How to Improve Women's Performance in Physics through Instructing Stereotype Threat

Regina Vollmeyer (R.Vollmeyer@paed.psych.uni-frankfurt.de)
Institute for Psychology, Goethe-University, PO Box 111932
60054 Frankfurt/Main, Germany

Anita Püttmann (Puettmann@paed.psych.uni-frankfurt.de)
Institute for Psychology, Goethe-University, PO Box 111932
60054 Frankfurt/Main, Germany

Margarete Imhof (Imhof@uni-mainz.de)
Psychological Institute, Gutenberg-University, Staudinger Weg 9
55099 Mainz, Germany

Abstract

Research into learning physics has repeatedly demonstrated lower motivation and poorer performance for female students than for male students. To attempt to reduce gender differences in strategy use, flow, and performance we used a stereotype threat manipulation. In a 2 x 2-design study (instruction x gender) with 37 11th grade students (20 female) we tested two groups: A control group and a stereotype information group who were told the stereotype was invalid. Both groups had to study a learning program in physics on torque. Pre-tests included prior knowledge and initial motivation. We recorded online exploration behavior to find strategy indicators. After learning, the students took a knowledge test. The results were consistent with our hypothesis: Female students in the stereotype information group reported a higher probability of success compared to females in the control group; they employed more effective strategies, experienced stronger flow, and demonstrated more knowledge. Females in the stereotype information group did not differ from males in either group.

Keywords: gender; learning; motivation.

Introduction

Not only in everyday life but also in empirical research there are many reports that female students perform worse than male students in mathematics and physics when comparing grades or test results (Frey, Asseburg, Carstensen, Ehmke, & Blum, 2007). When learning outcomes differ, the variables which account for this may be lying early in the learning process. Such variables for example could be interest (Females are less interested in mathematics and physics than men, e.g., Fredricks & Eccles, 2002) or self-concept (Females believe that they are not good at mathematics and science; Schütte, Frenzel, Asseburg, & Pekrun, 2007). Since Steele and Aronson's (1995) introduction of the concept *stereotype threat* another possible explanation has been that women are aware of the prevalent stereotype asserting that they have low aptitude for physics. This then leads them to perform poorly due to fear of confirming that stereotype. The aim of our research is to find further evidence as to which variables might explain gender differences in physics performance.

Stereotype Threat

In this study we will focus on the phenomenon of stereotype threat as an explanation for gender-related performance differences in physics. Schmader, Johns and Forbes' (2008) process model of how stereotype threat affects performance can explain why groups perform more poorly if such a stereotype is activated. They applied their model to examples such as that African-Americans achieved lower scores in intelligence tests when they had been told that the test was diagnostic of their intelligence (Steele & Aronson, 1995). Schmader et al. argue that all situations of stereotype threat activate three core concepts: the concept of one's in-group, the concept of the ability domain in question, and the self-concept. If all three core concepts are in balance (Heider, 1958) (e.g., My group, who are females, is good at physics; I am like other females; I am good at physics) there should be no threat. However, as soon as an imbalance occurs (e.g., My group, who are females, is poor at physics; I am not like other females; I am good at physics), the individual has to resolve the imbalance. According to Schmader et al. this imbalance creates negative thoughts which absorb capacity from working memory and lead to performance decrement.

However, once the group with the negative stereotype is told that the stereotype is wrong, the imbalance should disappear. To test this, Johns, Schmader and Martens (2005) told female subjects about that their anxiety during a mathematics test may be the result of a negative stereotype and had nothing to do with their ability. This manipulation lead to gender differences in mathematics disappearing. We will use the same manipulation to attempt to reduce gender differences in a physics task.

Learning Physics with a Computer Program

Females not only perform worse than males in physics, but it is also claimed in learning with computers. However, empirical results are inconsistent. Roy, Taylor, and Chi (2004) found that male students retrieved more task-relevant information in an online task than female students. In a formatting task, Shapka and Ferrari (2003) could not find gender differences. Schaumburg (2004) reported gender differences in knowledge about standard software if explicit computer instruction was missing, but not after explicit instruction. In our own study (Imhof, Vollmeyer, & Beierlein, 2007) students had to redesign a Power Point presentation. In this task male students could reconstruct more features of the presentation than female students. Female students who had to solve statistical problems with an unknown statistics program solved fewer problems than male students (Vollmeyer & Imhof, 2007). As a consequence of our previous work, we wanted to study gender differences in domains in which males outperformed females, such as physics and computer learning. Through the information that a negative stereotype against women exists we wanted to reduce the gender differences. However, we wanted to strengthen the effect through the instruction to work against the stereotype threat.

To investigate what female students exactly do that is different to male students, Püttmann (2008) asked students to learn using a physics program on a computer. Compared to a standard learning text a computer program offers the advantage that we can track the learning behavior with a spy program and thus can describe students' strategies to learn the content of the physics program. Another advantage is that we can enrich the program with interactive graphics. A meta-analysis by Höffler and Leutner (2005) demonstrated that there is a medium positive effect of animation use on learning performance.

In her own study, Püttmann (2008) collected variables relevant to the cognitive-motivational process model (Vollmeyer & Rheinberg, 1999, 2000) to describe possible gender differences when learning with the physics program. She found that male students reported a more positive motivation before they began to learn (stronger interest, higher probability of success) and also more flow-experience during learning. In addition, their way to use the program and its interactive graphics differed from the females' use: Male students used the graphics earlier, more frequently, and for longer periods of time. An analysis of quality revealed a more effective use of the graphics for male students. As a consequence, Püttmann found a strong effect of gender on the final knowledge test. With the results of this study in mind we predicted that these gender differences will be reduced through a stereotype threat instruction.

What we do not know is how such an instruction affects male students. When male students read that there is a stereotype claiming that males do better in physics, a stereotype lift might occur. Although such an effect has seldom been shown to be significant, Walton and Cohen (2003) confirmed such a stereotype lift in a meta-analysis. Therefore, we assumed that our instruction will also slightly improve males' performance compared to the not instructed males. However, our main focus is on the females' learning.

Predictions

The following hypotheses were stated for our study:

Hypothesis 1: Female students who learn that it is only a stereotype that women perform worse in physics than men have a more positive initial motivation than the female control group. Specifically, we assumed their belief in probability of success would increase. As the value or the attractiveness of the task is not changed, interest in the task should remain the same.

Hypothesis 2: Females in the stereotype group experience more flow during learning than the female control group. This may be due to the higher probability of success assessment.

Hypothesis 3: Females in the stereotype group will have a higher probability of success and therefore use the interactive graphics sooner, longer and more efficiently than the female control group.

Hypothesis 4: Females in the stereotype group acquire more knowledge with the help of the physics program than the female control group.

The hypotheses are only explicated for the women in our population. We assume we would reduce the gender effect for females but we leave open whether we can make it disappear. Methodologically, we also would test null hypotheses if we propose that females in the stereotype group perform as well as males in the control group. However, in the Results we will report comparisons between males and females in the stereotype group.

Regarding the comparison between the male stereotype group and the male control group, we expected a small, non significant increase in performance as mentioned earlier as being stereotype lift.

Method

Participants

Thirty seven 11^{th} grade students (20 female, 17 male, age: *M* = 16.9 years old) