups up to four by female experimenters. IATs were administered on portable Macintosh computers. Explicit measures were completed on paper-pencil questionnaires. To 4th graders, IAT instructions were explained orally and questionnaire items were read out. All participants started with the IATs, and the order of IATs was constant for all participants. The gender identity IAT was completed first, followed by the math identity IAT and the gender stereotype IAT. After the IATs, explicit measures were applied in the order described above. Finally, participants were debriefed and rewarded with small gifts. The study lasted about 25 minutes.

Response keys were Y (located where the Z is on English keyboards) for left and N for right responses. False reactions were indicated by a flashing "F!". Each IAT started with two practice tasks of 8 trials each. The 3rd and 5th IAT task were combined tasks comprising 2 practice trials followed by 48 to-be-analyzed trials each, with stimuli of all four categories of a given IAT presented in a random order. Combined tasks of the gender identity IAT (the first IAT) comprised 8 additional practice trials that were removed from analyses. The 4th IAT task served for practising the reversed classification and comprised 24 trials in order to minimize task order effects (cf. Nosek et al., 2005). While all combined tasks contain two item pairs, (e.g., *girls-self* and *boys-other)*, we refer to a combined task only with one item pair for the sake of abbreviation (e.g., *girls-self*). Task orders within IATs were balanced as follows. Girls beginning with *girls-self* in the gender identity IAT started with *German-self* in the math identity IAT and with *German-girls* in the gender stereotype IAT and vice versa. Similarly, boys who completed the *boys-self* task first then started with the *math-self* task and the *math-boys* task.

2.1.1.4 Design

Dependent variables were IAT effects in the gender identity, math-gender stereotype, and math identity IAT. Gender and grade (4th, 7th and 9th grade) were treated as independent variables. Given the control factor IAT task order a 2 x 3 x 2 between-subjects design emerged. Large gender differences in IAT effects with an effect size of $f = .50$ could be detected with an $\alpha = .05$ and a sample size of $N = 40$ within each grade with a power of $1 - \beta = .87$.

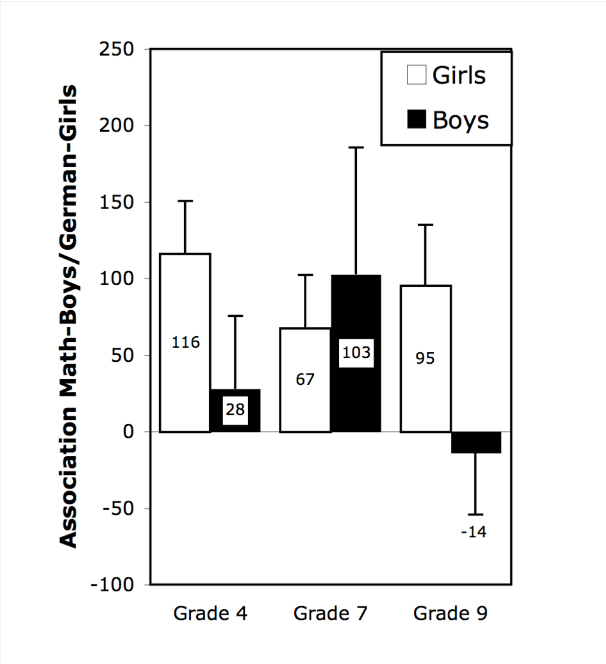
2.1.2 Results

Unless indicated differently, statistical tests in Study 1 and 2 were conducted with $\alpha = .05$. Therefore, individual *p*-values are not reported for statistically significant effects. The indicator of the effect size, R_p^2, is numerically identical to partial Eta squared and is an estimate of the proportion of explained variance after partialling out other factors in the design (Cohen, 1977). All IAT effects were computed similarly to IAT D effects (Greenwald, Nosek, & Banaji, 2003). Error reaction times were included in the analyses, but no values were recoded or error penalty used. For computing IAT effects in each IAT, the difference between each participant's average reaction times in the two combined tasks was divided by the participant's overall standard deviation of the response latencies in these tasks. Whereas these ipsatized effects were used in statistical analyses, figures show millisecond differences between the combined tasks.

2.1.2.1 IAT Analyses

Gender identity IAT. For reliability estimation, separate IAT effects were computed for trials with odd vs. even position numbers. Pearson correlations between IAT_{even} effects and IAT_{odd} effects were r = .69 for 4th graders and r = .84 for 7th and 9th graders, showing a satisfactory reliability. Positive difference scores indicate an association *self-girls*. Applying a known-groups approach to test whether IATs worked in all age groups, all participants were expected to show an association *self–own gender*. Indeed, a 2 (gender) x 3 (grade: 4th vs. 7th vs. 9th) x 2 (task order: *girls-self* first vs. *girls-other* first) ANOVA on IAT effects showed a main effect of gender, F (1,128) = 91.6, R^2_p = .42. Further, there was an interaction gender x grade, F (2,128) = 5.45, R^2_p = .08. Simple main effects of gender within grades revealed smaller gender differences in grade 4 (M_{girls} = 101 ms, M_{boys} = -59 ms), F (1,128) = 12.01, R^2_p = .09, than in grade 7 (M_{girls} = 135 ms, M_{boys} = -172 ms), F (1,128) = 36.79, R^2_p = .22, or in grade 9 (M_{girls} = 125 ms, M_{boys} = -113 ms), F (128) = 45.74, R^2_p = .26. Thus, the gender IAT worked in all age groups. The only other effect found (all other Fs < 1.68) was a main effect of task order, F (1,128) = 16.39, R^2_p = .11, indicating that IAT effects were biased in the direction of the task done first.

Gender stereotype IAT. Correlations between IAT_{even} effects and IAT_{odd} effects were r = .80 for 4th graders and r = .84 for 7th- and 9th graders, revealing satisfactory reliability. Larger IAT effects indicate stronger stereotypic associations *math-boys* and *German-girls*.



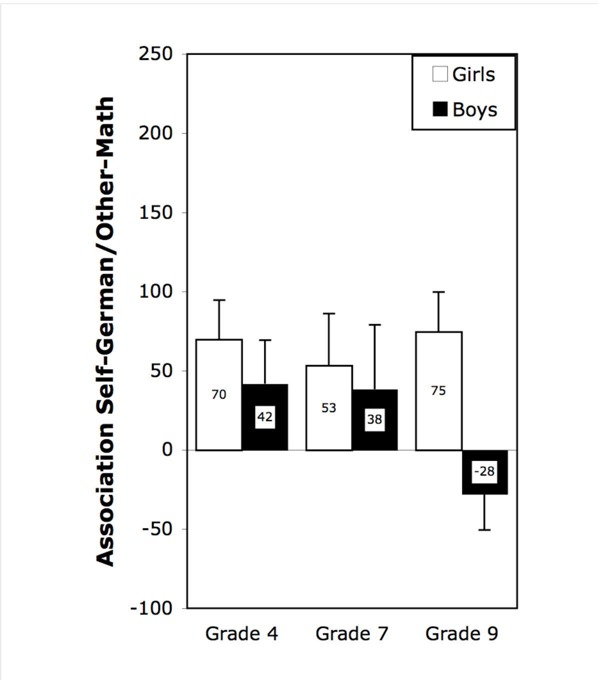


Figure 1: Response latency differences (in ms) in IATs, separately for gender and school grades in Study 1. Error bars reflect standard errors.

According to the upper part of Figure 1, these associations seem to be more pronounced in girls than in boys. Girls in grade 4 and 9, but not in grade 7, seem to reveal stronger stereotypic associations than boys of their age. This gender difference appears to be largest in grade 9 where boys, on average, did not show any stereotypic association. The 2 x 3 x 2 ANOVA revealed a large overall stereotypic association of *math-boys* and *German-girls*, F (1,128) = 25.17, R^2_p = .16, and a main effect of task order, F (1,128) = 22.75, R^2_p = .15 (all other Fs < 3.28). When we tentatively examined the observed gender difference in grade 9, the simple main effect of gender within grade was significant, F (1,128) = 4.86, R^2_p = .04. Six one-sample t-tests against 0 with an adjusted α = .008 to avoid an overall increase of α level (Bortz, 1999) did not reveal significant stereotypic associations in boys of any age group, all $|t|$s < 1.72. Female 7th graders did not reveal stereotypic associations when a rigorous α was used, t (21) = 2.07, p = .05, whereas girls in grade 4 and 9 showed significant stereotyping, t (31) = 4.15, R^2_p = .36, and t (20) = 2.96, R^2_p = .30. Thus, the math-gender stereotype IAT effect was driven by girls.

Math identity IAT. Correlations between IAT_{even} effects and IAT_{odd} effects were r = .52 for 4th graders and r = .65 for 7th- and 9th graders, revealing a rather low reliability. Larger IAT effects indicate stronger associations *self-German* and *other-math*. According to the lower part of Figure 1, these associations seem to be prevailing among girls in all grades, whereas boys, on average, show less pronounced associations of *self* with either domain. Gender differences appear largest in grade 9. The 2 x 3 x 2 ANOVA revealed the expected main effect of gender, F (1,128) = 17.16, R^2_p = .12, and a main effect of task order, F (1,128) = 53.39, R^2_p = .29. Girls showed stronger associations *self-German* than boys. Additionally, we found an interaction gender x grade, F (2,128) = 3.11, R^2_p = .05 (all other Fs < 1.81). Simple main effects of gender within grades revealed a gender difference only in grade 9, F (1,128) = 17.43, R^2_p = .12. Six one-sample t-tests against 0 (α = .008) showed that boys, on average, did not show significant IAT effects, all $|t|$s < 1.61. Girls showed an association *self-German*, t (31) = 2.92, R^2_p = .22, for grade 4, t (21) = 2.76, R^2_p = .27 for grade 7 (one-tailed, p = .006), and t (20) = 3.65, R^2_p = .40, for grade 9.

2.1.2.2 Explicit measures

Explicit math-gender stereotypes. First, the four ability ratings were combined in one index. Ratings of boys' giftedness for German were subtracted from ratings of girls' giftedness for German, and ratings of girls' giftedness for math were subtracted from ratings of boys' giftedness for math. These differences were averaged resulting in a gender stereotype score comparable to the IAT effect. Second, another index combining the two items measuring math and German gender stereotypes with *boys* and *girls* as anchor points was formed. As these two indexes were sufficiently correlated, $r = .48$, they were averaged. Means are displayed in Table 1. In contrast to implicit stereotypes, explicit gender stereotypes were comparable for boys and girls of all grades. A 2 (gender) x 3 (grade) ANOVA on the averaged stereotype index revealed no effects (all Fs < 1). As six one-sample t-tests against 0 ($\alpha = .008$) revealed, girls and boys in all grades showed significant gender stereotypes, all *ts* (one-tailed) ≥ 2.61 with $R^2_p s \geq .25$.

Table 1 Mean Explicit Math-Gender Stereotypes, Math and German Ability Self-Concepts, and Perceived Math-Gender Stereotypes (Study 1 and 2).

			Math-gender stereotypes	Math ability	German ability	Perceived math-gender stereotypes
Study 1	Grade 4	*Boys*	0.92 (0.94)	4.19 (0.86)	3.32 (1.02)	
		Girls	0.71 (0.85)	3.53 (0.95)	3.61 (1.00)	
	Grade 7	*Boys*	0.87 (1.02)	3.61 (0.69)	2.82 (1.04)	
		Girls	0.76 (1.04)	2.92 (1.18)	3.21 (1.06)	
	Grade 9	*Boys*	0.80 (1.40)	3.33 (0.86)	3.02 (0.62)	
		Girls	0.88 (0.90)	3.08 (0.89)	3.32 (0.95)	
Study 2	Grade 7	*Boys*	0.95 (1.04)	3.57 (0.98)	3.15 (0.92)	1.44 (1.72)
		Girls	0.63 (0.98)	2.99 (0.97)	3.22 (0.79)	1.81 (1.39)
	Grade 9	*Boys*	1.21 (1.31)	3.37 (1.15)	2.95 (0.96)	2.28 (1.48)
		Girls	1.22 (1.08)	2.96 (1.00)	3.26 (0.87)	2.81 (1.40)

Note. Higher values in the math and German ability self-concept scales indicate higher ability ratings with possible values between 1 and 5. Higher values in explicit stereotypes represent stronger math-gender stereotypes with possible values between -4 and 4. Standard deviations are in parentheses.

Explicit ability self-concepts. Both the math and the German ability self-concept scale showed a good internal consistency, Cronbach's $\alpha = .84$. Replicating previous findings (cf. Marsh, 1989), we expected, and found, gender differences in ability self-concepts, with a 2 x 3 ANOVA showing a higher math self-concept in boys than in girls, $F\ (1,133) = 11.03$, $R^2_p = .08$. Additionally, older participants showed less favourable math ability ratings than younger participants, $F\ (2,133) = 7.74$, $R^2_p = .10$ (all other Fs < 1). Also, girls showed a more favourable German self-concept than boys, F (1,132) = 3.76, R^2_p = .03 (all other Fs < 2.72).

Enrolment intentions. Enrolment intentions for math classes revealed similar gender differences as ability self-concepts. A 2 (gender) x 2 (grade: 7^{th} vs. 9^{th}) ANOVA on the intention to drop math courses showed a main effect of gender, F (1,76) = 4.06, R^2_p = .05 (all other Fs < 1.96), with girls having stronger intentions to drop math. No other gender differences were found. The same ANOVA on the intention to drop German courses showed only a main effect of grade, F (1,76) = 4.52, R^2_p = .06 (all other Fs < 1), with 9^{th} graders being less prone to drop German. For grammar school students who were additionally asked about taking advanced math and German courses, no gender differences were found, either (all Fs < 1.39). Further, there were no gender effects on future enrolment intentions in high school regarding German or math courses in 4^{th} graders (all Fs < 2.03).

School grades. The latest class test grades of 4^{th} graders in dictation and composition were averaged to form the index for their latest German class test, whereas the latest report grades in dictation, composition and reading were combined as an index for the latest German report grade. For all participants, the latest report grade and class test grade were averaged separately for German and math as they were highly related, r = .57 for German and r = .73 for math. Our expectations of better German grades in girls, but no gender differences in math grades were confirmed. A 2 (gender) x 3 (grade) ANOVA on German grades revealed better grades for girls than for boys, F (1,131) = 8.45, R^2_p = .06. Additionally, a main effect of grades emerged, with the youngest children having better grades than older children, F (2,131) = 9.80, R^2_p = .13 (all other Fs < 1). The same ANOVA on math grades did not reveal a gender difference (F < 1), but only a main effect of grade, F (2,131) = 13.16, R^2_p = .17, with 4^{th} graders receiving better grades (all other Fs < 1.70).

In sum, explicit ability self-concepts were quite traditional, with boys showing higher self-concepts and also higher enrolment intentions for math than girls, and girls revealing higher German self-concepts than boys. Girls achieved better German grades than boys, but boys did not outperform girls on math grades.

2.1.3 Discussion

In a nutshell, girls revealed implicit math-gender stereotypes in each grade, whereas boys did not show implicit stereotypes in any grade. Gender differences in implicit math-gender stereotyping could be found in grade 9. Girls demonstrated the association *self-German* already in grade 4, and a gender difference with girls having stronger *self-German* associations could be observed in grade 9. Boys did not show an association of *self* with *math* or *German* at any age. Further, the expected association of *self-own gender* in all participant groups showed that children were able to deal with IATs. Therefore, even in a sample of 9-year-olds IATs with a limited range of simple words as stimuli can be used.

In the math identity IAT, girls showed an implicit affinity to German already in grade 4. The gender difference became significant in grade 9 with girls showing a *self-German* association and boys showing no association with either academic domain. Nosek et al. (2002b) obtained similar results with a *math* vs. *arts* identity IAT in adults. Gender differences in implicit math identity thus seem to develop during puberty and can still be found in adulthood. Thus, our results are in line with the finding that girls refrain from math particularly when reaching puberty, with their self-concepts and interests becoming more gender-specific (Hannover, 1991; Hyde et al., 1990b).

Most importantly, implicit math-gender stereotypes were already found in female 4th graders, and in girls attending the other grades, but not in boys. Explicit stereotype measures did not mirror these gender differences. Further, previous studies investigating implicit math-gender stereotypes in adults did not find gender differences, either (Nosek et al., 2002a; 2002b). Stronger implicit gender stereotypes in girls are plausible, for example as a by-product of repeated stereotype threat experiences. However, some caution is required as these implicit stereotypes might depict a particularly strong stereotype activation during this study. The experimenters observed that some boys were enthusiastic about doing a computerized task whereas some girls made timid remarks about computers at the beginning of the study. Neither boys nor girls had any problems accomplishing the IATs. Nevertheless, gender stereotypes may have been particularly salient for girls, leading to a larger IAT effect in the gender stereotype IAT. This stereotype activation in girls may have been facilitated by the activation of the gender identity in the practice IAT and also when indicating the own gender before starting the first IAT. To rule out this activation explanation, Study 2 was a replication in a more gender-neutral setting.

2.2 Study 2

The main aim of Study 2 was to replicate the findings of Study 1 in a more gender-neutral setting, avoiding both computerized tasks and a gender-related practice IAT. We assessed whether girls would again reveal stronger implicit math-gender stereotypes than boys. Given a replication, a situation-based explanation of this gender difference would be ruled out. Further, girls were expected to show a stronger self-language association than boys.

The most important modifications of the procedure were (i) using a paper-and-pencil IAT (see Teachman, Gapinski, Brownell, Rawlins, & Jeyaram, 2003); (ii) the gender-neutral practice IAT assessed the associations between trees vs. mushrooms and big vs. small; and (iii) the experimenters stressed in their oral instructions that no ability tests would be accomplished in order to avoid concerns about math tests in girls. As a minor change, the concept label *German* was replaced by language, because the term *German* might activate not only representations of the school subject, but also of German nationality. Further, participants had to indicate their gender at the end of the study to avoid subtle gender priming, and the order of the math-gender stereotype IAT and the math identity IAT was counterbalanced.

2.2.1 Method

2.2.1.1 Participants

Data of $N = 430$ participants attending the 7^{th}- or 9^{th}-grade of various Western German grammar schools were collected. Permissions to conduct the study were granted by school directors and parents. The adolescents participated in the study voluntarily during regular school hours. Participants were excluded from analyses if they had higher error rates than 35% in at least one IAT sheet or if they had completely finished at least one such IAT sheet. After eliminating data of 17 7^{th} graders and 17 9^{th} graders, IAT effects of the gender stereotype and the math identity IAT were checked for outliers. One additional participant with a math identity IAT effect 3 SD below the mean was excluded. Altogether, data of 186 7th graders (mean age = 13 years 0 months; 102 girls, 85 boys) and 209 9th graders (mean age = 15 years 0 months; 119 girls, 90 boys) were included in the analyses.

2.2.1.2 Materials

Implicit measures. Each IAT consisted of four sheets, two sheets for each combined task. A sheet contained two columns of 35 items each, and concept labels were printed in bold on the top of a column. In every column, stimuli appeared in a different random order. Participants did not perform additional practice tasks. Concept labels and stimuli can be found in Table A1 in Appendix 1.

Explicit measures. Explicit measures regarding ability self-concepts, enrolment intentions, school grades, and gender stereotypes were identical to those employed for 7th and 9th graders in Study 1. Additionally, one item pair measuring math-gender stereotypes more subtly was used. In this item pair with *boys* and *girls* as anchor points, participants had to estimate to what extent most other people, in general, hold gender stereotypes regarding German and math (i.e., perceived stereotypes). Translations of the explicit measures can be found in Appendix 1.

2.2.1.3 Procedure

After giving their informed consent, all students of a class participated simultaneously in the study. Female experimenters provided oral instructions and handed out the booklets containing IATs and questionnaires. Participants were given 30 s to classify as many items as possible on an IAT sheet without skipping items or correcting mistakes. For example, in the stereotype-congruent task of the gender stereotype IAT, participants ticked the left side for stimuli belonging to *boys* or *math* and the right side for stimuli belonging to *girls* or *language*. Participants were asked to make small ticks instead of crosses to avoid spending too much time per item. First, the practice IAT was completed, and all participants started with the *trees-big/mushrooms-small* task, followed by *trees-small/mushrooms-big*. The math-gender stereotype IAT and the math identity IAT followed in counterbalanced order. A distractor task was used after the first critical IAT in order to prevent carry-over effects. This task consisted of a 2-minute visual search task in which differences between several similar drawings should be detected. Task order was counterbalanced for boys and girls, as in Study 1. After the IATs, participants completed the explicit measures in the order described above. Finally, participants were thanked and debriefed. The study lasted about 30 minutes.

2.2.1.4 Design

Dependent variables were IAT effects in the math-gender stereotype IAT and the math identity IAT. School grade (7^{th} vs. 9^{th}) and gender were treated as independent variables. Task order and IAT order as control factors yielded a 2 x 2 x 2 x 2 between-subjects design. Medium-sized gender differences in IAT effects with an effect size of $f = .25$ could be detected with an $\alpha = .05$ and a sample size of $N = 170$ within each grade with a power of $1 - \beta = .90$.

2.2.2 Results

To compute IAT effects, we first determined the number of correctly classified items on each IAT sheet. Second, two difference scores were computed for each IAT based on the first or second sheets of an IAT task, respectively. Third, each difference score was divided by the constituent with the higher value in order to control for participants' individual speed. Correlations of these two single IAT effects were used for reliability estimation, and the final IAT effect was computed by averaging these two values. Whereas these IAT effects were used for statistical analyses, differences of correctly classified items in the IAT tasks (per 30 s) are depicted in Figure 2.

The practice IAT worked. Boys and girls of both grades showed large associations of *trees-big/mushrooms–small*. All four one-sample *t*-tests against 0 with an adjusted $\alpha = .0125$ reached significance, *ts* between $t\ (101) = 16.08$, $R^2_p = .72$, and $t\ (84) = 18.52$, $R^2_p = .80$.

2.2.2.1 IAT Analyses

Gender stereotype IAT. IAT effects obtained from the first and second sheets of the combined IAT tasks were correlated with a satisfactory $r = .69$. Positive IAT effects indicate associations *math-boys* and *language-girls*. As can be seen in Figure 2, in line with our expectation, girls seem to show stronger implicit math-gender stereotypes than boys. In fact, only girls reacted faster in the *math-boys* task than in the *math-girls* task. Further, implicit stereotypes seem to be stronger in female 9th- than 7th graders. A 2 (gender) x 2 (grade: 7th vs. 9th) x 2 (task order) x 2 (IAT order) ANOVA on IAT effects revealed an overall stereotypic association of math-boys and language-girls, F (1,380) = 11.65, R^2_p = .03, and confirmed the expected gender difference, F (1,380) = 13.89, R^2_p = .04. Further, participants starting with the *math-boys* task showed larger IAT effects, F (1,380) = 28.70, R^2_p = .07. The only other effect was an interaction task order x IAT order, F (1,380) = 6.24, R^2_p = .02 (all other Fs < 3.69). Simple main effects of gender within grades confirmed stronger stereotypic associations in girls in grade 7, F (1,380) = 5.68, R^2_p = .02, and in grade 9, F (1,380) =8.39, R^2_p = .02. Four one-sample t-tests against 0 with α = .0125 showed significant IAT effects only for female 9th graders, t (118) = 5.61, R^2_p = .21, and female 7th graders, t (101) = 2.43 (p = .02, one-tailed), R^2_p = .06. Thus, the math-gender stereotype IAT effect was driven by girls.

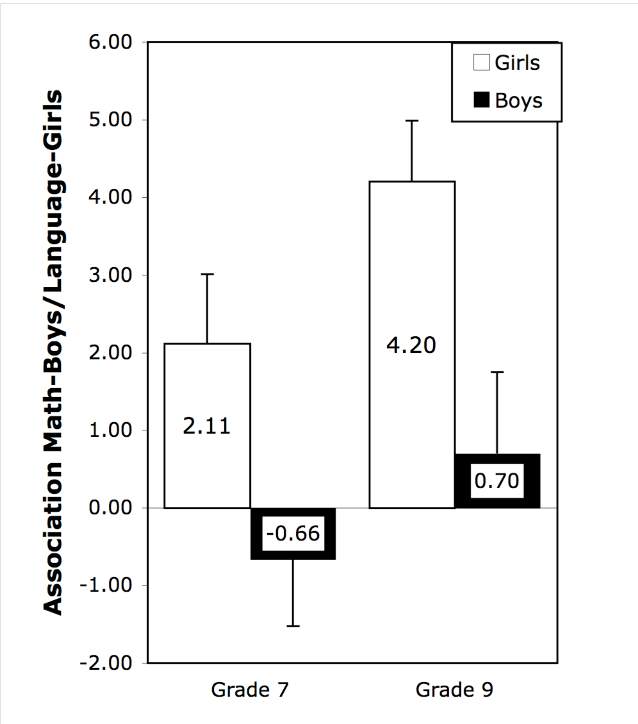


Figure 2: Differences of numbers of correctly classified items in the two tasks of the math-gender stereotype IAT (per 30 s), separately for gender and school grades in Study 2. Error bars reflect standard errors.

Math identity IAT. IAT effects obtained from the first and second sheets of the combined IAT tasks were correlated significantly, $r = .35$, but yielded insufficient reliability. Neither participant group revealed, on average, any associations, with no effects in four one-sample t-tests against 0 ($\alpha = .0125$), all $|t|$s < 2.10, and a 2 x 2 x 2 x 2 ANOVA, all Fs < 2.01. However, despite missing associations on the group level, IAT effects revealed interindividual variance and were included in the correlational analyses.

2.2.2.2 Explicit measures

Explicit gender stereotypes. As in Study 1, two stereotype indices were computed. As they were sufficiently correlated, $r = .60$, a mean stereotype score was computed (see Table 1). A 2 (gender) x 2 (grade: 7^{th} vs. 9^{th}) ANOVA yielded only a main effect of grade, with stronger stereotypes in grade 9, F (1,392) = 14.46, $R^2_p = .04$ (all other Fs < 2.17). Again, four one-sample t-tests against 0 ($\alpha = .0125$) revealed significant stereotyping for boys and girls in all grades, all ts > 6.48 with $R^2_p s \geq .29$.

Further, participants estimated other people's stereotypes regarding math and German. Again, higher values represent stronger stereotypes (see Table 1). The 2 x 2 ANOVA yielded both a main effect of grade, F (1,392) = 36.91, $R^2_p = .09$, and a main effect of gender, F (1,392) = 9.02, $R^2_p = .02$ (all other Fs < 1). Both 9^{th} graders and girls perceived stronger stereotypes in their environment. Four one-sample t-tests against 0 ($\alpha = .0125$) revealed significant stereotyping in all participant groups, all ts > 7.69 with $R^2_p s \geq .41$.

Explicit ability self-concepts. Internal consistencies were Cronbach's $\alpha = .81$ for the math self-concept scale and Cronbach's $\alpha = .79$ for the German self-concept scale. Means are presented in Table 1. As expected, a 2 x 2 ANOVA on the math self-concept revealed a main effect of gender, with boys reporting a higher math ability self-concept than girls, F (1,392) = 22.33, $R^2_p = .05$ (all other Fs < 1.28). Girls showed a higher German ability self-concept than boys, F (1,392) = 4.58, $R^2_p = .01$ (all other Fs < 1.92).

Enrolment intentions. Intentions to drop math or German courses were strongly negatively related to the intentions to choose the subject as an advanced course, $r = -.62$ (N = 396) for math and $r = -.51$ (N = 396) for German. After recoding, these two items were combined for math and German separately. The 2 x 2 ANOVA on the math enrolment index showed higher enrolment intentions in boys than in girls, F (1,392) = 28.18, R^2_p = .07 (all other Fs < 1). The same ANOVA on the German enrolment intention index revealed higher enrolment intentions in girls than boys, F (1,392) = 10.77, R^2_p = .03 (all other Fs < 1).

School grades. As in Study 1, the latest class test and report grade were averaged separately for math and German as these grades were correlated, r = .62 for math and r = .53 for German. A 2 x 2 ANOVA on German grades revealed that girls earned better German grades than boys, F (1,389) = 18.62, R^2_p = .05 (all other Fs < 3.38). No gender differences could be observed in math grades (all Fs < 1.25).

2.2.3 Summary of findings

Overall, in Study 2, female 7th- and 9th graders again showed stronger implicit math-gender stereotypes than their male classmates. Boys attending grade 7 or 9 did not reveal any stereotypic associations in the paper-and-pencil IAT whereas girls showed large stereotypic effects. Thus, stronger math-gender stereotyping in girls than in boys seem to be a robust finding as it was replicated under more gender-neutral conditions and with modifications of the measurement procedure. In the math identity IAT, neither boys nor girls showed, on average, any effects, perhaps due to the paper-and-pencil IAT's lower sensitivity.

As in Study 1, both boys and girls reported math-gender stereotypes, however, with older participants revealing stronger stereotypes. Interestingly, girls perceived stronger stereotyping in their environment than boys. Self-concept measures and school grades again revealed common gender differences. Girls reported higher ability self-concepts and enrolment intentions in German than boys and also received better German grades than boys. Though boys showed higher ability self-concepts and higher enrolment intentions in math than girls, boys and girls did not differ in their math grades.

2.3 Relations between implicit math-gender stereotypes and math-related outcomes

For girls, stronger implicit math-gender stereotypes should be related to a stronger identification with language vs. math, to higher enrolment intentions for language vs. math classes, and to better German vs. math grades (cf. Nosek et al., 2002b). Additionally, we assessed whether implicit math-gender stereotypes show incremental validity in predicting these outcomes when included in regression analyses together with explicit stereotypes. Further, we tested whether relations of implicit gender stereotypes with other math-related factors are stronger for girls than for boys.

Data preparation. In order to maximize statistical power, data sets of both studies were combined. The 4th graders in Study 1 were too few to be included in the combined data set. Within each of the two data sets (Study 1 vs. Study 2), z-values of IAT effects in the gender stereotype and math identity IAT were calculated separately for participant groups starting with either task order. In a second step, data sets were merged.

Table 2: Correlations between Implicit and Explicit Measures, Separately for Boys and Girls. Data of Study 1 and Study 2 are Combined.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) Implicit stereotypes		.25	.17	.16	.17	.16	.22
(2) Implicit math identity	(-.01)		.13	.20	(.04)	.15	(.06)
(3) Explicit math identity	(-.08)	(.08)		.29	(.11)	.78	.68
(4) Explicit stereotypes	.24	(-.09)	-.28		.33	.29	.25
(5) Perceived stereotypes	.33	-.16	(-.01)	.37		(.10)	(.09)
(6) Enrolment preferences	(-.12)	.19	.74	-.36	(-.15)		.63
(7) Grade differences	-.15	(.01)	.65	-.19	(-.06)	.51	

Note. Correlations are controlling for source (Study 1 vs. 2) and task order within IATs. Grade differences reflect differences between mean German and math grades, with higher values indicating better grades in German than in math. Values above the diagonal refer to girls, values beneath the diagonal to boys. Numbers in parentheses: Correlation is not statistically significant.

2.3.1 Implicit-explicit relations

Math-gender stereotypes. For boys and girls taken together, implicit gender stereotypes were correlated with explicit gender stereotypes ($r = .19$, $N = 476$) and with perceived stereotypes ($r = .27$, $N = 396$; Study 2 only). The size of these relations is in the expected order of magnitude (Hofmann, Gawronski, Gschwender, Le, & Schmitt, 2005; Lane et al., 2007). Separate correlations for boys and girls can be obtained from Table 2.

Math identity. First, an explicit math identity index was computed by subtracting the math self-concept from the German self-concept. Higher values indicate a higher self-concept in German vs. math. For boys and girls taken together, implicit and explicit math identity measures were significantly, but low correlated ($r = .13$, $N = 476$), though math identity should not be a socially sensitive construct.

2.3.2 Relations of math-gender stereotypes with math identity

Implicit math-gender stereotypes were related to implicit and explicit math identity measures only for girls, not for boys (see Table 2). For girls, stronger stereotyping was linked to a stronger identification with language compared to math. Not only implicit, but also explicit stereotypes were related to implicit math identity for girls and to explicit math identity for boys and girls.

In a second step, we investigated whether implicit gender stereotypes predict unique variance in math identity.[2] Two independent hierarchical regressions were carried out with implicit (regression 1) and explicit (regression 2) math identity as criteria and explicit and implicit gender stereotypes as predictors. Separate analyses were conducted for boys and girls. First, the explicit gender stereotype was entered; second, the implicit gender stereotype was added. Predicting both the implicit and the explicit math identity in girls, both implicit and explicit gender stereotypes showed unique predictive power (see Table 3). For boys, only explicit gender stereotypes predicted explicit math identity.

---

[2] Implicit math identity showed only low relations to achievement and enrolment preferences. This might be due to the rather poor measurement properties of the paper-pencil math identity IAT. Implicit math identity was not related to school grades and only barely related to enrolment preferences for boys, $r = .19$, and girls, $r = .15$. Hierarchical regressions with enrolment preferences as criterion and explicit and implicit math identity as predictors revealed unique predictive power for implicit math identity only for boys, $\beta = .13$, $t = 2.78$, not for girls.

Table 3: Beta Weights From Hierarchical Regressions Predicting Math-Related Outcomes, Separately for Boys and Girls (Study 1 and 2).

	Dependent Variables	Step 1		Step 2		
		Explicit gender stereotypes	Adj. R^2	Explicit gender stereotypes	Implicit gender stereotypes	Adj. R^2
Girls	Implicit math identity	.20	.04	.16	.23	.08
	Explicit math identity	.29	.08	.27	.13	.09
	School grades	.25	.06	.22	.19	.09
	Enrolment preferences	.29	.08	.27	.12	.09
Boys	Implicit math identity	--	--	--	(-.01)	.00
	Explicit math identity	-.28	.07	-.27	(-.02)	.07
	School grades	-.19	.03	-.16	(-.11)	.04
	Enrolment preferences	-.36	.13	-.35	(-.04)	.13

Note. Non-significant predictors ($p < .05$) are put in parentheses.

2.3.3 Relations of math-gender stereotypes with achievement

A difference score between mean math and German grades was computed for use in the correlational and regression analyses, with higher values indicating better grades in German than in math. Stronger implicit gender stereotypes were related with better German vs. math grades for girls, $r = .22$, and worse German vs. math grades for boys, $r = -.15$.

Separate hierarchical regressions were carried out for boys and girls with school grades as criterion and explicit and implicit gender stereotypes as predictors. In the girls' sample, again both explicit and implicit stereotypes were significant predictors, whereas for boys only explicit stereotypes predicted variance in school grades (see Table 3).

2.3.4 Relations of math-gender stereotypes with enrolment preferences

A difference score for German and math enrolment intention indices was computed with higher values indicating an enrolment preference for German. Implicit gender stereotypes were related to the enrolment preference only for girls. Stronger stereotypes were linked to a stronger enrolment preference for German over math classes, $r = .16$.

Hierarchical regressions with the enrolment preference as criterion and explicit and implicit gender stereotypes as predictors showed that, for girls, again both explicit and implicit gender stereotypes were significant predictors of enrolment preference. For boys, only explicit stereotypes predicted enrolment preferences (see Table 3). In sum, implicit math-gender stereotypes showed incremental validity beyond explicit stereotypes in predicting math identity, school grades and enrolment preferences only for girls.

2.4 Discussion

We investigated implicit math-gender stereotypes, implicit math identity, and their relations to math withdrawal. Implicit math-gender stereotypes were detected already among 9-year-old girls. Adolescent girls showed stronger implicit stereotypes than adolescent boys, who, on average, did not reveal any stereotypic associations in the IAT. Explicit gender stereotypes did not capture these gender differences. Implicit stereotypes showed unique predictive power beyond explicit stereotypes in predicting implicit and explicit math identity, enrolment preferences, and school grades for girls, but not boys. Further, girls aged 9 showed, on average, already implicit associations *German-self* and *math-other* (Study 1) while boys did not show any implicit affinity to math or *language*/*German* at any age. Gender differences in implicit math identity were significant in grade 9.

The finding that implicit math-gender stereotypes can be found in girls already at the age of 9 is consistent with other research. For example, elementary school girls associated high spelling skills with girls (Heyman & Legare, 2004). Further, girls aged 6-10 years rated women as less interested and less competent in math than men (Steele, 2003). However, girls in the same study did not report any negative math stereotypes regarding girls, and the restriction of the math-gender stereotype to adults could also be observed in a somewhat more indirect measure. In that measure, girls had to specify the gender of a mathematically talented person they heard about in a short story. Girls thought of an adult mathematician most often as a man, but supposed the mathematically talented child most frequently to be a girl (see also Ambady et al., 2001).

At first sight, these findings of elementary school girls showing no negative math stereotypes regarding their in-group seem to be inconsistent with the implicit stereotypes measured with an IAT. However, the IAT might have revealed stereotypes the girls were not able or willing to tell. Further, the IAT might not capture a clear distinction between math stereotypes regarding *girls* vs. *women*, but merely measure associations with the basic category *female*. Thus, implicit math-gender stereotypes in younger—and also older—girls may partly reflect their knowledge of women participating less in math-intensive fields. Finally, as both math and language stereotypes contribute to the gender stereotype IAT effect, the IAT effect in the youngest girls might be based on strong *language-girls* associations even in the absence of pronounced *math-boys* associations. Implicit math-gender stereotypes in older girls are less surprising as these girls may already have been confronted with math-gender stereotypes.

Elementary school boys did not show, on average, implicit math-gender stereotyping. This result is consistent with other findings. For example, boys attending grade 1-8 did not report any math-gender stereotypes (Ambady et al., 2001; Steele, 2003). A more indirect measure also yielded matching evidence. In the gender specification task by Steele (2003) alluded to above, boys aged 6-10 supposed both adults and children excelling in math or spelling to be male. Thus, elementary school boys revealed a strong in-group bias rather than gender stereotyping along cultural representations.

Adolescent boys showed weaker implicit math-gender stereotypes than adolescent girls, and on average, male 7^{th}- and 9^{th} graders did not reveal any stereotypic associations. This data pattern was unique to the implicit stereotype, whereas boys and girls showed comparable explicit gender stereotypes. At the same time, girls may experience gender stereotype activations more often than boys do, for example when facing stereotype threat during demanding math or science tests. These repeated stereotype activations might have produced stronger implicit stereotypes in girls than in boys. Nevertheless, it is somewhat surprising that adolescent boys did not show, on average, stereotypic association in the IATs as these boys have most likely been exposed to gender stereotypes during their socialization (e.g., Jacobs & Eccles, 1992). Further, a discrepancy remains between the gender differences in implicit stereotyping in our adolescent sample and the absence of such gender differences in the adult samples reported by Nosek et al. (2002a, 2002b). It might be possible that boys catch up with girls on implicit stereotyping not until early adulthood, for example after meeting predominantly male fellow students in technical study programs. Note, however, that the studies also differed in the concepts used in the IATs. Whereas Nosek et al. chose *arts* or *liberal arts* to represent the verbal academic domain, we used *German* or *language* because these terms were appropriate for children and adolescents. The concept (*liberal*) *arts* might have a strong female connotation particularly for men leading to strong implicit stereotypes in both genders. These inconsistencies should be reconciled in further research, for example by disentangling implicit stereotypes regarding math and language with a tool like the Go/No-go Association Task (Nosek & Banaji, 2001).

Whereas Nosek et al. (2002b) detected strong relations between implicit math-gender stereotyping and performance for men ($r = .51$), this relation was small for boys in the current studies. Further, we found no relations between implicit gender stereotypes and math identity for boys. However, though boys did not show, on average, any stereotypic associations, their stereotype IAT effects were not meaningless. In addition to the small, but significant relation between implicit stereotypes and school grades, boys' implicit math-gender stereotypes were also related with their explicit gender stereotypes and their perceptions of other people's stereotypes. These findings support the validity of the gender stereotype IATs also in the boys' sample, and gender differences in implicit math-gender stereotyping can be interpreted as a valid finding.

For girls, implicit math-gender stereotypes consistently predicted unique variance in implicit and explicit math identity, enrolment preferences, and school grades. These relations underpin the validity of these IATs. The present research shows that such relations are not confined to adults, but implicit stereotypes share common variance with achievement-related variables also in 13- to 15-year-old girls. As a limitation of our studies, the correlational nature of our data does not allow investigations of possible causal relations. Whereas activated gender stereotypes may shape self-concept variables or performance, personal ability estimations or achievement might also influence stereotypic associations about gender.

A prospective study should resolve this question (cf. Kiefer & Sekaquaptewa, 2007b).

2.5 Conclusion

Implicit math-gender stereotypes could be detected in 9-year-old girls and turned out to be strong in female adolescents. Boys, on average, did not reveal any implicit math-gender stereotyping. Girls' implicit gender stereotypes demonstrated unique predictive power in predicting math identity, enrolment preferences, and achievement. Together with an early implicit affinity to German/language vs. math, these findings suggest that implicit processes exert their influence on girls already at an early age and diminish their commitment to math-intensive fields. It should be the aim of educational policies to counteract implicitly operating biases already in young girls, and psychological research has to identify effective intervention strategies to abolish girls' and women's refraining from math and science (e.g., Dasgupta & Asgari, 2004). Role models such as female outstanding mathematicians like the famous Ruth Lawrence as well as female math professionals with a more attainable level of success will surely play a large role in changing career-related cognitions and decisions in female students.

The studies reported above could not explain why girls revealed stronger implicit math-gender stereotypes than boys. In IATs, implicit gender stereotypes regarding math and language are combined and cannot be separated. The stronger implicit stereotypes in girls might be based on stronger implicit *math-boys* or/and *language-girls* stereotypes. Do girls possess particularly strong associations representing a disadvantage (i.e., math-gender stereotypes) or an advantage (i.e., language-gender stereotypes) of their own gender? This topic was assessed in the research presented in the following chapter.

3 Separating Implicit Math-Male and Language-Female Stereotypes: Implicit Associations are Self-Serving for Boys and Men, but not for Girls and Women

Women are still underrepresented in math-intensive careers and earn only a small percentage of university diplomas (e.g., 15.4% of computer science diplomas in Germany in 2004) in these fields (Ramm & Bargel, 2005). In addition to ability self-concepts (e.g., Eccles, 1994), gender stereotypes regarding academic domains contribute to this gender gap in career choices. For example, negative stereotypes concerning women's math ability can undermine women's performance and interest in math (e.g., Davies et al., 2002; Spencer et al., 1999).

In order to analyze the impact of math-gender stereotypes, it should be investigated to what extent students have internalized these stereotypes. Both male and female students often disavow math-gender stereotypes when asked directly, revealing no explicit math-gender stereotypes (e.g., Ambady et al., 2001; Hyde et al., 1990b; Steele, 2003). However, despite counterstereotypic self-reports, students might possess negative stereotypes regarding women's math abilities. These stereotypes can be conceptualized as associations between gender and stereotypic attributes, for example *math-male* and *language-female*, and associations can differ in strength. These so-called implicit stereotypes can be activated automatically without intention or control, and they may influence behaviour without the person's awareness of that specific impact (see Greenwald & Banaji, 1995).

According to previous research, women and men showed implicit math-gender stereotypes measured by Implicit Association Tests (IATs) (Greenwald et al., 1998). Men and women revealed similar degrees of *math-male* and *arts-female* associations (Nosek et al., 2002a; Nosek et al., 2002b). Implicit math-gender stereotypes also demonstrated unique predictive power regarding math-related outcome variables. Stronger implicit stereotypes were related to a stronger math preference, higher math identification, and better math performance for men. For women, stronger stereotypes were related to a lower math preference, lower math identification, and lower math performance (Nosek et al., 2002b). In a prospective study with female university students, stronger implicit math-gender stereotypes measured with IATs predicted lower math performance and lower interests in math-related careers (Kiefer & Sekaquaptewa, 2007b). Further, implicit math-gender stereotypes measured with IATs appeared to moderate stereotype threat effects (Kiefer & Sekaquaptewa, 2007a).

Crucial career decisions are made during school years. Therefore, children's and adolescents' implicit math-gender stereotypes should also be considered. In Study 1 and 2 (Chapter 2) of the present research, implicit math-gender stereotypes were assessed with IATs in children and adolescents aged, on average, 9, 13, and 15 years. Girls aged 9 years who were attending grade 4 already revealed associations *math-boys* and *language-girls*, and adolescent girls aged 13 and 15 years attending grade 7 and 9 showed stronger implicit math-gender stereotypes than boys. For adolescent girls, but not boys, implicit math-gender stereotypes were related to explicit and implicit identification with language relative to math, enrolment preferences for language over math classes, and school grades favoring language over math.

Stronger implicit math-gender stereotypes in adolescent girls than boys call for an explanation, particularly as gender differences could not be found in adult samples (Nosek et al., 2002a; Nosek et al., 2002b). However, implicit associations of the concepts *math* and *language* are intertwined in the IAT effect and cannot be separated within the IAT (Nosek et al., 2005). Adolescent girls showing stronger implicit stereotypes may, as compared to boys, have stronger associations *math-boys*, they may have stronger associations *language-girls*, or both. Girls may have acquired stronger associations *math-boys* than boys because math-gender stereotypes might be activated in girls more often than in boys. For example, repeated stereotype threat experiences during demanding math or science tests might have strengthened the association *math-boys* in girls to a greater degree than in boys. On the other hand, girls outperform boys on various verbal tasks and also have higher verbal self-concepts than boys (Hyde & Kling, 2001; Hyde & Linn, 1988; Marsh, 1989), and being aware of this might have led to particularly strong associations *language-girls* in girls compared to boys.

In the present research we used Go/No-Go Association Tasks (GNATs) (Nosek & Banaji, 2001) to investigate whether the stronger implicit math-gender stereotypes in girls than boys can be traced to stronger *math-boys* and/or *language-girls* associations. GNATs have been developed as a measurement tool for implicit associations of a single concept with an attribute pair (e.g., *math* with *male* vs. *female*). Similarly to IATs, GNATs consist of two tasks. For example, in the stereotype-congruent task of a math-gender GNAT, participants have to respond to math- or male-stimuli by pressing the spacebar and ignore female-stimuli and other distractor stimuli. In the stereotype-incongruent task, responses are required to math- or female-stimuli, and participants are instructed to ignore male-stimuli and other distractor stimuli. Participants with strong *math-male* (vs. *math-female*) associations should react faster in the *math-male* than in the *math-female* task. In Study 3, two separate GNATs were applied to measure implicit *math-boys* and *language-girls* stereotypes[3] in 9th graders. In a third GNAT, the implicit identification with math vs. language was assessed (i.e., implicit self-concept). We investigated gender differences in implicit stereotypes and in the implicit self-concept. In Study 4, we tested with a sample of university students whether the findings obtained with 9th-graders could be generalized to adults.

3.1 Study 3

Implicit gender stereotypes regarding math and language were assessed separately with GNATs in a sample of 9th-graders. We tested whether girls have stronger *math-boys* and/or *language-girls* associations than boys. In a GNAT regarding academic self-concepts, we expected girls to have stronger *language-self* associations than boys. Thus, the self-concept GNAT should reveal known-groups validity as it has been demonstrated with math (vs. arts) identity IATs (Nosek et al., 2002b).

[3] To avoid confusions in terminology, the expression “implicit math-gender stereotypes“ refers only to stereotypes measured with IATs, combining both math- and language-associations. The expressions “Implicit *math-boys (men)* stereotypes“ or “implicit *language-girls (women)* stereotypes“ refer to stereotypes captured by GNATs.

Explicit math and language gender stereotypes were collected for comparison reason. We investigated explicit ability self-concepts and school grades to test whether our sample shows the usual gender differences in these measures. We expected gender differences in the math self-concept to favor boys and gender differences in the language (or the respective school subject, in our case, German) self-concept to favor girls (see Hannover, 1991; Marsh, 1989). Further, we expected girls to outperform boys on German grades, but we expected no gender differences in math grades (cf. Hannover, 1991; Kimball, 1989).

Additionally, we investigated the relations of implicit gender stereotypes with other math- and language-related outcomes. Similarly to IAT findings, implicit and explicit gender stereotypes should be correlated (see Lane et al., 2007). We also expected relations between implicit stereotypes and explicit ability self-concepts or grades of a given domain. For example, stronger implicit *math-boys* associations might be related to a lower math ability self-concept in girls. Further, relations between implicit gender stereotypes with variables pertaining to the other academic domain should be explored.

3.1.1 Method

3.1.1.1 Participants

Initially, we collected data of $N = 195$ participants attending 9th grades in various grammar schools (highest school track, Gymnasium) in East and West Germany and one Western German secondary school (intermediate school track, Realschule). Permissions to conduct the study were granted by school directors and parents. The adolescents participated in the study voluntarily during their regular school hours. Eight participants who had higher error rates than 30% in at least one combined GNAT task were removed from analysis. Altogether, $N = 187$ participants (mean age = 14 years 10 months; 91 boys and 96 girls) were included in the analysis. Of the 187 participants, 90 (48%) grew up in East Germany, 75 (40%) in West Germany, and 22 (12%) have lived in both parts of the country. The sample comprised 158 (84%) students from grammar school and 29 (16%) from secondary school. Both regional provenance and school track had no effects on implicit and explicit measures so that they were not included as factors in the analyses. Only few participants (14; 7.5%)

indicated that German was not their native language. However, theses students were too few to allow any conclusions, and excluding them did not change any results. Therefore, we kept these participants to have a greater statistical power.

3.1.1.2 Materials

Implicit measures. The math-gender GNAT consisted of the concept *math* and the concept pair *boys* vs. *girls*. The language-gender GNAT used the concepts *language* and *boys* vs. *girls*, the self-concept GNAT *self* together with *math* vs. *language*. Stimuli were selected to bear as few additional connotations as possible in order to maximize concept associations rather than stimulus associations in the GNAT effect (cf. Steffens et al., in press). Further, two distractor stimuli related to the broader concept *school* were used in the GNATs (*school break, school bus*) (see Table A2 in the Appendix 2 for a complete list of stimuli). Adding stimuli of a super-ordinate category to the no-go trials contributed to a somewhat larger GNAT effect (Nosek & Banaji, 2001).

Explicit ability self-concepts in math and German. Math self-concept was assessed with four items, for example, "I learn things quickly in math". All explicit ratings were made on 5-point scales, with lower values indicating a stronger agreement. This scaling corresponded to school grades in Germany with lower grades indicating better evaluations. Parallel items were used for the German ability self-concept.

School grades. Participants were asked to indicate their latest class test and report grades in math and German.

Explicit gender stereotypes. First, participants had to estimate the giftedness of boys and girls in math or German on four statements, for example, "Girls are often talented for doing math". Two items captured comparative gender stereotypes regarding math and German, using *girls* and *boys* as anchor points. In two further items, participants had to estimate to what extent they perceive most other people, in general, to have gender stereotypes regarding math or German. Again, *boys* and *girls* were used as anchor points. These items referring to perceived stereotypes might capture gender stereotypes more subtly.

Self-report computer skills. Participants had to rate their computer skills in three items, for example, “I am familiar with computers”.

Demographical questions. Demographical questions were presented at the end of the study in order to avoid gender priming effects. Translations of the explicit measures can be found in Appendix 2.

3.1.1.3 Procedure

After giving their informed consent, participants were tested in groups up to seven by a female experimenter. GNATs and explicit measures were administered on iBooks. All participants started with the gender stereotype GNATs and then completed the self-concept GNAT. After the GNATs, participants filled out the self-report measures in the order described above. Finally, all participants were debriefed and rewarded with small gifts. The procedure lasted about 25 minutes.

Concept labels were visible during a GNAT task, and stimuli were flashed in for 1000 ms. Participants had to press the space bar as fast as possible if a stimulus belonged to one of the concept labels (go trial). If a stimulus did not belong to either concept, participants were instructed to do nothing, and the stimulus disappeared after 1000 ms (no-go trial). False responses were indicated by a flashing “F!”. For example, in the stereotype-congruent task of the math-gender GNAT, participants had to respond to *boys*- or *math*-stimuli, but not to girls- or distractor stimuli. In the stereotype-incongruent task, responses were required for *math*- or *girls*-stimuli, but not for *boys*- or distractor stimuli. These critical tasks are also referred to as combined tasks.

Half of the stimuli in a combined GNAT task required a go-response, whereas the other half, comprising stimuli of the opposite gender and distractor stimuli, consisted of no-go trials. Each combined task consisted of 60 trials (+ 2 practice trials at the beginning). Practice tasks requiring responses to only one concept (6 trials) were employed before a combined task if new concepts were introduced.

The order of the stereotype GNATs (math vs. language) was counterbalanced. Further, one half of the participants started with the stereotype-congruent task in both stereotype GNATs (*language-girls*, *math-boys*), the other half with the stereotype-incongruent task (*language-boys*, *math-girls*). The self-concept GNAT was completed after the stereotype GNATs, and the task order (*self-math* vs. *self-language* first) was counterbalanced within each of the four conditions obtained from counterbalancing the stereotype GNATs.

3.1.1.4 Design

Dependent variables were GNAT effects in the math-gender, language-gender and the self-concept GNAT. For all GNATs, gender was treated as independent variable. For the gender stereotype GNATs, two additional control factors (GNAT order and task order) were included, resulting in a 2 x 2 x 2 between-subjects design. For the self-concept GNAT, task order was included as a control factor so that a 2 x 2 between-subjects design was obtained. Medium-sized gender differences in GNAT effects with an effect size of $d = 0.5$ can be detected with $\alpha = .05$ and a total sample size of $N = 180$ with a power of $1 - \beta = .92$.

3.1.2 Results

Unless indicated differently, statistical tests in Study 3 and 4 were conducted with $\alpha = .05$. Therefore, individual p-values are not reported for statistically significant effects. R^2_p is reported as an indicator of the effect size. R^2_p is numerically identical to partial Eta squared and is an estimate of the proportion of explained variance after partialling out other factors in the design (Cohen, 1977). Error reaction times were included in analyses. For computing the GNAT effect in each GNAT, the difference between each participant's average reaction times in the two combined tasks was divided by the participant's overall standard deviation of the response latencies in these tasks.

3.1.2.1 GNAT Analyses

Math-gender GNAT. For reliability estimations, separate GNAT effects were computed for go trials with odd position numbers in the combined tasks (i.e. trials with position numbers 1, 3 to 29) and go trials with even position numbers (i.e. position numbers 2, 4 to 30). Please note that the other half of the trials in a combined task were no-go trials requiring no response. The $GNAT_{odd}$ effect and the $GNAT_{even}$ effect were correlated with $r = .42$, revealing a rather low reliability.

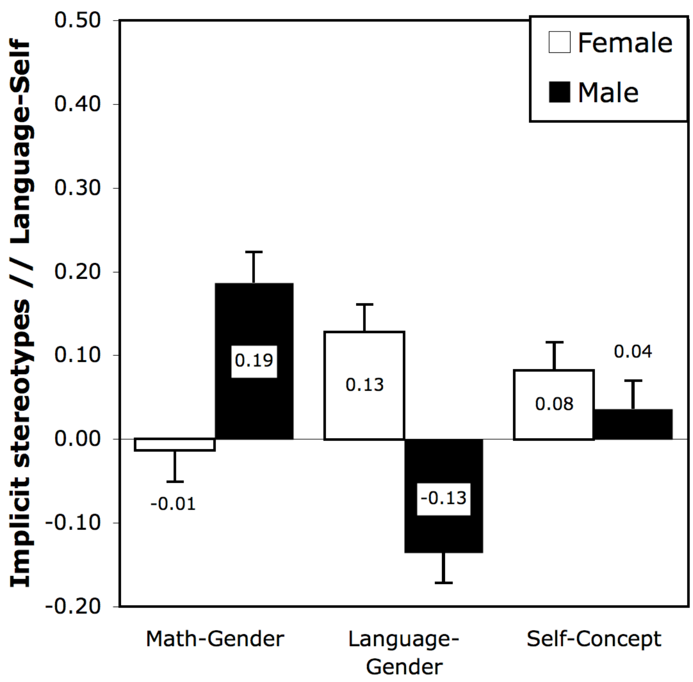


Figure 3: GNAT effects in the math-gender stereotype GNAT, the language-gender stereotype GNAT, and the academic self-concept GNAT, separately for boys and girls (9th graders) in Study 3. Positive GNAT effects indicate associations math-boys, language-girls, and language-self. Error bars reflect standard errors.

Larger GNAT effects indicate stronger stereotypic associations *math-boys*. According to the left part of Figure 3, only boys revealed *math-boys* associations, whereas girls, on average, did not show any stereotypic associations. This impression was confirmed by a 2 (gender) x 2 (task order) x 2 (GNAT order) ANOVA on the math-gender GNAT effect with a main effect of gender, F (1,179) = 16.11, R_p^2 = .08. Additionally, the main effect of task order reached significance, F (1,179) = 15.90, R_p^2 = .08, indicating that GNAT effects were biased in the direction of the task done first (all other Fs < 1.32). Further, GNAT effects were tested against 0 in one-sample t-tests separately for boys and girls with α = .025 to avoid an increase of overall α-level (Bortz, 1999). The GNAT effects found in boys differed significantly from 0, t (90) = 5.01, R_p^2 = .22.

Language-gender GNAT. The $GNAT_{odd}$ effect and the $GNAT_{even}$ effect were correlated with $r = .34$, revealing a low reliability. Positive GNAT effects indicate *language-girls* associations. According to the middle part of Figure 3, girls showed *language-girls* associations, whereas boys showed *language-boys* associations. The 2 x 2 x 2 ANOVA confirmed this gender difference, F (1,179) = 31.62, $R^2_p = .15$. Further, a main effect of task order emerged with GNAT effects being biased in the direction of the first task, F (1,179) = 22.45, $R^2_p = .11$. An interaction gender x task order also reached significance, F (1,179) = 4.08, $R^2_p = .02$ (all other Fs < 3.68). As one-sample t-tests against 0 ($\alpha = 0.025$) revealed, both the *language-girls* associations in girls, t (95) = 3.89, $R^2_p = .14$, and the *language-boys* associations in boys differed significantly from 0, t (90) = -3.65, $R^2_p = .13$.

Self-concept GNAT. The $GNAT_{odd}$ effect and the $GNAT_{even}$ effect were correlated with $r = .21$, revealing an unsatisfactory reliability. Positive GNAT effects indicate an association *language-self*. According to the right part of Figure 3, only girls showed an association *language-self*, whereas boys did not reveal any associations. However, a 2 (gender) x 2 (task order) ANOVA revealed only a main effect of task order, F (1,183) = 10.32, $R^2_p = .05$, indicating that GNAT effects were biased in the direction of the task done first (all other Fs < 1). As one-sample t-tests against 0 ($\alpha = 0.025$) revealed, the *language-self* association in girls differed significantly from 0, t (95) = 2.45, $R^2_p = .06$.

3.1.2.2 Explicit measures

Explicit gender stereotypes. To obtain the score for explicit *math-boys* stereotyping, boys' giftedness ratings in math were subtracted from girls' giftedness ratings with a higher value indicating stronger stereotype endorsement. Similarly, for the explicit *language-girls* stereotype score, girls' giftedness ratings were subtracted from boys' giftedness ratings. In a second step, these indices were transformed to a value range between 1 and 5 and averaged with the respective comparative stereotype item. One-sample t-tests against the neutral value of the scale (i.e., 3) were carried out separately for boys and girls ($\alpha = 0.025$). Both boys and girls endorsed gender stereotypes regarding math and language (all ts > 3.8 with $R^2_p s \geq .13$). Further, both boys and girls perceived other people bearing gender stereotypes regarding math and language, as one-sample t-tests ($\alpha = 0.025$) against the neutral value of the scale (i.e., 3) revealed (all *t*s > 4.3 with $R^2_p s \geq .17$). As the only gender difference in the self-report stereotypes, girls perceived stronger *math-boys* stereotypes in their environment than boys, *t* (185) = 2.82, $R^2_p = .04$ (see Table 4).

Table 4: Mean Explicit Math-Male Stereotypes, Language-Female Stereotypes, Perceived Math-Male Stereotypes, and Perceived Language-Female Stereotypes (Study 3 and 4).

		Math-male stereotypes	Language-female stereotypes	Perceived math-male stereotypes	Perceived language-female stereotypes
Grade 9	*Male*	3.29 (0.69)	3.54 (0.63)	3.52 (1.15)	3.90 (0.82)
	Female	3.26 (0.66)	3.67 (0.59)	3.97 (1.04)	4.06 (0.88)
University students	*Male*	3.62 (0.45)	3.68 (0.46)	4.45 (0.60)	4.27 (0.77)
	Female	3.73 (0.55)	3.83 (0.51)	4.56 (0.66)	4.47 (0.609)

Note. Higher values in explicit stereotypes indicate stronger stereotype ratings with possible values between 1 and 5. Standard deviations are in parentheses.

Explicit ability self-concepts. All self-concept items were recoded with higher values indicating higher ability estimations. Both the math and the German ability self-concept scale revealed high internal consistencies, with Cronbach's $\alpha = .91$ for the math and Cronbach's $\alpha = .86$ for the German scale. Whereas girls showed the typical advantage in the German self-concept compared to boys, *t* (185) = 2.85, $R^2_p = .04$, boys and girls did not differ in their math self-concepts, *t* (185) = 1.32. The items for the rating of one's own computer skills were highly reliable, Cronbach's $\alpha = .89$. Girls rated their own computer skills lower than did boys, *t* (185) = -3.87, $R^2_p = .07$ (see Table 5).

Table 5: Mean Explicit Math and German Ability Self-Concepts and Ratings of the Own Computer Skills (Study 3 and 4).

		Math ability	German ability	Computer skills
Grade 9	*Male*	3.19 (1.02)	3.26 (0.77)	4.25 (0.97)
	Female	3.39 (1.03)	3.60 (0.86)	3.69 (1.00)
University students	*Male*	3.00 (1.12)	3.34 (0.95)	3.75 (1.02)
	Female	3.08 (1.22)	3.67 (0.92)	3.19 (0.91)

Note. Values were recoded so that higher values indicate higher ability ratings with possible values between 1 and 5. Standard deviations are in parentheses.

School Grades. As the latest report and class test grades were correlated for German, $r = .46$, and math, $r = .42$, we calculated means for school grades in each domain. School grades were recoded with higher values indicating higher performance. In our sample, girls earned better math and German grades than boys, $t\ (185) = 3.29$, $R^2_p = .06$, for German and $t\ (185) = 2.50$, $R^2_p = .03$, for math grades.

In sum, explicit ability self-concepts and school grades in German showed the typical advantage for girls over boys. Contrary to the findings cited above, girls and boys did not differ in their math self-concepts, and girls earned even better math grades than boys. However, girls again seemed to underestimate their math ability because they did not rate their math ability higher than boys did despite their better achievements. This result is consistent with the more common findings of girls receiving comparable math grades as boys, but showing a lower math self-concept. Further, girls revealed lower computer skills ratings than boys.

3.1.2.3 Correlational Analyses

Correlations between implicit gender stereotypes and other math- and language-related variables were assessed in order to test the validity of these implicit measures. However, the low reliabilities of our GNATs are a drawback as they severely restrain correlation sizes. Thus, correlations tended to be rather low, and the pattern of correlations seems somewhat irregular as not all expected relations reached significance. All significance tests regarding correlations in Study 3 and 4 were one-tailed unless indicated differently. To control for task order effects within GNATs, we computed z-values of GNAT effects separately for participants starting with either task order. These z-values were used for correlational analyses in Study 3 and 4.

Correlations were observed mostly between implicit and explicit – and also perceived – stereotypes, often across the two academic domains. Stronger implicit *language-girls* stereotypes were related to stronger perceived *language-girls* stereotypes, $r = .20$, and to stronger perceived *math-boys* stereotypes, $r = .22$. Implicit *math-boys* stereotypes and explicit *math-boys* stereotypes were related with $r = .23$. Separate analyses for boys and girls revealed a series of additional correlations between implicit *math-boys* stereotypes and other variables, particularly for girls. For girls, stronger implicit *math-boys* stereotypes were related to stronger explicit *language-girls* stereotypes, $r = .25$, stronger perceived *language-girls* stereotypes, $r = .20$, and stronger perceived *math-boys* stereotypes, $r = .18$. Further, stronger implicit *math-boys* stereotypes were related to a lower explicit math self-concept, $r = -.18$, and to lower computer skills ratings, $r = -.19$, in girls. For boys, stronger implicit *math-boys* stereotypes were additionally related to worse German grades, $r = -.30$.

3.1.3 Discussion

Gender differences were observed in both stereotype GNATs. In the math-gender GNAT, only boys revealed a stereotypic association *math-boys*, whereas girls, on average, did not show any associations regarding math. In the language-gender GNAT, girls showed an association *language-girls*, whereas boys revealed a counter-stereotypic association *language-boys*. In the self-concept GNAT, only girls showed an association *language-self*, but the gender difference did not reach significance.

The current findings in the stereotype GNATs can offer a plausible post-hoc explanation for girls showing stronger implicit math-gender stereotypes than boys measured with IATs (see Chapter 2). In the IAT effect, associations *math-boys* and *language-girls* are combined to a joint stereotype score. Girls having *language-girls* associations together with no stereotypic math-associations should nevertheless show implicit stereotyping in the IAT due to the stereotypic language-associations. On the other hand, stereotypic math-associations and counter-stereotypic language-associations should form a low IAT effect for boys. Boys possess gendered associations regarding math and language, but these associations might be obscured in the IAT as both associations are combined in the IAT effect.

The gender differences found in the stereotype GNATs could not be replicated in the explicit stereotypes. Boys and girls did not differ in their endorsement of math and language gender stereotypes, and girls perceived even stronger *math-boys* stereotypes in their environment than boys did. Additionally, the implicit association *language-boys* in boys is not consistent with their explicit view of language as a girls' domain. However, implicit stereotypes were related to a variety of other variables, supporting the validity of these GNATs. The relation between implicit and explicit *math-boys* stereotypes is not surprising, as implicit and explicit measures of the same construct are likely to be related (Nosek, 2005). The relation between implicit and perceived language stereotypes is not surprising because both measures can be conceived as subtle stereotype measures (see Bohner & Wänke, 2002). Further, many correlations could be observed across the academic domains. For example, stronger implicit *math-boys* stereotypes were related to stronger explicit (and also perceived) *language-girls* stereotypes in girls. Thus, gender stereotypes regarding math and language are not independent, but reveal a general tendency for a gendered perception of these academic domains.

3.2 Study 4

In a subsequent study, we tested whether our findings could be replicated in an adult sample. Men were expected to show stronger implicit *math-men* stereotypes than women, and women should reveal stronger implicit *language-women* stereotypes than men.

3.2.1 Method

3.2.1.1 Participants

The initial sample comprised 192 students at the University of Jena. Three participants with higher error rates than 30% in at least one combined GNAT task were excluded from analyses. Altogether, data of 189 participants (mean age = 22.1 years, range = 18 - 35 years; 71 men, 118 women) were included in the analyses. Of these participants, 133 (70%) have lived mainly in East Germany, 21 (11%) mainly in West Germany, and 35 (19%) in both parts of the country. Among women, 85 (72%) studied a nonmath major (i.e., liberal arts, social sciences, law), whereas 33 (28%) studied a math-intensive major (i.e., economics, science, math, medicine, engineering, and computer science). Among men, 41 (58%) studied a nonmath major and 30 (42%) a math-intensive major. Only 7 participants (3.7%) indicated that German was not their native language. As results did not change after excluding these participants, their data was included in the analyses. Participants received either course credit or a chocolate bar for their participation.

3.2.1.2 Materials

Implicit measures. As a minor change to Study 3, the concept labels *men* and *women* were used for the gender dimension in the stereotype GNATs to be appropriate for adults. Concepts and stimuli of the self-concept GNAT remained unchanged. Further, *break* and *dorm* were used as distractors in all three GNATs (see Appendix 2 for a complete list of stimuli).

Explicit measures. Explicit measures regarding German and math ability self-concepts, self-report computer skills, and demographical questions were the same as in Study 3. As a minor change in the stereotype items, the terms *boys* and *girls* were replaced by *men* and *women*. Further, participants had to indicate the last report grades in math and German at school. Translations of the explicit measures can be found in Appendix 2.

3.2.1.3 Procedure

After giving their informed consent, participants were tested in groups up to seven by a female experimenter. GNATs and self-report measures were administered on iBooks. The length and the counterbalancing of the GNATs were identical to Study 3. After the GNATs, participants completed the self-report measures in the same order as described in Study 3. After the study, participants were debriefed and rewarded. The study lasted about 20 minutes.

3.2.1.4 Design

Design and power estimation were identical to Study 3. In a second step, we conducted combined analyses of the adolescents' and the university students' sample, resulting in a 2 (gender) x 2 (task order) x 2 (GNAT order) x 2 (age group) between-subjects design for the stereotype GNATs and a 2 (gender) x 2 (task order) x 2 (age group) between-subjects design for the self-concept GNAT. Given a generalizability of our findings, the factor age group should not show any interactions with gender.

3.2.2 Results

GNAT effects were computed as described in Study 3 and can be obtained from Figure 4.

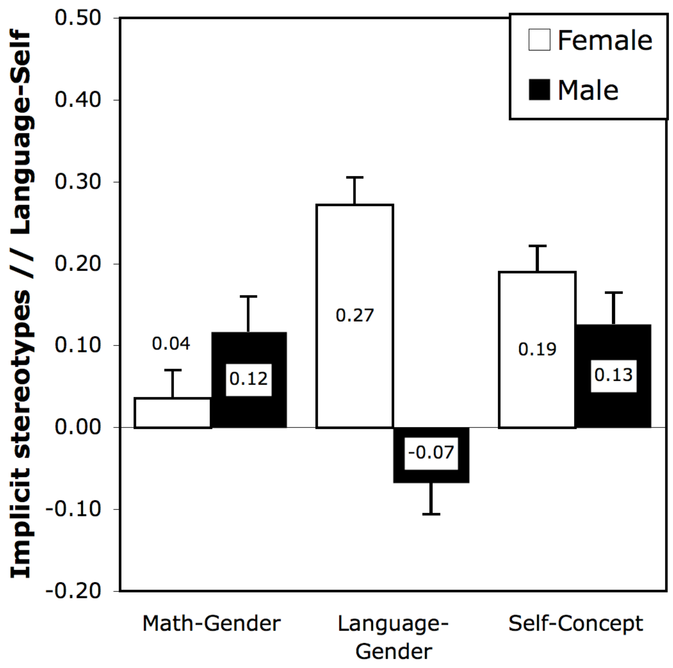


Figure 4: GNAT effects in the math-gender stereotype GNAT, the language-gender stereotype GNAT, and the academic self-concept GNAT, separately for men and women in Study 4. Positive GNAT effects indicate associations math-men, language-women, and language-self. Error bars reflect standard errors.

3.2.2.1 GNAT Analyses

Math-gender GNAT. $GNAT_{odd}$ effects and $GNAT_{even}$ effects were correlated with r = .39, revealing a rather low reliability. According to the left part of Figure 4, men showed stronger associations *math-men* than women. However, a 2 x 2 x 2 ANOVA revealed only a main effect of task order, F (1,181) = 13.78, $R^2_p = .07$, indicating that GNAT effects were biased in the direction of the task done first (all other Fs < 2.75). However, separate one-sample t-tests against 0 for men and women ($\alpha = .025$) revealed, as expected, stereotypic associations in men, t (70) = 2.73, $R^2_p = .10$, but not in women, t (117) = 1.04.

To test the comparability of the findings among adolescents and adults, we also analyzed both data sets together. The 2 x 2 x 2 x 2 ANOVA yielded the expected main effect of gender, F (1,360) = 15.80, $R^2_p = .04$, and a main effect of task order, F (1,360) = 29.59, $R^2_p =$

.08 (all other *F*s < 2.50). No interaction gender x age group could be found so that gender differences were basically comparable for adolescents and university students. One sample *t*-tests against 0 separately for all male and female participants (α = .025) revealed stereotypic associations for the male sample, *t* (161) = 5.54, R_p^2 = .16, but not for the female sample (*t* < 1).

Language-gender GNAT. $GNAT_{odd}$ effects and $GNAT_{even}$ effects were correlated with *r* = .43, revealing a rather low reliability. According to the middle part of Figure 4, women revealed strong associations *language-women*, whereas men showed an association *language-men*. A 2 x 2 x 2 ANOVA yielded the expected main effect of gender, *F* (1,181) = 37.39, R_p^2 = .17. Further, a main effect of task order emerged, *F* (1,181) = 6.37, R_p^2 = .03, with GNAT effects biased in the direction of the first task (all other *F*s < 1.13). According to two one-sample *t*-tests against 0 (α = .025), women revealed significant associations *language-women*, *t* (117) = 8.01, R_p^2 = .35, whereas the association *language-men* in the men's sample did not reach significance, *t* (70) = -1.72, *p* = .09 (two-tailed).

In the combined analysis of adolescents' and university students' data, main effects of gender, *F* (1,360) = 69.17, R_p^2 = .16, and task order, *F* (1,360) = 25.24, R_p^2 = .07, emerged. Again, no interaction gender x age group could be found. The main effect age group with university students showing stronger stereotypes than adolescents, *F* (1, 360) = 8.86, R_p^2 = .02 (all other *F*s < 2.90), should not be interpreted as the university students' sample consisted of more women than men. One-sample *t*-tests against 0 separately for all male and female participants (α = .025) revealed a *language-female* association for female participants, *t* (213) = 8.54, R_p^2 = .25, and a *language-male* association for male participants, *t* (161) = -3.91, R_p^2 = .09.

Self-concept GNAT. $GNAT_{odd}$ effects and $GNAT_{even}$ effects were correlated with *r* = .35, revealing only a low reliability. According to the right part of Figure 4, women showed somewhat stronger *language-self* associations than men. However, a 2 x 2 ANOVA revealed no effects (all *F*s < 1.74). Two one-sample *t*-tests against 0 (α = .025) revealed *language-self* associations both in men, *t* (70) = 3.29, R_p^2 = .13, and women, *t* (117) = 5.99, R_p^2 = .23.

In the combined sample with adolescents and university students, neither a main effect of gender nor a gender x age group interaction could be found, but only a main effect of task order, F (1,368) = 9.40; $R^2_p = .03$. The main effect of age group with university students showing stronger *language-self* associations than adolescents, F (1,368) = 7.92, $R^2_p = .02$ (all other Fs < 2.66), should not be interpreted as the university student sample consisted of more women than men. One-sample t-tests against 0 for all male and female participants ($\alpha = .025$) revealed *language-self* associations for both female participants, t (213) = 6.08, $R^2_p = .15$, and male participants, t (161) = 2.98, $R^2_p = .05$.

3.2.2.2 Explicit measures

Explicit gender stereotypes. Stereotype scores were computed as in Study 3. One-sample t-tests against the neutral point of the scale (i.e., 3) separately for men and women ($\alpha = .025$) revealed that both men and women endorsed gender stereotypes regarding math and language (all ts > 11.52 with $R^2_p s \geq .65$). Similar t-tests showed that both men and women perceived math and language gender stereotypes in their environment (all ts > 13.80 with $R^2_p s \geq .73$). Women endorsed stronger *language-women* stereotypes than men, t (187) = 2.04, $R^2_p = .02$, and also perceived stronger *language-women* stereotypes in other people than men, t (187) = 2.04, $R^2_p = .02$ (see Table 4).

Explicit ability self-concepts and school grades. Whereas men and women did not differ in their math self-concepts and their last math grades at school, both ts < 1, women revealed higher verbal self-concepts than men, t (187) = 2.40, $R^2_p = .03$, and also reported better German grades at school, t (187) = 3.10, $R^2_p = .05$. Women rated their own computer skills lower than men did, t (187) = -3.91, $R^2_p = .08$ (Cronbach's $\alpha = .89$) (see Table 5).

3.2.2.3 Correlational Analyses

Data preparation and significance tests were carried out as described in Study 3. Regarding correlations between implicit and explicit measures of the same construct, only the implicit self-concept and the difference score between explicit German and math self-concepts were related, $r = .23$. For men, stronger implicit *language-women* stereotypes were further related to better math grades, $r = .20$, and also to better German grades, $r = .24$ (two-tailed), with the latter correlation showing a stereotype-inconsistent direction. Additionally, stronger associations *language-self* in men were related to less pronounced explicit *math-men* stereotypes, r = -.23, and also to lower estimates of the own computer skills, r = -.22. For women, stronger implicit *math-men* stereotypes were related to stronger *language-self* associations, r = .17, and to lower math self-concepts, r = -.22. Additionally, women with stronger *language-self* associations perceived stronger *math-men* stereotypes, r = .21, and also stronger *language-women* stereotypes, r = .22, in the environment. Stronger *language-self* associations in women were also related to better German grades, r = .23, and to better math grades, r = .26 (two-tailed). The latter correlation showed an unexpected direction.

3.2.3 Summary of findings

In sum, implicit gender stereotypes and implicit self-concepts were comparable in the adolescents' and university students' samples. Though the gender difference in the implicit *math-men* stereotype did not reach significance among university students, only men showed a *math-men* association. Women showed again stronger *language-female* associations than men. Contrary to adolescents, *language-men* associations in men did not reach significance, but revealed only a tendency. In the self-concept GNAT, women showed again only descriptively stronger *language-self* associations than men, with both men and women revealing significant *language-self* associations.

Women reported comparable explicit *math-men* self-concepts and math school grades as men, but revealed higher German self-concepts and better German school grades than men. Similarly to implicit *language-women* stereotypes, women endorsed language stereotypes to a stronger degree than men and also perceived more language stereotyping than men. Unlike Study 3, correlations could be found predominantly between the implicit self-concept and all instances of stereotype measures. Additionally, implicit *math-men* stereotypes were again negatively related to explicit and implicit math self-concepts in women.

3.3 Discussion

We investigated implicit *math-male* stereotypes, implicit *language-female* stereotypes and implicit self-concepts regarding math vs. language with GNATs. Participants were adolescents attending grade 9 and university students. Boys revealed stronger implicit *math-boys* stereotypes than girls, who, on average, did not show any stereotypic associations regarding math. Though this gender difference did not reach significance in the adult sample, results were comparable as only men revealed *math-men* associations. Generally, female participants showed stronger implicit *language-female* stereotyping than male participants. Adolescent boys revealed even a significant *language-boys* association; *language-men* associations in men did not reach the conventional level of significance. In the self-concept GNAT, girls and women revealed *language-self* associations. However, their implicit self-concepts did not differ significantly from self-concepts shown by boys and men, with men also revealing a *language-self* association. Altogether, the pattern of GNAT effects along with gender differences found in adolescents could be widely replicated in the adult sample.

The gender differences in the implicit *math-male* stereotypes were not mirrored in the self-report *math-male* stereotypes. As the only gender difference, girls perceived even stronger *math-boys* stereotypes in their environment than boys. On the contrary, women (but not girls) revealed stronger explicit (and also perceived) *language-women* stereotypes, which is in line with females showing stronger implicit *language-female* stereotypes. Girls and women reported higher German self-concepts than boys and men along with comparable math self-concepts as male participants. These findings resemble the association of *self* with *language* (vs. *math*) in female participants, but not in boys.

Implicit gender stereotypes measured with GNATs among adolescents are consistent with their implicit math-gender stereotypes measured with IATs (see Chapter 2). In these IATs, adolescent girls revealed stronger stereotypes *math-boys* and *language-girls* than boys. Whereas no stereotypic associations regarding math and strong *language-girls* associations should lead to a stereotypic IAT effect in girls, stereotypic *math-boys* associations and counterstereotypic *language-boys* associations in boys should cancel each other out, explaining the low stereotype IAT effect in adolescent boys.

Nosek et al. (2002a, 2002b) did not find gender differences in math-gender stereotypes measured with IATs in adults. They employed the concept (*liberal*) *arts*, whereas the concept *language* had been used for the verbal academic domain in the present Studies 2-4. Though speculative, the concept (*liberal*) *arts* might be strongly linked to the concept *female* for both women and men. Thus, as counterstereotypic associations were missing in the IATs administered by Nosek et al., men have also demonstrated implicit math-gender stereotypes. However, we chose the term *language* to have appropriate concepts also for adolescents.

Implicit math and language stereotypes as well as the implicit self-concept were correlated with achievement-related variables despite the rather low reliability of the implicit measures. Implicit gender stereotypes were related to their explicit counterparts (Study 3), and also the implicit and explicit self-concept measures were related (Study 4). For both adolescents and adults, implicit gender stereotypes of one academic domain were related to variables pertaining to the other academic domain. For example, stronger implicit *language-girls* stereotypes were related to stronger perceived math stereotypes (Study 3), and men reporting better school math grades revealed more pronounced implicit associations *language-women* (Study 4). Thus, (implicit and explicit) stereotypes regarding math and language seem to reflect a general tendency for a gendered perception of the academic domains. However, as we have only correlational data, we cannot investigate any causal relationships.

The separate measurements of implicit math and language stereotypes yielded findings that were unexpected at the first sight. For example, the association *language-boys* in boys seem to be odd given their traditional explicit *language-girls* stereotypes. Further, the lack of implicit *math-male* stereotyping in girls and women seems contradictory in the light of numerous stereotype threat effects (e.g., Spencer et al., 1999). However, the observed associations regarding math and language in females and males can be reconciled. We postulate that the implicit associations regarding math and gender in girls and women have an exceptional position, as they do not express group-serving (and by extension, self-serving) math associations with the own gender. This seems to reflect the negativity of the math-gender stereotypes for girls and women. Regarding the other associations, female participants revealed strong associations *language-own gender*, whereas male participants showed *math-own gender* associations. Whereas the *language-own gender* associations in men did not reach the conventional level of significance, boys showed significant *language-own gender* associations. Boys' implicit associations regarding math and language are consistent with their ingroup bias in gender ascriptions to persons who were good at math or spelling (Steele, 2003). In that study, young boys aged 6-10 years supposed both adults and children excelling either in math or in spelling to be male. In a study investigating gender-stereotypic traits, for both women and men, implicit gender stereotypes regarding the own gender were biased in a self-favourable direction. For example, men revealed stronger *men-powerful/women-weak* associations than women, particularly if power-related words were positive and weakness-related words negative in valence (Rudman et al., 2001). Thus, a self-serving (or self-enhancing) component may be quite common in implicit gender stereotyping, even for adolescents and adults.

As the only case, girls and women, on average, never revealed self-serving gender associations regarding math. As a post-hoc explanation, this missing *math-female* association might be an indicator for women's or girls' vulnerability to math-gender stereotypes. If girls or women experience failure in a difficult math test and/or if math-gender stereotypes are made salient (e.g., by providing stereotypic test descriptions), implicit *math-male* stereotyping might increase rather easily and exert its detrimental influence as no self-serving implicit associations can act as buffer.

According to this interpretation, boys or men should not be that much affected by language-gender stereotypes. Up to now, negative effects of language-gender stereotyping on men have been rarely demonstrated (e.g., Keller, 2007). Further, it might be the case that language-gender stereotypes are not as threatening for men as math-gender stereotypes might be for women. For example, bad performance in a task introduced as fitting to women's abilities increased participants' perception of the male target person as masculine (Reinhard, Stahlberg, & Messner, in press). It is quite probable that men would even appreciate this consequence rather than fear it.

Future research should address the role of women's implicit *math-male* stereotypes in performance situations. Associations of math with male vs. female do not seem to be chronically activated in girls and women. However, as implicit stereotypes are malleable (see Blair, 2002), these math-gender stereotypes might become activated in achievement situations containing stereotypic cues and impede math performance in girls and women.

3.4 Conclusion

At the first sight, the lack of implicit *math-male* associations in females might give the impression that this implicit bias might be only weak. However, the missing implicit *math-male* associations in girls and women seem to be a special case not revealing self-serving associations. This might be interpreted as an indicator for a vulnerability to math-gender stereotypes.

In the following chapter, an experiment will be presented that addressed boundary conditions for the activation of *math-men* stereotypes in women. A stereotypic vs. non-stereotypic description of an announced math test and the later exposure to that test were the independent variables. Effects of test descriptions and test exposure on implicit *math-men* stereotypes measured by a math-gender GNAT were examined in a sample of female university students.

4 Stereotypic Math Test Descriptions and Subsequent Exposure to a Math Test Activate Math-Gender Stereotypes in Women

Though women have caught up with men on post-secondary education during the last decades, only a small percentage of women enter math-intensive fields like computer science or engineering (U.S. Department of Education, 2000). Gender stereotypes concerning women's alleged incompetence in math have a negative impact on women. For example, activating math-gender stereotypes during a math test can impair women's math performance and undermine their interest in math (e.g., Davies et al., 2002; Spencer et al., 1999). This can be explained by women's concerns of being judged in terms of the stereotype or confirming the negative stereotype about their group (i.e., stereotype threat, see Steele, Spencer, & Aronson, 2002 for a review).

When investigating to what extent students have internalized math-gender stereotypes, both male and female students often disavow these stereotypic beliefs when asked directly (Ambady et al., 2001; Hyde et al., 1990b). However, even individuals who do not endorse math-gender stereotypes on an explicit level may possess negative stereotypes regarding women's math aptitude. These stereotypes can be conceptualized as associations between gender and stereotypic attributes, for example, *math-male* and *language-female*, and associations can differ in strength. These so-called implicit stereotypes can be activated automatically without intention or control, and they may influence behavior without the person's awareness of that specific impact (see Greenwald & Banaji, 1995).

Men and women showed implicit math-gender stereotypes measured with Implicit Association Tests (IATs) (Greenwald et al., 1998), revealing associations *math-male* and (*liberal*) *arts-female* (Nosek et al., 2002a; Nosek et al., 2002b). These implicit gender stereotypes also demonstrated unique predictive power regarding math-related outcomes. Stronger stereotypes were related to stronger math preferences, stronger math identification and better math performance for men. For women, stronger stereotypes were related to weaker math preferences, weaker math identification and to worse math performance (Nosek et al., 2002b). Further, implicit math-gender stereotypes could predict women's math performance in a prospective study and appeared to moderate stereotype threat effects (Kiefer & Sekaquaptewa, 2007a, 2007b).

Differently from the IAT research described above, implicit math-gender and language-gender stereotypes were investigated separately with Go/No-go Association Tasks (GNATs) (Nosek & Banaji, 2001) as presented in Chapter 3 of the present research. Participants were 15-year-old adolescents attending grade 9 and university students with various majors. Whereas boys and men revealed implicit *math-male* associations, girls and women did not demonstrate any *math-male* associations, but only strong *language-female* associations. Additionally, boys revealed a counterstereotypic (and also self-serving) *language-male* association.

Apparently, girls and women do not posses, on average, chronically activated *math-male* associations. However, many studies have demonstrated that implicit stereotypes are malleable and can be influenced by various internal (e.g., motivation, attention) or external (e.g., situational cues) factors (see Blair, 2002; Dasgupta & Asgari, 2004). Thus, implicit *math-male* stereotypes might become activated and exert their influence on math-related outcomes. This view is supported by findings of stereotype activation in the context of stereotype threat effects. Steele and Aronson (1995, Study 3) demonstrated stereotype activation in participants during a stereotype threat-evoking situation. Further, the decrease in women's math performance was mediated by the level of stereotype activation measured with a lexical decision task (Davies et al., 2002). Altogether, both the malleability of implicit stereotypes and the stereotype activation during stereotype threat imply that implicit *math-male* stereotypes should be observed in women under certain circumstances. Identifying boundary conditions for a stereotype activation would have also practical relevance, as detrimental stereotype activations, for example in testing situations, could be avoided easier.

We focused in our research on two situational factors that might affect stereotype activation, namely (i) a stereotypic math test description and (ii) exposure to a difficult math test. Taken from stereotype threat research, the relevance of math-gender stereotypes for a given test can be stressed by describing it as usually producing (vs. not producing) gender differences (e.g., Spencer et al., 1999, Study 2). An instruction characterizing a test as usually showing gender differences is one variant of stereotype threat instructions, whereas stressing the equality of genders in a task is a common non threat-instruction (see Maass & Cadinu, 2003, for an overview). Thus, a stereotypic test description postulating gender differences might be sufficient for stereotype activation. However, it might be the case that additionally taking a demanding math test would be required for stereotype activation. Exposure to a demanding math test is sufficient to evoke stereotype threat effects even in absence of any other remarks about the test (Spencer et al., 1999, Study 3). Further, difficulties during such a math test might increase the applicability of math-gender stereotypes to the situation in the eyes of the female test-takers. On the contrary, the test description negating gender differences makes math-gender stereotype less applicable to the testing situation, and women should not demonstrate implicit *math-male* stereotypes, even after difficulties with the test.

4.1 Study 5

In the current experiment, female university students took a test allegedly measuring math ability and completed a GNAT measuring math-gender stereotypes. At the beginning of the study, the participants read that the math test usually showed (or did not show) gender differences. The math-gender GNAT was administered either directly after reading the test description and before taking the math test (math test second), or after the math test (math test first). Either all women who have read the stereotypic test description or only women who additionally took the math test before completing the GNAT should reveal implicit *math-male* stereotypes. Further, women who have read the non-stereotypic test description should not reveal an activation of math-gender stereotypes in the GNAT.

Regarding the math test, we expected a stereotype threat effect. Women who have read the stereotypic test description should perform worse in the test than women who have obtained the non-stereotypic test description. Further, we tested whether the performance decrements due to stereotype threat would be mediated by the level of stereotype activation captured by the GNAT.

4.1.1 Method

4.1.1.1 Participants

Participants were 128 female students at the University of Jena (mean age = 22.1 years, range = 18 - 33 years). The majority of the participants (79%, 101 participants) studied nonmath majors (i.e., liberal arts, social sciences, law). Math-intensive majors (i.e., economics, science, math, medicine, engineering, and computer science) had been chosen by 21% (27) of our participants. Participants received either course credit or a chocolate bar for their participation.

4.1.1.2 Materials

Implicit stereotype measure. The math-gender GNAT consisted of the concept *math* (together with the stimuli computation and equation) and the concept pair *men* (stimuli: men, boys) vs. *women* (stimuli: women, girls). Further, two distractor stimuli (break, apartment) were used (cf. Nosek & Banaji, 2001).

Math test. In the first part of the alleged math ability test, four word problems from the Brain Twister Test (Lienert, 1964) and four word problems from the Mathematical Thinking Tasks 10+ (Bartel, Hylla, & Süllwold, 1970) were presented in a mixed order. The Brain Twister Test is supposed to measure creative problem solving and logical thinking in students taking A-levels, requiring only little mathematical knowledge. The Mathematical Thinking Tasks 10+ assess mathematical giftedness in students attending grade 10. The second part of the test contained 13 items from the matrices subtest of the Intelligence Structure Test 2000 (Amthauer, Brocke, Liepmann, & Beauducel, 2000). These items were sorted in an ascending order of difficulty. In the I-S-T 2000, the matrices are one of three subtests measuring figural intelligence (e.g., the ability to deal with two- or three-dimensional objects or to detect logical relations between figures, see Amthauer et al., 2000, p. 80). These items are similar to Raven’s Advanced Progressive Matrices (Raven, 1962), which have been successfully used to demonstrate stereotype threat effects regarding intellectual ability in black students and lower SES students (Brown & Day, 2006; Croizet et al., 2004).

Self-report measures. The identification with math was measured by the agreement to the statement “It is personally important to me how good I am at math”. The explicit math ability self-concept was measured by the participants’ agreements to the statements “I am gifted for math” and “I am good at math”. Further, participants had to rate the importance of math for their major “My major is very math-intensive”. All explicit ratings were made on 5-point scales, with higher values indicating a stronger agreement. Participants were asked about their last math grade at school and, if applicable, at university. Explicit *math-male* stereotypes were captured by one item with *men* and *women* as anchor points. Participants were probed for suspicion, and after filling out demographic questions, they had to indicate whether their math test had been characterized as usually producing or not producing gender differences. This item served as manipulation check. Translations of the explicit measures can be found in Appendix 3.

4.1.1.3 Procedure

After giving their informed consent, participants were tested in groups up to seven by a female experimenter. First, participants were informed that they were to take a math ability test, which was introduced as a part of a newly developed cognitive skills training for university students. Participants were randomly assigned to one of the test description conditions. Half of the participants read that the performance of men and women usually differs (vs. does not differ) in this test. Then, participants completed the math test and the GNAT in a counterbalanced order. The math test was handed out in booklets, and participants worked on it for 20 minutes. After the math test and the GNAT, self-report measures were applied in the order described above. Finally, all participants were thanked and fully debriefed. The study lasted about 30 minutes.

The GNAT was administered on iBooks. Concept labels were visible during a GNAT task, and stimuli were flashed in for 1000 ms. Participants had to press the space bar as fast as possible if a stimulus belonged to one of the concept labels (go trial). If a stimulus did not belong to either concept, participants were instructed to do nothing, and the stimulus disappeared after 1000 ms (no-go trial). False responses were indicated by a flashing “F!”. In the stereotype-congruent task, participants had to respond to male- or math-stimuli, but not to female-stimuli or distractor stimuli. In the stereotype-incongruent task, responses were required for female- or math-stimuli, but not for male- or distractor stimuli. These tasks are referred to as combined tasks.

Half of the stimuli in a combined task required a go-response, whereas the other half, comprising stimuli of the opposite gender and distractor stimuli, consisted of no-go trials. The combined tasks consisted each of 60 trials (+ 2 practice trials at the beginning). The GNAT started with two practice tasks (6 trials) classifying math-stimuli (practice task 1) and female- (or male-) stimuli (practice task 2). Before the second combined task, responses to the other gender were practiced in a practice task (6 trials). One half of the participants received the *math-male* combined task first, the other half the *math-female* combined task.

4.1.1.4 Design

Dependent variables were the stereotype activation measured by the GNAT and the performance in the math test. For the GNAT, the test description (stereotypic vs. non-stereotypic) and the order of the math test vs. the GNAT (math test first or second) were the independent variables. Further, the task order within the GNAT (*math-male* vs. *math-female* first) was added as a control factor so that a 2 x 2 x 2 between-subjects design was obtained. For the math test performance, the test description was the independent variable, and the order of the math test vs. the GNAT was included as a control factor, resulting in a 2 x 2 between-subjects design. Medium-sized effects of test descriptions on the math test and the stereotype activation with an effect size of $f = .25$ can be detected with $\alpha = .05$ and a total sample size of $N = 128$ with a power of $1 - \beta = .80$.

4.1.2 Results

Unless indicated differently, statistical tests were conducted with $\alpha = .05$. Therefore, individual p-values are omitted for significant effects. R_p^2 is reported as an indicator of the effect size. R_p^2 is numerically identical to partial Eta squared and is an estimate of the proportion of explained variance after partialling out other factors in the design (Cohen, 1977).

4.1.2.1 Manipulation Check

We performed chi-square analyses on participants' responses whether their math test had been described as usually producing gender differences or not. The majority of the participants (67% in the stereotypic test description condition and 89% in the non-stereotypic test description condition) remembered the information correctly, Chi-Square (1) = 38.99. No participant mentioned any suspicions about a relation between the test description and the math performance or the math-gender stereotyping.

4.1.2.2 Math-gender GNAT

For computing the GNAT effect, the difference between each participant's average reaction times in the two combined tasks was divided by the participant's overall standard deviation of the response latencies in these tasks. Error reaction times were included in the analyses.

For reliability estimations, separate GNAT effects were computed for go trials with odd position numbers in the combined tasks (i.e., trials with position numbers 1, 3 to 29) and go trials with even position numbers (i.e., position numbers 2, 4 to 30). Please note that the other half of the trials in a combined task were no-go trials requiring no response. The $GNAT_{odd}$ effect and the $GNAT_{even}$ effect were correlated with r = .45, revealing a rather low reliability.

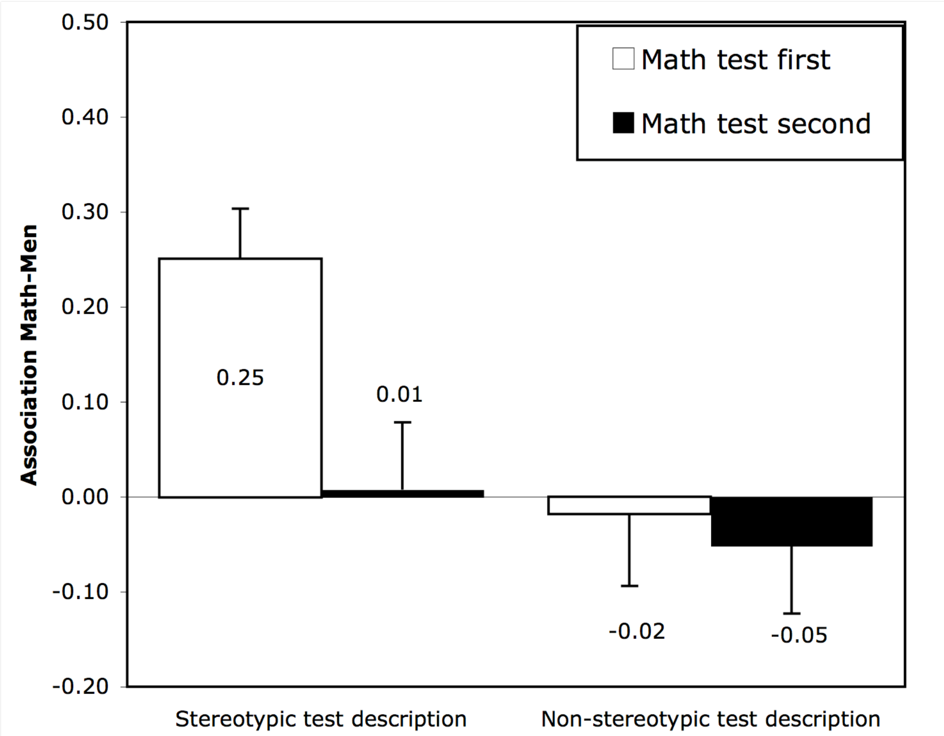


Figure 5: GNAT effects in the math-gender stereotype GNAT, separately for test description condition and order of GNAT vs. math test. Positive GNAT effects indicate associations math-men. Error bars reflect standard errors.

Larger GNAT effects indicate stronger stereotypic associations *math-men*. According to Figure 5, only women who read the stereotypic test description and who additionally worked on the math test first revealed an activation of math-gender stereotypes. We tested a contrast assuming stereotype activation only in the experimental condition "stereotypic test description + math test first", but not in the other three combinations of the two independent variables "test description" and "order of math test vs. GNAT". Contrast weights were specified as 3 -1 -1 -1, and the contrast reached significance, F (1,124) = 12.81. Participants in the critical experimental condition showed, on average, a GNAT effect of $M_{critical}$ = 0.25, whereas the mean GNAT effect in the other three conditions taken together was M_{other} = -0.02. In a second step, we tentatively investigated the contrast separately for women with nonmath vs. math-intensive majors. The contrast reached significance both for women studying nonmath majors, F (1,97) = 6.67, $M_{critical}$ = 0.24; M_{other} = 0.02, and for women studying math-intensive majors, F (1,23) = 9.93, $M_{critical}$ = 0.28; M_{other} = -0.23. Thus, both women with nonmath and math-intensive majors revealed *math-male* stereotyping after reading the stereotypic test description and additionally working on the math test. Further, women with math-intensive majors seemed to reveal, on average, counterstereotypical *math-women* associations in the other three experimental conditions.

4.1.2.3 Math test performance

A 2 (test description) x 2 (order of math test vs. GNAT) ANOVA on the number of correctly solved items in the math test (word problems and matrices taken together) revealed no effects (all Fs < 1.72). For exploratory reasons, we included the type of major (nonmath vs. math-intensive) as a between-subjects factor in the analysis. The 2 x 2 x 2 ANOVA revealed as the only effect an interaction test description x type of major, F (1,120) = 8.07, R^2_p = .06 (all other Fs < 1.10). Simple main effects of test description within type of major revealed a significant stereotype threat effect for women with nonmath majors, F (1,120) = 7.52, R^2_p = .06. Women studying nonmath majors solved more items (47%) after reading the non-stereotypic than the stereotypic (38%) test description.[4] However, women studying math-intensive majors revealed a tendency to solve more items (51%) after receiving the stereotypic than the non-stereotypic (35%) test description, F (1,120) = 3.42, p = .07.[5]

[4] For women studying nonmath majors, exploratory analyses revealed that the observed stereotype threat effect was based on performance differences in the matrices (52% vs. 66% solved matrices in the stereotypic vs. the non-stereotypic test description condition). A 2 (test description) x 2 (order of math vs. GNAT) ANOVA on correctly solved matrices yielded a main effect of test description, F (1,97) = 8.93, R^2_p = .08 (all other Fs < 1). Women studying nonmath majors solved only few word problems (overall mean M = 16.58%) without any test description effects, F < 1.

[5] The slightly better performance of women with math-intensive majors after the stereotypic than the non-stereotypic test description was based on the word problem performance (27% vs. 15% correct after the stereotypic vs. the non-stereotypic test description). A 2 x 2 ANOVA revealed a main effect of test description, F (1,23) = 7.33, R^2_p = .24. Additionally, a main effect of order of math test vs. GNAT could be observed with a better performance if the GNAT

For women studying nonmath majors, we tested whether the stereotype activation captured by the GNAT was related to the observed performance decrements in the math test. Only women starting with the math test were included in the analyses, as this order appeared to be one precondition for the math-gender stereotype activation. Women studying a math-intensive major and starting with the math test in our experiment were too few to allow any conclusions. However, a partial correlation controlling for the task order within the GNAT did not reveal any relations between implicit stereotyping and the number of solved math test items, partial $r = -.08$, $N = 47$.[6] Thus, we could not detect a mediation of the stereotype threat effect via the stereotype activation among women with nonmath majors.

4.1.2.4 Explicit measures

The explicit math ability self-concept, math identification, the rating of math importance for the own major, and the explicit *math-male* stereotype did not vary as a function of the test description (all *t*s < 1). Women studying math-intensive vs. nonmath majors did not differ in their explicit *math-male* stereotypes ($t < 1$). However, women with math-intensive majors revealed a higher explicit math ability self-concept, $t\ (126) = 3.99$, $R^2_p = .11$, a higher math identification, $t\ (126) = 3.67$, $R^2_p = .10$, and higher ratings of math importance for their major, $t\ (126) = 6.32$, $R^2_p = .24$, than women studying nonmath majors.

---

was completed first, $F\ (1,23) = 4.31$, $R^2_p = .16$ (all other *F*s < 1). No effects of test description could be observed in the matrices ($F < 1$), and women with math-intensive majors solved, on average, 55% of the matrices.

[6] The partial correlation between the GNAT effect and the number of solved matrices, controlling for the task order within the GNAT, did not reach significance, either, partial $r = -.05$, $N = 47$.

4.2 Discussion

We investigated boundary conditions for the activation of math-gender stereotypes in women. The situational factors under scrutiny were (i) a stereotypic vs. non-stereotypic test description and (ii) the exposure to a demanding math test. Women studying nonmath and math-intensive majors showed a stereotype activation measured by a math-gender GNAT only after both a stereotypic test description and an exposure to the math test before completing the GNAT. Further, female students of nonmath majors performed worse in the math test after receiving the stereotypic rather than the non-stereotypic test description, showing a stereotype threat effect. However, that performance decrease was not related to the stereotype activation captured by the GNAT. Women studying math-intensive majors revealed a tendency to perform better in the math test after receiving the stereotypic rather than the non-stereotypic test description, showing a tendency for a reversed stereotype threat effect.

The activation of math-gender stereotypes in women required both a stereotypic test description stressing gender differences in the math test and an exposure to that test. Obviously, reading the stereotypic test description alone was too weak to elicit implicit *math-male* stereotyping. This stereotype activation occurring only under certain circumstances is consistent with the lack of chronically activated *math-male* associations in women (see Chapter 3). Women reading the non-stereotypic test description did not reveal any stereotype activation, even if they worked on the math test before completing the GNAT. In line with numerous stereotype threat effect findings, a test description negating gender differences and thus making gender stereotypes irrelevant had also a protective effect precluding stereotype activation. The present study did not test whether the exposure to a demanding math test alone leads to stereotype activation. However, the stereotype activation would not be inevitable in that case, but could be precluded by an instruction rendering the stereotype irrelevant to the test.

A similar activation of math-gender stereotypes in women or girls might occur not only in the laboratory, but also in everyday educational settings. Math-gender stereotypes might easily become salient to women or girls during math or science tests, for example triggered by the minority status of women in many science or math classes (cf. Inzlicht & Ben-Zeev, 2000; Murphy, Steele, & Gross, 2007). Instructions stressing a gender-fair construction of a test should work protectively outside the laboratory, too, and further research has to identify such effective test instructions.

Interestingly, both women studying nonmath and math-intensive majors were susceptible to the stereotype activation. Women studying math-intensive majors associated *math* with *men* and not with *women* after both the stereotypic test description and the exposure to the math test even though they reported a more favourable math ability self-concept than women with nonmath majors and though they performed slightly better in the math test after the stereotypic test description. This quite robust stereotype activation might indicate that also women with math-intensive majors can be vulnerable to math-gender stereotypes. This idea is supported by strong stereotype threat effects on math performance in highly math-identified women (e.g., Cadinu, Maass, Frigerio, Impagliazzo, & Latinotti, 2003; Spencer et al., 1999).

Women with math-intensive majors performed better after reading the stereotypic than the non-stereotypic test description. Keller (2007) explained such reversed stereotype threat effects by motivational states of the individuals. For example, perceiving a testing situation as a gain/no-gain situation with maximum goals activated, individuals are expected to use approach strategies and be particularly sensitive to positive outcomes. According to the Regulatory Focus Theory (Higgins, 1998), such a self-regulation is based on a so-called promotion focus. In this case, negative stereotypes are supposed to be less central or should even lead to challenge (Keller, 2007). Perceiving a testing situation as a loss/non-loss situation containing high pressure, individuals should be more sensitive to negative outcomes (i.e., acting according to a prevention focus), and negative stereotypes should have a greater impact. Keller (2007) demonstrated that negative stereotypic expectancies led to worse performance if a prevention focus as opposed to a promotion focus was activated. The current experiment might have activated a promotion focus among the participants as all results remained anonymous and no (potentially negative) individual feedback had been promised. Further, women with math-intensive majors might have perceived the stereotypically characterized math test as a challenge.

It would be interesting to investigate math test performance and stereotype activation in a more high-stakes situation. A more self-relevant and consequential math test might elicit a prevention focus (see Keller & Dauenheimer, 2003), and also women with math-intensive majors might reveal stereotype threat effects. It might be particularly intriguing to investigate math-gender stereotype activation in this context. As negative expectancies seem more crucial to situations eliciting a prevention focus (Keller, 2007), stereotype activation might play a larger role for achievement in high-stakes than in less self-relevant testing situations.

4.3 Conclusion

Implicit *math-male* stereotyping could be demonstrated in female university students after they had received a stereotypic math test description and additionally worked on that test. Both women with nonmath and math-intensive majors revealed this stereotype activation so that it seems to be rather robust. However, together with this stereotype activation, both a typical (women studying nonmath majors) as well as a reversed (women studying math-intensive majors) stereotype threat effect on math test performance could be observed. Thus, additional factors (e.g., regulatory focus) have to be considered when investigating effects of math-gender stereotypes on women’s math performance.

5 General Discussion

In this chapter, the main findings of the present studies will be summarized and discussed. Measurement issues as well as theoretical and practical implications of the findings will be outlined. Further, suggestions regarding future research will be presented.

5.1 Summary of findings

Still today, a large gender gap can be observed in math-intensive fields like computer science or engineering (U.S. Department of Education, 2000). Math-gender stereotypes diminish both math performance and math interest in women (Davies et al., 2002; Spencer et al., 1999). As explicit math-gender stereotypes might be easily distorted due to self-presentational concerns or personal egalitarian standards, implicit math-gender stereotypes were assessed in the present studies. As the math-gender gap is rooted in decisions or choices made in school years, implicit math-gender stereotypes were investigated in children (Study 1) and adolescents (Study 1-3). University students participated in Study 4 and 5.

In the first part of this empirical research (Chapter 2), the onset of implicit math-gender stereotypes in elementary school children and relations of implicit math-gender stereotypes with math-related outcomes were investigated. In Study 1, implicit math-gender stereotypes and implicit math identity were assessed in students attending grade 4,7, and 9. IATs were employed as measures of implicit math-gender stereotypes (associations *math-boys* and *German-girls*) and implicit math identity (associations *self-math* and *other-German*) (Nosek et al., 2002b). *German* was used as concept for the verbal domain as it is the common term for the respective school subject. Implicit math-gender stereotypes could be detected already among girls attending grade 4. Girls revealed implicit math-gender stereotypes in all grades, whereas boys, on average, did not show implicit associations *math-boys* and *German-girls* in any grade. Gender differences in implicit math-gender stereotyping reached significance in grade 9. In the math identity IAT, girls showed an implicit affinity to German (*self-German* and *other-math*) already in grade 4. Boys did not reveal any implicit associations with math or German at any age, and the gender difference with girls showing stronger *self-German* and *other-math* associations than boys reached significance in grade 9.

Study 2 served as conceptual replication of the main finding of Study 1 (i.e., stronger implicit math-gender stereotypes in adolescent girls than in adolescent boys). Using paper-and-pencil IATs instead of computerized IATs and the concept label *language* instead of *German*, adolescent girls attending grade 7 and 9 again revealed stronger implicit math-gender stereotypes than boys. Thus, the gender difference in implicit math-gender stereotypes seems to be rather robust. In regression analyses, implicit math-gender stereotypes revealed incremental validity beyond explicit math-gender stereotypes in predicting math-related outcomes only for adolescent girls, but not for adolescent boys. For girls attending grade 7 and 9, stronger implicit math-gender stereotypes were related to a stronger implicit identification with language vs. math, a higher explicit German vs. math ability self-concept, higher enrolment preferences for German over math courses, and better German as compared to math grades. In sum, girls - but not boys - revealed implicit math-gender stereotypes, and these stereotypes were related to their math withdrawal.

In the second part of the research (Chapter 3), implicit math-gender and language-gender stereotypes were assessed separately with GNATs (Nosek & Banaji, 2001). As could be observed in Study 1 and 2, adolescent girls revealed stronger implicit math-gender stereotyping in the IATs than boys, calling for an explanation. Stereotype IAT effects reflect the joint strength of the associations *math-boys* and *language-girls*, and these math- vs. language stereotypes cannot be separated within IATs. In Study 3, implicit math-gender and language-gender stereotypes were assessed with GNATs in 9th graders. In the math-gender GNAT, boys revealed stronger stereotyping than girls. Only boys revealed a *math-boys* association, and girls, on average, did not show any associations. In the language-gender GNAT, girls showed implicit associations *language-girls*, whereas boys revealed counterstereotypic *language-boys* associations. These GNAT results imply a plausible post-hoc interpretation of the larger math-gender stereotype IAT effect in girls than in boys. Girls having, on average, no stereotypic associations regarding math and simultaneously bearing *language-girls* associations should reveal a stereotypic IAT effect. However, in boys' stereotype IAT effects, the stereotypic *math-boys* associations and the counterstereotypic *language-boys* associations should cancel out each other, resulting in a low combined stereotype score.

In Study 4, it was tested whether the pattern of implicit math-gender and language-gender stereotyping obtained with 9th graders could be generalized to adults. Implicit math-gender and language-gender stereotypes were investigated with GNATs in university students with a wide variety of majors. Only men – but not women – revealed an association *math-men*, though this gender difference did not reach significance. Further, women showed stronger *language-women* associations than men. However, the counterstereotypic *language-men* association in men did not reach the conventional level of significance. Thus, the results found in 9th graders (Study 3) could be basically replicated in a sample of university students. At the first sight, it might be surprising to detect no chronically activated *math-male* associations in girls and women. However, the weak implicit associations of math with gender in females can be reconciled. These non-significant associations are the only ones revealing no self-serving associations of the own gender with the academic domains. Though speculative, this finding might indicate the vulnerability of girls and women for math-gender stereotyping.

In the third part of this research program (Chapter 4), boundary conditions for implicit *math-male* stereotyping in women were examined. According to Study 4, women do not seem to have chronically activated *math-male* stereotypes. However, implicit stereotypes are malleable (cf. Blair, 2002) and stereotype activation plays a role in stereotype threat effects (Davies et al., 2002; Steele & Aronson, 1995). Therefore, stereotype activation should be detected in women under certain circumstances. In Study 5, the sample consisted of female university students. Both women studying nonmath and math-intensive majors revealed implicit *math-male* stereotyping in a math-gender GNAT after receiving a stereotypic math test description and additionally completing that difficult math test. Further, women studying nonmath majors revealed a stereotype threat effect in the math test performance, showing a worse performance after reading the stereotypic than the non-stereotypic test description. However, women studying math-intensive majors revealed a (descriptive) reversed stereotype threat effect, performing better after reading the stereotypic than the non-stereotypic test description. Thus, when investigating the effects of math-gender stereotyping on women's math performance, additional factors have to be taken into account. For example, depending on the regulatory focus of an individual, negative stereotypic expectancies might be converted into challenge (Higgins, 1998; Keller, 2007).

5.2 Methodological issues

First, the unique contributions of implicit stereotypes beyond explicit gender stereotyping will be outlined. Second, some consequences resulting from IATs as combined measures will be discussed. In detail, math-gender stereotype IAT effects are based on two academic domains, and this has to be considered when the meaning of these IAT effects is interpreted. In the third part of this section, the interpretation of zero values in IATs and GNATs will be discussed.

5.2.1 What could the implicit stereotype measures reveal beyond the explicit ones?

The question whether the implicit stereotype measures could reveal any information the explicit measures were not able to capture is an important criterion for evaluating the implicit (or any new) measures. Both the math-gender stereotype IAT and the two gender stereotype GNATs were related to a variety of math- and language-related variables, demonstrating their criterion validity. Most importantly, the math-gender IAT revealed even incremental validity beyond explicit math-gender stereotypes predicting math-related outcomes for girls (Study 1 and 2).

As further contributions of the implicit stereotype measures, a variety of gender differences was observed in implicit, but not in explicit gender stereotyping. In Study 1 and 2, strong joint associations *math-boys* and *language-girls* measured with IATs could be observed only in girls, but not in boys, leading to further investigations. However, in an explicit stereotype score computed analogous to IAT effects, both boys and girls endorsed these stereotypes to a similar extent. In the GNATs measuring implicit math-gender and language-gender stereotypes (Study 3), boys revealed stronger implicit *math-boys* stereotypes than girls and even counterstereotypic *language-boys* associations. These gender differences were not mirrored in explicit gender stereotypes. Implicit gender stereotypes measured with GNATs seem to yield interesting self-serving math- and language-associations in boys (and men) and the lack of self-serving math-associations in girls and women. In Study 5, only implicit – but not explicit – *math-men* stereotypes varied as a function of the test description. Thus, the strength of implicit *math-men* associations could serve as an indicator for stereotype activation.

5.2.2 What did the stereotype GNATs reveal beyond the stereotype IAT?

In the adolescents' sample taken from Study 1 and 2, IATs revealed unique predictive power in explaining math withdrawal in girls and thus demonstrated their usefulness as measurement tools. However, when interpreting the meaning of the stereotype IAT effect, one has to consider that it reflects the combined strength of *math-boys* and *language-girls* associations. As the stereotype GNATs in Study 3 could show, boys' implicit associations of math and language with their own gender were obscured in their low IAT effect. Knowing only the low stereotype IAT effects in boys and the larger stereotype IAT effects in girls, this might lead to wrong conclusions, for example, assuming *math-boys* associations only in girls, but not in boys. This assumption would even be plausible as females – but not males – are affected by math-gender stereotype threat. However, the math-gender stereotype GNAT revealed the opposite for both adolescents and adults. Thus, when interpreting IAT effects, conclusions referring to the associations of only one concept should be avoided (cf. Nosek et al., 2005). For investigating associations of one concept with another concept pair, GNATs might be applied.

5.2.3 May we rely on the metrics of our implicit measures?

When concluding that women do not have any implicit *math-male* associations or that boys reveal counterstereotypic *language-boys* associations, zero values of GNAT effects are interpreted. However, one has to keep in mind that zero values in IATs and GNATs are not fully trusted. The absence of response latency differences does not necessarily mean that the associations *math-boys* vs. *math-girls* in GNATs or the association pairs *math-boys*/*language-*

girls and *math-girls/language-boys* in IATs do not differ in strength. Stimulus and label properties can influence IAT effects making zero values difficult to interpret (e.g., Rothermund & Wentura, 2004; Steffens & Plewe, 2001). Stimulus sets of each concept pair used in the current studies were selected to be roughly equal in valence and not differently associated with the other dimension in the IAT or GNAT. For example, a *language*-stimulus like dictation has been avoided due to its negative valence, and thriller or rhetoric might have been too strongly linked to the concept boys or men. However, while it was tried to measure concept associations under avoidance of strong stimulus influences, distortions of IAT effects, for example, due to salience asymmetry in concepts and stimuli could not be ruled out (Rothermund & Wentura, 2001; Rothermund & Wentura, 2004). The same might hold for GNATs though these issues have not yet been investigated in GNATs.

Asendorpf, Banse, and Mücke (2002) have suggested an approach to avoid the ambiguity of mean IAT effects and their zero values. According to their position, IATs might be used only as tools for assessing individual differences. Though the underlying cognitive processes in IATs have to be understood in order to judge IATs' internal validity, the focus should not be on the (potentially ambiguous) mean IAT effects, but only on IATs' capacities to predict outcome variables. Relying only on interindividual or intergoup differences in IAT or GNAT effects appears to be a rather cautious approach. Nevertheless, the conclusions that adolescent girls have stronger implicit math-gender stereotypes than boys, and these stereotypes predict girls' math withdrawal (Study 1 and 2), remain valid even in a more restrained interpretation of IAT effects. In Study 5, women in the critical experimental condition "stereotypic test description + math test first" revealed stronger implicit stereotyping as compared to women in the other experimental conditions. Thus, a more cautious interpretation of GNAT effects would not change this conclusion, either. Interpretations regarding self-serving math- or language associations on the basis of GNAT effects (Study 3 and 4) would be more controversial. Detecting self-serving language-associations in males and postulating the absence of self-serving math-associations in females require the interpretation of mean GNAT effects and their zero values. However, most psychological measures possess an arbitrary metrics leading to insecurities about the meaning of the zero point. As this problem is neither confined to IATs or GNATs nor to difference scores in a broader sense, the zero values of these implicit measures may be interpreted, keeping in mind that such interpretations have to be done cautiously.

5.3 Implicit math-gender stereotypes and math-related outcomes

The present studies investigated relations between implicit math-gender stereotypes and math- and language-related outcome variables. However, causality issues regarding adolescents' implicit math-gender stereotypes and outcome variables have not been explored in the current studies. These questions should be addressed in future research. Adolescent girls revealing a relation between implicit math-gender stereotypes and math-related outcomes might be – even unconsciously – influenced by their implicit *math-boys* and *language-girls* associations. However, a causal relationship in the opposite direction or even a bidirectional influence are also possible. For example, girls might generalize their personal ability self-concepts or school grades to their gender group and reveal, as a consequence, implicit math-gender stereotypes (Nosek et al., 2002b). Further, it might be assumed that girls' implicit math-gender stereotypes also reflect their observations of the math and German grades earned by their classmates. Many studies (e.g., Hannover, 1991; Kimball, 1989) demonstrated that girls earn comparable math grades as boys, but better German grades than boys. These gender differences might be translated into girls' IAT or GNAT effects. This question has not been investigated in the present studies. Nevertheless, this explanation cannot be ruled out, and other classmates' math and German grades are environmental cues that might also contribute to implicit math-gender stereotyping.

Environmental cues combined with personal experience can evoke implicit stereotyping in women, as it was demonstrated by the increased implicit *math-male* stereotyping during stereotype activation (Study 5). It would be interesting to assess the duration of that stereotype activation. Further, repeated stereotype activations might lead to chronical *math-male* associations in women or girls (Gawronski & Bodenhausen, 2006).

In the current studies, relations between implicit math-gender stereotypes and math-related outcomes seemed to be stronger for females than for males. In Study 1 and 2, math-gender stereotype IATs revealed their incremental validity only in the girls' sample. In Study 3 and 4, implicit *math-male* associations showed more relations with outcome variables for females than for males. However, implicit stereotypes measured with IATs and GNATs revealed meaningful correlations – predominantly with their explicit counterparts and also school grades – for boys and men, too. This speaks for the validity of these stereotype measures also in the males' samples. Contrary to the present findings, Nosek et al. (2002b) found relations between implicit math-gender stereotypes and math-related outcomes also for male students. Future research should investigate potential moderator variables that might influence the relations between implicit gender stereotypes and outcome variables. Nosek et al. (2002b) already explored the impact of gender identification (see also Greenwald et al., 2002), and other possible variables might be, for example, stereotype awareness or stigma awareness (Brown & Pinel, 2003).

5.4 Practical implications and future directions

Implicit math-gender stereotyping could be observed already among elementary school girls aged 9 years. Thus, interventions to weaken girls' implicit stereotypes should be implemented already at elementary school. As implicit stereotypes depend on situational cues (see Study 5), teachers and parents should avoid stereotypic hints, particularly in the context of achievement situations. Additionally, it should be investigated how achievement situations in math or science could be shaped in a gender-fair way, corresponding to the non-stereotypic test descriptions applied in stereotype threat research.

Though somewhat speculative, it might be interesting whether learning about the effects of implicit math-gender stereotypes could impair their subtle and detrimental influences on females. For example, stereotype threat effects – which are conceptually related to implicit stereotypes, as these stereotypes can mediate or moderate stereotype threat – could be counteracted by teaching women about this phenomenon (Johns, Schmader, & Martens, 2005). In that study, participants in the standard test condition were introduced that they were to take "a standardized math test for the study of gender differences in mathematics performance" (Johns et al., 2005, p. 176). In the "teaching-intervention condition", participants received the additional comment that women might feel anxious during the math test. Further, that anxiety might be unrelated to females' personal math ability, but simply caused by negative cultural stereotypes. Thus, the female participants learned that they could attribute their test anxiety to math-gender stereotypes, and not to their actual math ability. Whereas men outperformed women in the standard test condition, women performed equally to men in the teaching-intervention condition. Similarly to the attributional processes diminishing stereotype threat effects, women might learn that a part of their uneasiness when, for example, considering math course enrolment could stem from their implicit stereotypes, but not from accurate ability judgments. Women might learn that they have internalized stereotypes existing in their environment and become influenced by these stereotypes even if they do not consciously endorse these beliefs. This approach – if practicable – would not directly aim at reducing (or avoiding) implicit math-gender stereotypes, but impede the influence of these stereotypes on women. However, both pathways seem to be important.

6 Conclusion

Implicit gender stereotypes regarding math and language were detected in children, adolescents, and university students, and these gender stereotypes revealed a series of characteristics. First, implicit math-gender stereotypes measured with IATs were found already in 9-year-old girls who also showed an implicit affinity to German as opposed to math. Thus, implicit cognitions regarding math and language (or the respective school subject, German) can be detected early. Second, implicit math-gender stereotypes revealed unique predictive power in explaining math withdrawal in female adolescents. Showing incremental validity beyond explicit math-gender stereotypes, math-gender stereotype IATs proved to be useful measurement tools.

Third, implicit math-gender and language-gender stereotypes did not only reveal culturally shared stereotypes, but also a group-serving (and, by extension, also self-serving) component. Adolescent boys (and basically also men) revealed self-serving associations of their own gender with both academic domains as measured with GNATs. Girls and women revealed only a strong *language-female* association. The lack of chronically activated math-associations in women and girls is cautiously interpreted as a lack of self-serving or protective associations, signaling their vulnerability to math-gender stereotypes.

Fourth, implicit stereotypes are highly dependent on the environmental context. In women, implicit *math-men* associations could be observed only after a stereotypic math test description and an additional exposure to that test. Explicit *math-men* stereotype measures did not capture this stereotype activation. As this stereotype activation could be observed both in women studying nonmath and math-intensive majors, this activation seems to be rather robust. However, as both a stereotype threat effect and a reversed stereotype threat effect could be observed in the context of that *math-men* stereotype activation, additional factors have to be considered when investigating the effects of math-gender stereotyping on females' math performance.

Future research regarding achievement-related behaviour during childhood and adolescence should benefit from the inclusion of implicit stereotype measures. Variables that mediate or moderate the influence of implicit math-gender stereotypes on math-related outcomes should be identified. Further, interventions aimed at diminishing implicit math-gender stereotypes should be developed, and they will most likely rely on environmental changes like introducing a sufficient number of female math or science role models (cf. Dasgupta & Asgari, 2004).

7 References

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

American Association of University Women. (1999). *Gender gaps: Where schools still fail our children*. New York, NY: Marlowe.

Amthauer, R., Brocke, B., Liepmann, D., & Beauducel, A. (2000). *Intelligenz-Struktur-Test 2000 (Intelligence Structure Test 2000)*. Göttingen: Hogrefe.

Asendorpf, J. B., Banse, R., & Mücke, D. (2002). Double dissociation between implicit and explicit personality self-concept: The case of shy behavior. *Journal of Personality and Social Psychology, 83*, 380-393.

Banaji, M. R., & Hardin, C. D. (1996). Automatic stereotyping. *Psychological Science, 7*, 136-141.

Banse, R., Seise, J., & Zerbes, N. (2001). Implicit attitudes towards homosexuality: Reliability, validity, and controllability of the IAT. *Zeitschrift für Experimentelle Psychologie, 48*, 145-160.

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

Bartel, H., Hylla, E., & Süllwold, F. (1970). *Mathematische Denkaufgaben 10+ (Mathematical thinking tasks 10+)*. Weinheim: Beltz.

BBC News Archive. (1985). Retrieved 29.09.2007, from http://news.bbc.co.uk/onthisday/hi/dates/stories/july/4/newsid_2492000/2492853.stm

Benbow, C. P. (1988). Sex differences in mathematical reasoning ability in intellectually talented preadolescents: Their nature, effects, and possible causes. *Behavioral and Brain Sciences, 11*, 169-232.

Benbow, C. P., & Stanley, J. C. (1983). Sex differences in mathematical reasoning ability: More facts. *Science, 222*, 1029-1031.

Bhanot, R., & Jovanovic, J. (2005). Do parents' academic gender stereotypes influence whether they intrude on their children's homework? *Sex Roles, 52*, 597-607.

Blair, I. V. (2001). Implicit stereotypes and prejudice. In G. B. Moskowitz (Ed.), *Cognitive social psychology: The Princeton symposium on the legacy and future of social cognition* (pp. 359-374). Mahwah, NJ, USA: Lawrence Erlbaum Associates, Inc.

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

Bleier, R. (1988). The plasticity of the human brain and human potential. *Behavioral and Brain Sciences, 11*, 184-185.

Bohner, G., & Wänke, M. (2002). *Attitudes and attitude change*. Hove, UK: Psychology Press.

Bortz, J. (1999). *Statistik für Sozialwissenschaftler (Statistics for the social sciences)*. Berlin: Springer.

Brody, L. E., & Mills, C. J. (2005). Talent search research: What have we learned? *High Ability Studies, 16*, 97-111.

Brown, R. P., & Day, E. A. (2006). The difference isn't black and white: Stereotype threat and the race gap on Raven's Advanced Progressive Matrices. *Journal of Applied Psychology, 91*, 979-985.

Brown, R. P., & Pinel, E. C. (2003). Stigma on my mind: Individual differences in the experience of stereotype threat. *Journal of Experimental Social Psychology, 39*, 626-633.

Bussey, K., & Bandura, A. (1999). Social cognitive theory of gender development and differentiation. *Psychological Review, 106*, 676-713.

Cadinu, M., Maass, A., Frigerio, S., Impagliazzo, L., & Latinotti, S. (2003). Stereotype threat: The effect of expectancy on performance. *European Journal of Social Psychology, 33*, 267-285.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (revised ed.). Hillsdale, NJ, USA: Lawrence Erlbaum Associates, Inc.

Croizet, J.-C., Despres, G., Gauzins, M.-E., Huguet, P., Leyens, J.-P., & Meot, A. (2004). Stereotype threat undermines intellectual performance by triggering a disruptive mental load. *Personality and Social Psychology Bulletin, 30*, 721-731.

Dasgupta, N., & Asgari, S. (2004). Seeing is believing: Exposure to counterstereotypic women leaders and its effect on the malleability of automatic gender stereotyping. *Journal of Experimental Social Psychology, 40*, 642-658.

Davies, P. G., Spencer, S. J., Quinn, D. M., & Gerhardstein, R. (2002). Consuming images: How television commercials that elicit stereotype threat can restrain women academically and professionally. *Personality and Social Psychology Bulletin, 28*, 1615-1628.

Eccles, J. S. (1994). Understanding women's educational and occupational choices. *Psychology of Women Quarterly, 18*, 585-609.

Eccles, J. S., Freedman-Doan, C., Frome, P. M., Jacobs, J., & Yoon, K. S. (2000). Gender-role socialization in the family: A longitudinal approach. In T. Eckes & H. M. Trautner (Eds.), *The developmental social psychology of gender*. New Jersey: Erlbaum Associates.

Eysenck, H. J. (1988). O tempora, o mores! *Behavioral and Brain Sciences, 11*, 189-190.

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition: Their meaning and uses. *Annual Review of Psychology, 54*, 297-327.

Friedman, L. (1989). Mathematics and the gender gap: A meta-analysis of recent studies. *Review of Educational Research, 59*, 185-213.

Frome, P. M., & Eccles, J. S. (1998). Parents' influence on children's achievement-related perceptions. *Journal of Personality and Social Psychology, 74*, 435-452.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin, 132*, 692-731.

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

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review, 109*, 3-25.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology, 74*, 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*, 197-216.

Hannover, B. (1991). Zur Unterrepräsentanz von Mädchen in Naturwissenschaften und Technik: Psychologische Prädiktoren der Fach- und Berufswahl (Underrepresentation of girls in the natural sciences and technology: Psychological predictors of subject choice and for choice of occupation). *Zeitschrift für Pädagogische Psychologie/ German Journal of Educational Psychology, 5*, 169-186.

Hansford, B. C., & Hattie, J. A. (1982). The relationship between self and achievement/performance measures. *Review of Educational Research, 52*, 123-142.

Heyman, G. D., & Legare, C. H. (2004). Children's beliefs about gender differences in the academic and social domains. *Sex Roles, 50*, 227-236.

Higgins, E. T. (1998). Promotion and prevention: Regulatory focus as a motivational principle. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 30, pp. 1-46). San Diego, CA: Academic Press.

Hofmann, W., Gawronski, B., Gschwender, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Personality and Social Psychology Bulletin, 31.*

Hugenberg, K., & Bodenhausen, G. V. (2003). Implicit prejudice and the perception of facial threat. *Psychological Science, 14*, 640-643.

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990a). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin, 107*, 139-155.

Hyde, J. S., Fennema, E., Ryan, M., & Frost, L. A. (1990b). Gender comparisons of mathematics attitudes and affect: A meta-analysis. *Psychology of Women Quarterly, 14*, 299-324.

Hyde, S. J., & Kling, K. L. (2001). Women, motivation, and achievement. *Psychology of*

Women Quarterly, 25, 364-378.

Hyde, S. J., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin, 104*, 53-69.

Inzlicht, M., & Ben-Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychologcial Science, 11*, 365-371.

Ivie, R., & Ray, K. N. (2005). Women in physics and astronomy, 2005. AIP Report Number R-430.02. Retrieved 12.5.2007, 2007, from www.aip.org/statistics

Jacobs, J. E., & Eccles, J. S. (1992). The impact of mothers' gender-role stereotypic beliefs on mothers' and children's ability perceptions. *Journal of Personality and Social Psychology, 63*, 932-944.

Johns, M., Schmader, T., & Martens, A. (2005). Knowing is half the battle: Teaching stereotype threat as a means of improving women's math performance. *Psychologcial Science, 16*, 175-179.

Keller, C. (2001). Effect of teachers' stereotyping on students' stereotyping of mathematics as a male domain. *The Journal of Social Psychology, 14*, 165-173.

Keller, J. (2007). When negative stereotypic expectancies turn into challenge or threat: The moderating role of regulatory focus. *Swiss Journal of Psychology, 66*, 163-168.

Keller, J., & Dauenheimer, D. (2003). Stereotype threat in the classroom: Dejection mediates the disrupting threat effect on women's math performance. *Personality and Social Psychology Bulletin, 29*, 371-381.

Kiefer, A. K., & Sekaquaptewa, D. (2007a). Implicit stereotypes and women's math performance: How implicit gender-math stereotypes influence women's susceptibility to stereotype threat. *Journal of Experimental Social Psychology, 43*, 825-833.

Kiefer, A. K., & Sekaquaptewa, D. (2007b). Implicit stereotypes, gender identification, and math-related outcomes: A prospective study of female college students. *Psychological Science, 18*, 13-18.

Kimball, M. M. (1989). A new perspective on women's math achievement. *Psychological Bulletin, 105*, 198-214.

Köller, O., Daniels, Z., Schnabel, K., & Baumert, J. (2000). Kurswahlen von Mädchen und Jungen im Fach Mathematik: Die Rolle des fachspezifischen Selbstkonzepts und Interesses (Course selection of girls and boys in mathematics: The role of academic self-concept and interest). *Zeitschrift für Pädagogische Psychologie/ German Journal of Educational Psychology, 14*, 26-37.

Kühnen, U., Schiessl, M., Bauer, N., Paulig, N., Pöhlmann, C., & Schmidthals, K. (2001). How robust is the IAT? Measuring and manipulating implicit attitudes of East- and West-Germans. *Zeitschrift für Experimentelle Psychologie, 48*, 135-144.

Lane, K. A., Banaji, M. R., Nosek, B. A., & Greenwald, A. G. (2007). Understanding and using the implicit association test: IV. What we know (so far) about the method. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes: Procedures and controversies* (pp. 59-102). New York: The Guilford Press.

Lienert, G. A. (1964). *Denksport-Test (Brain Twister Test)*. Göttingen: Hogrefe.

Maass, A., & Cadinu, M. (2003). Stereotype threat: When minority members underperform. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 14, pp. 243-275). New York: Taylor and Francis Group.

Marsh, H. W. (1989). Age and sex effects in multiple dimensions of self-concept: Preadolescence to early adulthood. *Journal of Educational Psychology, 81*, 417-430.

Marsh, H. W. (1990). The causal ordering of academic self-concept and academic achievement: A multiwave, longitudinal path analysis. *Journal of Educational Psychology, 82*, 646-656.

Marsh, H. W., & Yeung, A. S. (1997). Causal effects of academic self-concept on academic achievement: Structural equation models of longitudinal data. *Journal of Educational Psychology, 89*, 41-54.

McConnell, A. R., & Leibold, J. M. (2001). Relations among the Implicit Association Test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology, 37*, 435-442.

McFarland, S. G., & Crouch, Z. (2002). A cognitive skill confound on the Implicit Association Test. *Social Cognition, 20*, 483-510.

Mierke, J., & Klauer, K. C. (2003). Method-specific variance in the Implicit Association Test. *Journal of Personality and Social Psychology, 85*, 1180-1192.

Murphy, M. C., Steele, C. M., & Gross, J. J. (2007). Signaling threat: How situational cues affect women in math, science, and engineering settings. *Psychologcial Science, 18*, 879-885.

National Research Council. (2006). *To recruite and advance: women students and faculty in science and engineering*. Washington, D.C.: The National Academies Press.

National Science Foundation. (2006). Women, minorities, and persons with disabilities in science and engineering. Retrieved 14.02.2007, from http://www.nsf.gov/statistics/wmpd/sex.htm

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231-259.

Nosek, B. A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General, 134*, 565-584.

Nosek, B. A., Banaji, M., & Greenwald, A. G. (2002a). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics, 6*, 101-115.

Nosek, B. A., & Banaji, M. R. (2001). The Go/No-go Association Task. *Social Cognition, 19*, 625-666.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002b). Math = male, me = female, therefore math ≠ me. *Journal of Personality and Social Psychology, 83*, 44-59.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the Implicit Association Test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin, 31*, 166-180.

Parsons, J. S., Adler, T., & Meece, J. L. (1984). Sex differences in achievement: A test of alternate theories. *Journal of Personality and Social Psychology, 46*, 26-43.

Ramm, M., & Bargel, T. (2005). *Frauen im Studium. Langzeitstudie 1983 - 2004 (Female students. Long-term study 1983 - 2004)*. Bonn, Berlin: Bundesministerium für Bildung und Forschung (BMBF)/ Federal Ministry of Education and Research.

Raven, J. C. (1962). *Advanced Progressive Matrices, Sets I and II.* London: H. K. Lewis.

Reinhard, M. A., Stahlberg, D., & Messner, M. (in press). Failure as an asset for high-status persons - Relative group performance and attributed occupational success. *Journal of Experimental Social Psychology*.

Rothermund, K., & Wentura, D. (2001). Figure-ground asymmetries in the Implicit Association Test (IAT). *Zeitschrift für Experimentelle Psychologie, 48*, 94-106.

Rothermund, K., & Wentura, D. (2004). Underlying processes in the Implicit Association Test: Dissociating salience from associations. *Journal of Experimental Psychology: General, 133*, 139-165.

Rudman, L. A., & Borgida, E. (1995). The afterglow of construct accessibility: The behavioral consequences of priming men to view women as sexual objects. *Journal of Experimental Social Psychology, 31.*

Rudman, L. A., Greenwald, A. G., & McGhee, D. E. (2001). Implicit self-concept and evaluative implicit gender stereotypes: Self and ingroup share desirable traits. *Personality and Social Psychology Bulletin, 27*, 1164-1178.

Rustemeyer, R. (1999). Geschlechtstypische Erwartungen zukünftiger Lehrkräfte bezüglich des Unterrichtsfaches Mathematik und korrespondierende (Selbst-)Einschätzungen von Schülerinnen und Schülern (Teachers' expectations concerning gender-specific performance in mathematics and the corresponding self-evaluation of male and female students). *Psychologie in Erziehung und Unterricht/ Psychology in Education and Teaching, 46*, 187-200.

Rustemeyer, R., & Jubel, A. (1996). Geschlechtsspezifische Unterschiede im Unterrichtsfach Mathematik hinsichtlich der Fähigkeitseinschatzung, Leistungserwartung, Attribution sowie im Lernaufwand und im Interesse (Gender-related differences in mathematics concerning the subjective evaluation of one's own competence, expectation of performance, attributions, learning effort, and interest in the subject). *Zeitschrift für Pädagogische Psychologie/ German Journal of Educational Psychology, 10*, 13-25.

Schmukle, S. C., & Egloff, B. (2004). Does the Implicit Association Test for assessing anxiety measure trait and state variance? *European Journal of Personality, 18*, 483-494.

Sells, L. W. (1973). High school mathematics as the critical filter in the job market. In R. T. Thomas (Ed.), *Developing opportunities for minorities in graduate education* (pp. 37-39). Berkeley: University of California Press.

Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science, 10*, 80-83.

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

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

Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 34, pp. 379-440).

Steele, J. (2003). Children's gender stereotypes about math: The role of stereotype stratification. *Journal of Applied Social Psychology, 33*, 2587-2606.

Steffens, M. C. (2004). Is the Implicit Association Test immune to faking? *Experimental Psychology, 51*, 165-179.

Steffens, M. C., & Buchner, A. (2003). Implicit Association Test: Separating transsituationally stable and variable components of attitudes toward gay men. *Experimental Psychology, 50*, 33-48.

Steffens, M. C., Kirschbaum, M., & Glados, P. (in press). Avoiding stimulus confounds in Implicit Association Tests by using the concepts as stimuli. *British Journal of Social Psychology*.

Steffens, M. C., & Plewe, I. (2001). Items' cross-category associations as a confounding factor in the Implicit Association Test. *Zeitschrift für Experimentelle Psychologie, 48*, 123-134.

Steffens, M. C., & Schulze-Koenig, S. (2006). Predicting spontaneous Big Five behavior with Implicit Association Tests. *European Journal of Psychological Assessment, 22*, 13-20.

Stipek, D. J., & Gralinski, J. H. (1991). Gender differences in children's achievement-related beliefs and emotional responses to success and failure in mathematics. *Journal of*

Educational Psychology, 83, 361-371.

Teachman, B. A., Gapinski, K. D., Brownell, K. D., Rawlins, M., & Jeyaram, S. (2003). Demonstrations of implicit anti-fat bias: The impact of providing causal information and evoking empathy. *Health Psychology, 22*, 68-78.

Teachman, B. A., Gregg, A. P., & Woody, S. R. (2001). Implicit associations for fear-relevant stimuli among individuals with snake and spider fears. *Journal of Abnormal Psychology, 110*, 226-235.

Tiedemann, J., & Faber, G. (1995). Mädchen im Mathematikunterricht: Selbstkonzept und Kausalattributionen im Grundschulalter (Gender differences in elementary school children's self-concept and attributions in mathematics). *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 27*, 61-71.

U.S. Department of Education. (2000). *Educational equity for girls and women NCES 2000-030*. Washington, D.C.: U.S. Government Printing Office.

Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review, 107*, 101-126.

Yee, D. K., & Eccles, J. S. (1988). Parent perceptions and attributions for children's math achievement. *Sex Roles, 19*, 317-333.

Zemore, S. E., Fiske, S. T., & Kim, H. J. (2000). Gender stereotypes and the dynamics of social interaction. In T. Eckes & H. M. Trautner (Eds.), *The developmental social psychology of gender* (pp. 207-241). New Jersey: Erlbaum Associates.

Appendix

Appendix 1: Stimuli and Questionnaire Items for Study 1 and 2

Table A1. Translations of Concept Labels and Stimuli used in Study 1 and 2.

IAT					
Gender Identity - only Study 1-	*Labels*	Boys	Girls	I	Other
	Stimuli	Boys Son	Girls Daughter	I Me	Other Foreign
Math Identity	*Labels*	Math	German (Language)	I	Other
	Stimuli	Math Computation	German Language	I Me	Other Foreign (them)
Gender Stereotype	*Labels*	Boys	Girls	Math	German (Language)
	Stimuli	Boys Son	Girls Daughter	Math Computation	German Language

Note. Concept labels and stimuli in parentheses were used in Study 2. The Practice IAT in Study 2 consisted of the concept labels trees (stimuli: trees, maple) vs. mushrooms (stimuli: mushrooms, toadstool) and small (stimuli: small, tiny) vs. big (stimuli: big, huge).

Explicit measures

Finally there are some questions. Here you shall state how the following statements apply. There are no wrong or right answers because your opinion counts. Please answer spontaneously and honestly – you answers remain secret and will be used for research purposes only.

	Applies not at all				Applies totally
1. I like math.	1	2	3	4	5
2. I like German.	1	2	3	4	5
3. I am good at math.	1	2	3	4	5
4. I am good at German.	1	2	3	4	5
5. I learn things quickly in math.	1	2	3	4	5
6. I learn things quickly in German.	1	2	3	4	5
7. I would like to drop my math class.	1	2	3	4	5
8. I would like to drop my German class.	1	2	3	4	5
9. I can imagine taking advanced math classes for A-levels. (grade 7 & 9 only)	1	2	3	4	5
10. I can imagine taking advanced German classes for A-levels. (grade 7 & 9 only)	1	2	3	4	5
11. In high school I am going to choose many math classes. (grade 4 only)	1	2	3	4	5
12. In high school I am going to choose many German classes. (grade 4 only)	1	2	3	4	5

Remember: Your answers remain secret!

1. What was your last report grade in math? ________
2. What was your last report grade in German? ________
3. What was your last class test grade in math? ________
4. What was your last class test grade in German? ________

Here are again some statements that are about your opinion.
Answer also here spontaneously and honestly, please – Your answers remain secret!

	Applies not at all				Applies totally
1. Boys are often talented for doing math.	1	2	3	4	5
2. Girls are often talented for doing math.	1	2	3	4	5
3. Boys are often talented for doing German.	1	2	3	4	5
4. Girls are often talented for doing German.	1	2	3	4	5

***Math* is rather a typical:**
Boys' subject ☐ ☐ ☐ ☐ ☐ Girls' subject

***German* is rather a typical:**
Boys' subject ☐ ☐ ☐ ☐ ☐ Girls' subject

What do you think? How would most people judge on math?

***Math* is rather a typical:**
Boys' subject ☐ ☐ ☐ ☐ ☐ Girls' subject

What do you think? How would most people judge on German?

***German* is rather a typical:**
Boys' subject ☐ ☐ ☐ ☐ ☐ Girls' subject

Some data of your person:

I am a: Girl ☐ Boy ☐

Grade: ____

Age: _______

Thank you for your participation!

Appendix 2: Stimuli and Questionnaire Items for Study 3 and 4

Table A2. Translations of Concept Labels and Stimuli used in Study 3 and 4.

GNAT					
Math-Gender Stereotype	*Labels*	Math	Boys (Men)	Girls (Women)	
	Stimuli	Computation	Boys	Girls	Break
		Equation	Son (Men)	Daughter (Women)	School bus (Dorm)
Language-Gender Stereotype	*Labels*	Language	Boys (Men)	Girls (Women)	
	Stimuli	Poem	Boys	Girls	Break
		Composition	Son (Men)	Daughter (Women)	School bus (Dorm)
Academic Self-Concept	*Labels*	I	Language	Math	
	Stimuli	I	Poem	Computation	Break
		Me	Composition	Equation	School bus (Dorm)

Note. Concept labels and stimuli in parentheses were used in Study 4.

Explicit measures

Finally there are some questions. Here you shall state how the following statements apply. There are no wrong or right answers because your opinion counts. Please answer spontaneously and honestly – you answers remain secret and will be used for research purposes only.

	Applies totally				Applies not at all
13. I like math.	1	2	3	4	5
14. I like German.	1	2	3	4	5
15. I am good at math.	1	2	3	4	5
16. I am good at German.	1	2	3	4	5
17. I learn things quickly in math.	1	2	3	4	5
18. I learn things quickly in German.	1	2	3	4	5
19. I am talented for doing math.	1	2	3	4	5
20. I am talented for doing German.	1	2	3	4	5

Remember: Your answers remain secret!

1. What was your last report grade in math? ________
2. What was your last report grade in German? ________
3. What was your last class test grade in math? ________
4. What was your last class test grade in German? ________

Here are again some statements that are about your opinion.
Answer also here spontaneously and honestly, please – Your answers remain secret!

	Applies totally				Applies not at all
5. Boys are often talented for doing math.	1	2	3	4	5
6. Girls are often talented for doing math.	1	2	3	4	5
7. Boys are often talented for doing German.	1	2	3	4	5
8. Girls are often talented for doing German.	1	2	3	4	5

***Math* is rather a typical:**
Boys' subject ☐ ☐ ☐ ☐ ☐ Girls' subject

***German* is rather a typical:**
Boys' subject ☐ ☐ ☐ ☐ ☐ Girls' subject

What do you think? How would most people judge on math?

***Math* is rather a typical:**
Boys' subject ☐ ☐ ☐ ☐ ☐ Girls' subject

What do you think? How would most people judge on German?

***German* is rather a typical:**
Boys' subject ☐ ☐ ☐ ☐ ☐ Girls' subject

	Applies totally				Applies not at all
1. I am familiar with computers.	1	2	3	4	5
2. I often occupy myself with computers.	1	2	3	4	5
3. I like to occupy myself with computers.	1	2	3	4	5

Some data of your person:

I am a: Girl ☐ Boy ☐

Grade: ____

Age: ________

Is German your native language? Yes ☐ No ☐

Did you live only in East Germany or in West Germany so far?

- ☐ Only West Germany
- ☐ West and East Germany
- ☐ Only East Germany

Which school do you attend?

- ☐ Highest school track
- ☐ Intermediate school track

Thank you for your participation!

Appendix 3: Questionnaire Items for Study 5

Explicit measures

Finally you shall indicate how much the following statements apply. Please answer spontaneously and honestly – Your data remain anonymous and will be used for research purposes only.

	Applies not at all				Applies totally
21. It is personally important to me how good I am at math.	1	2	3	4	5
22. I am good at math.	1	2	3	4	5
23. I am gifted for math.	1	2	3	4	5
24. My major is very math-intensive.	1	2	3	4	5

What was the range of your last math grades?
School: __________ University (if applicable): __________

I think math is rather a …
men' subject ☐ ☐ ☐ ☐ ☐ women' subject

Did anything attract your attention at this experiment? Write down your ideas.

Following math problems I already knew:

Some data of your person:
Age: ___________

Subject of Studies: ____

Academic term: ________________________

Sex: ☐ male ☐ female

German is my native language: ☐ yes ☐ no

Please Remember: Did you work on a test condition in which had been…

- ☐ **no** alleged gender differences
- ☐ alleged gender differences

In a moment you will learn more about the background of this experiment. Thank you for your participation!

Summary of findings

Though women have caught up with men on post-secondary education, they are still underrepresented in math-intensive fields like computer science or engineering (U.S. Department of Education, 2000). Math-gender stereotypes have been identified to diminish both math performance and math interest in women (Davies et al., 2002; Spencer et al., 1999). Explicit (or self-report) math-gender stereotypes can be easily distorted due to social desirability concerns or personal egalitarian standards. Thus, implicit math-gender stereotypes have been investigated in the current studies. As main career decisions are made during school years, implicit math-gender stereotypes were assessed in children (Study 1) and adolescents (Study 1-3). University students participated in Study 4 and 5.

In the first part of this empirical research, the onset of implicit math-gender stereotypes in elementary school children and relations of implicit math-gender stereotypes with math-related outcomes were investigated. In Study 1, implicit math-gender stereotypes and implicit math identity were assessed in students attending grade 4,7, and 9. IATs were employed as measures of implicit math-gender stereotypes (associations *math-boys* and *German-girls*) and implicit math identity (associations *self-math* and *other-German*) (Nosek et al., 2002b). *German* was used as concept for the verbal domain as it is the common term for the respective school subject. Implicit math-gender stereotypes could be detected already among girls attending grade 4. Girls showed implicit math-gender stereotypes in all grades, whereas boys, on average, did not show implicit associations *math-boys* and *German-girls* in any grade. Gender differences in implicit math-gender stereotyping reached significance in grade 9. In the math identity IAT, girls showed an implicit affinity to German (*self-German* and *other-math*) already in grade 4. Boys did not reveal any implicit associations with math or German at any age, and the gender difference with girls showing stronger *self-German/other-math* associations than boys reached significance in grade 9.

Study 2 was conducted to replicate the finding of stronger implicit math-gender stereotypes in adolescent girls than in boys. Using paper-and-pencil IATs instead of computerized IATs and the concept label *language* instead of *German*, adolescent girls attending grade 7 and 9 again revealed stronger implicit math-gender stereotypes than boys. Thus, the gender difference in implicit math-gender stereotypes seems to be rather robust. Further, regression analyses were conducted for 7^{th} and 9^{th} graders from Study 1 and 2 taken together. Implicit math-gender stereotypes revealed incremental validity beyond explicit math-gender stereotypes in predicting math-related outcomes only for girls, but not for boys. For girls, stronger implicit math-gender stereotypes were related to a stronger implicit identification with language vs. math, a higher explicit German vs. math ability self-concept, higher enrolment preferences for German over math courses, and better German as compared to math grades. In sum, girls - but not boys - revealed implicit math-gender stereotypes in the IATs, and these stereotypes were related to their math withdrawal.

In the second part of the research, implicit math-gender and language-gender stereotypes were assessed separately with GNATs (Nosek & Banaji, 2001). As observed in Study 1 and 2, adolescent girls revealed stronger implicit math-gender stereotyping in the IATs than boys, and these findings called for an explanation. IAT effects capture the combined strength of the associations *math-boys* and *language-girls*, and these two stereotypes cannot be separated within IATs. In Study 3, implicit math-gender and language-gender stereotypes were assessed with GNATs in 9^{th} graders. Gender differences were observed in both stereotype GNATs. In the math-gender GNAT, boys revealed stronger stereotyping than girls. Only boys revealed a *math-boys* association, and girls, on average, did not show any associations. In the language-gender GNAT, girls showed implicit associations *language-girls*, whereas boys revealed counterstereotypic *language-boys* associations. These GNAT results imply a plausible post-hoc interpretation of the stronger implicit math-gender stereotypes in girls than in boys measured with IATs. Girls having, on average, no stereotypic associations regarding math and simultaneously bearing *language-girls* associations should reveal implicit stereotyping in a math-gender stereotype IAT. However, in boys' stereotype IAT effects, the stereotypic *math-boys* associations and the counterstereotypic *language-boys* associations should cancel out each other, resulting in a low combined stereotype score.

In Study 4, implicit math-gender and language-gender stereotypes were investigated with GNATs in university students of various majors. It was tested whether the findings obtained with 9th graders could be generalized to adults. Only men – but not women – revealed an association *math-men*, though this gender difference did not reach significance. Further, women showed stronger *language-women* associations than men. However, the counterstereotypic *language-men* association in men did not reach the conventional level of significance. Thus, the results found in 9th graders (Study 3) could be basically replicated in a sample of university students. At the first sight, it might be surprising to detect no chronically activated *math-male* associations in girls and women. However, the weak implicit associations of math with gender in females can be reconciled. These non-significant associations are the only ones revealing no self-serving associations of the own gender with the academic domains. Though speculative, this finding might indicate the vulnerability of girls and women for math-gender stereotyping.

In the third part of this research program, boundary conditions for implicit *math-men* stereotyping in women were examined. Girls and women do not seem to have chronically activated *math-male* stereotypes. However, implicit stereotypes are malleable (cf. Blair, 2002) and stereotype activation plays a role in stereotype threat effects (Davies et al., 2002; Steele & Aronson, 1995). Therefore, stereotype activation should be detected in girls and women under certain circumstances. In Study 5, the sample consisted of female university students. Both women studying nonmath and math-intensive majors revealed implicit *math-male* stereotyping in a math-gender GNAT after receiving a stereotypic math test description and additionally completing that math test. Further, women studying nonmath majors revealed a stereotype threat effect in the math test performance, showing a worse performance after reading the stereotypic than the non-stereotypic test description. However, women studying math-intensive majors revealed a (descriptive) reversed stereotype threat effect, performing better after reading the stereotypic than the non-stereotypic test description. Thus, when investigating the effects of math-gender stereotyping on the math performance of girls and women, additional factors have to be taken into account. For example, depending on the regulatory focus of a person, negative stereotypic expectancies can be transformed into challenge (cf. Higgins, 1998; Keller, 2007).

Zusammenfassung

Obwohl Frauen heute im selben Umfang wie Männer höhere Bildungsabschlüsse erreichen, sind sie in mathematisch-technischen Fächern wie Informatik oder Ingenieurswissenschaften immer noch unterrepräsentiert (U.S. Department of Education, 2000). Geschlechterstereotype in Bezug auf Mathematik können sowohl die Mathematikleistung von Frauen als auch ihr Interesse an diesem Fach mindern (Davies et al., 2002; Spencer et al., 1999). Explizite Maße für diese Geschlechterstereotype können leicht verfälscht werden, z.B. um sozial erwünscht oder egalitär zu antworten. Daher wurden in den vorliegenden Studien implizite Geschlechterstereotype erfasst. Da außerdem Berufs- oder Studienfachwahlen während der Schulzeit getroffen werden, wurden implizite Geschlechterstereotype bei Kindern (Studie 1) und Jugendlichen (Studie 1-3) untersucht. An den Studien 4 und 5 nahmen Studierende teil.

Im ersten Teil dieses Forschungsprojektes wurde untersucht, ab welcher Altersstufe implizite Geschlechterstereotype in Bezug auf Mathematik und Sprache nachgewiesen werden können. Außerdem wurde getestet, ob jene impliziten Stereotype mit Selbsteinschätzungen oder Leistungen in diesen Fächern zusammenhängen. In Studie 1 wurden implizite Geschlechterstereotype und die implizite Identifikation mit Mathematik vs. Sprache bei Schülerinnen und Schülern aus den Klassen 4, 7 und 9 untersucht. Implizite Assoziationstests (IATs) wurden als Messverfahren für implizite Geschlechterstereotype (Assoziationen *Mathe-Jungen* und *Deutsch-Mädchen*) und für die implizite Identifikation (Assoziationen *Ich-Mathe* und *Andere-Deutsch*) eingesetzt (Nosek et al., 2002b). Als Konzeptbezeichnung für den sprachlichen Bereich wurde „Deutsch" gewählt, da dies der gebräuchlichste Begriff für das betreffende Schulfach ist. Implizite Geschlechterstereotype in Bezug auf Mathematik und Sprache konnten bereits bei Mädchen in Klasse 4 nachgewiesen werden. Mädchen zeigten in allen untersuchten Klassenstufen implizite Geschlechterstereotype, während bei Jungen in keiner Klassenstufe implizite Geschlechterstereotype beobachtet werden konnten. Dieser Geschlechterunterschied wurde in Klasse 9 signifikant. Im IAT zur Identifikation mit Mathematik vs. Sprache zeigten Mädchen

bereits in Klasse 4 eine implizite Affinität zu Sprache (*Ich-Deutsch*, *Andere-Mathe*). Jungen hingegen zeigten in keiner Klassenstufe implizite Affinitäten zu Sprache oder Mathematik; der Geschlechterunterschied wurde in Klasse 9 signifikant.

Studie 2 diente zur konzeptuellen Replikation des Befundes, dass Mädchen stärkere implizite Geschlechterstereotype aufweisen als Jungen. Es wurden Papier-und-Bleistift-IATs anstelle von computerbasierten IATs eingesetzt, und die Konzeptbezeichnung „Deutsch“ wurde durch „Sprache“ ersetzt. Mädchen aus den Klassen 7 und 9 zeigten wiederum stärkere implizite Geschlechterstereotype als Jungen – dieser Befund scheint damit ziemlich robust zu sein. Für Regressionsanalysen wurden die Daten der Jugendlichen (Klasse 7 und 9) aus beiden Studien zusammengefasst. Implizite Geschlechterstereotype zeigten eine inkrementelle Validität zusätzlich zu expliziten Geschlechterstereotypen bei der Vorhersage mathematikbezogener Kriteriumsvariablen nur für Mädchen, nicht aber für Jungen. Bei Mädchen waren stärkere implizite Geschlechterstereotype mit einer stärkeren impliziten Affinität zu Sprache vs. Mathematik, einem höheren expliziten Fähigkeitsselbstkonzept in Deutsch vs. Mathematik, stärkeren Wahlabsichten für Deutsch- vs. Mathematikkurse und besseren Deutsch- vs. Mathematiknoten verbunden. Damit konnten bei Mädchen implizite Geschlechterstereotype in Bezug auf Mathematik und Sprache nachgewiesen werden, und diese Stereotype zeigten einen Zusammenhang mit einem Rückzug aus der Mathematik.

Im zweiten Teil dieses Forschungsprojektes wurden implizite Geschlechterstereotype in Bezug auf Mathematik und Sprache getrennt erfasst. Dafür wurden Go/No-go Association Tasks (GNATs) verwendet (Nosek & Banaji, 2001). Es sollte geprüft werden, auf welche Assoziationen die stärkeren impliziten Geschlechterstereotype bei Mädchen in den IATs zurückzuführen sind. Die IAT-Effekte der Stereotyp-IATs spiegeln die kombinierte Stärke der Assoziationen *Mathe-Jungen* und *Sprache-Mädchen* wider, und diese beiden Assoziationen können in IATs nicht getrennt untersucht werden. In Studie 3 wurden implizite Geschlechterstereotype in Bezug auf Mathematik und Sprache mit zwei GNATs bei Jugendlichen aus der Klasse 9 erfasst. In beiden Stereotyp-GNATs konnten Geschlechterunterschiede beobachtet werden. Im Mathematik-GNAT zeigten Jungen stärkere Geschlechterstereotype als Mädchen. Nur Jungen zeigten *Mathe-Jungen* Assoziationen, während Mädchen im Durchschnitt keinerlei stereotype Assoziationen zeigten. Im Sprache-GNAT zeigten Mädchen die Assoziationen *Sprache-Mädchen*, während Jungen sogar kontrastereotype *Sprache-Jungen* Assoziationen aufwiesen. Diese GNAT-Resultate liefern eine plausible post-hoc Erklärung für die Geschlechtsunterschiede im Stereotyp-IAT. Wenn Mädchen im Durchschnitt keinerlei Assoziationen *Mathe-Jungen*, wohl aber stereotype *Sprache-Mädchen* Assoziationen haben, so sollten sie auch insgesamt einen stereotypen IAT-Effekt zeigen. Bei Jungen sollten sich jedoch stereotype *Mathe-Jungen* Assoziationen und kontrastereotype *Sprache-Jungen* Assoziationen gegenseitig ausmitteln, und Jungen sollten damit einen niedrigen IAT-Effekt aufweisen.

In Studie 4 wurden implizite Geschlechterstereotype in Bezug auf Mathematik und Sprache mit GNATs bei Studierenden aus unterschiedlichen Fachbereichen untersucht. Es sollte geprüft werden, ob die Befunde aus der Jugendlichen-Stichprobe bei Erwachsenen repliziert werden können. Nur Männer zeigten *Mathe-Männer* Assoziationen, obwohl der Geschlechtsunterschied nicht signifikant wurde. Frauen zeigten stärkere *Sprache-Frauen* Assoziationen als Männer. Die *Sprache-Männer* Assoziation bei Männern erreichte allerdings nicht das gewünschte Signifikanzlevel. Damit konnten die Resultate der Jungendlichen-Stichprobe weitestgehend repliziert werden. Es mag überraschen, dass keine impliziten *Mathe-männlich* Assoziationen bei Mädchen und Frauen nachgewiesen werden konnten. Allerdings könnte man die (im Durchschnitt) sehr schwachen Assoziationen zwischen Mathe und Geschlecht bei Mädchen und Frauen so interpretieren, dass es die einzigen Fälle sind, in denen keine selbstwertdienlichen Assoziationen *Fach-eigenes Geschlecht* vorliegen. Möglicherweise könnte dies ein Indikator dafür sein, dass Mädchen und Frauen besonders vulnerabel für Geschlechterstereotype in Bezug auf Mathematik sind.

Im dritten Teil dieses Projektes (Studie 5) wurde untersucht, unter welchen Bedingungen Frauen eine implizite Assoziation *Mathe-Männer* zeigen, zumal diese Assoziation bei Frauen nicht chronisch aktiviert zu sein scheint. Im Allgemeinen sind implizite Geschlechterstereotype kontextsensitiv (siehe Blair, 2002), und Stereotypaktivierung spielt auch bei *Stereotype Threat* Effekten eine Rolle (Davies et al., 2002; Steele & Aronson, 1995). Demnach ist es plausibel, dass Frauen unter bestimmten

Bedingungen eine Stereotypaktivierung zeigen sollten. Sowohl Studentinnen aus mathematikfernen als auch aus mathematiklastigen Studienfächern zeigten eine Aktivierung des *Mathe-Männer* Stereotyps in einem GNAT, wenn sie zuvor eine stereotype Beschreibung eines Mathematiktests gelesen und diesen Test anschließend bearbeitet haben. Frauen aus mathematikfernen Fächern zeigten außerdem einen *Stereotype Threat* Effekt im Mathematiktest. Sie lösten weniger Aufgaben bei einer stereotypen als bei einer nicht-stereotypen Testbeschreibung. Studentinnen aus mathematiklastigen Fächern zeigten jedoch einen umgekehrten *Stereotype Threat* Effekt. Sie lösten – zumindest deskriptiv – mehr Aufgaben bei einer stereotypen als bei einer nicht-stereotypen Testbeschreibung. Damit muss man noch zusätzliche Faktoren berücksichtigen, wenn man den (potentiellen) Einfluss von Geschlechterstereotypen auf die Mathematikleistung untersuchen will. Mögliche Variablen sind z.B. motivationale Zustände einer Person, welche negative stereotype Erwartungen in Herausforderungen umwandeln, z.B. der *regulatory focus* einer Person (siehe Higgins, 1998; Keller, 2007).

Lebenslauf

Petra Jelenec

Persönliche Daten:
Geburtstag: 07.04.1975
Geburtsort: Bergneustadt
Familienstand: ledig
Nationalität: slowenisch

Ausbildung:

10/94 – 04/01	Studium der Psychologie an der Universität Trier Abschluss: Diplom
08/85 - 05/94	Wüllenwebergymnasium Bergneustadt Abschluss: Abitur
09/81 - 06/85	Kath. Grundschule Bergneustadt

Berufliche Tätigkeit:

04/05 – heute	Wissenschaftliche Mitarbeiterin in der Abteilung für Kognitive Psychologie und Soziale Kognition (Prof. Dr. M.C. Steffens) an der Friedrich-Schiller-Universität Jena
01/03 – 01/05	Projektmitarbeiterin im DFG-Projekt „Modellbasierte Analysen von Kontextanbindung und Integration als Determinanten der Erinnerungsleistung an ausgeführte Aktivitäten und Handlungen“ an der Universität Trier (Leitung: Prof. Dr. M.C. Steffens)
01/02 – 12/02	Tätigkeit als diplomierte wissenschaftliche Hilfskraft in der Abteilung für Allgemeine und Kognitive Psychologie an der Universität Trier und Inhaberin eines Promotionsstipendiums der Landesgraduiertenförderung Rheinland-Pfalz
04/01 – 12/01	Tätigkeit als diplomierte wissenschaftliche Hilfskraft in der Abteilung für Allgemeine und Kognitive Psychologie an der Universität Trier

Jena, den 11. Januar 2008 Petra Jelenec

Ehrenwörtliche Erklärung

Hiermit erkläre ich, dass mir die Promotionsordnung der Fakultät für Sozial- und Verhaltenswissenschaften an der Friedrich-Schiller-Universität Jena bekannt ist.

Weiterhin erkläre ich, dass ich die vorliegende Dissertation selbst und ohne unzulässige Hilfe Dritter angefertigt habe. Keine weiteren Personen waren bei der Auswahl und Auswertung des Materials beteiligt. Alle benutzten Hilfsmittel und Quellen sind in der Arbeit angegeben.

Ich habe weder die Hilfe eines Promotionsberaters in Anspruch genommen, noch haben Dritte unmittelbar oder mittelbare geldwerte Leistungen von mir für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der Dissertation stehen.

Die Arbeit wurde weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt. Ich habe weder früher noch gegenwärtig an einer anderen Hochschule eine Dissertation eingereicht.

Ich versichere, dass ich nach bestem Wissen und Gewissen die Wahrheit gesagt habe und nichts verschwiegen habe.

Jena, den 11. Januar 2008 Petra Jelenec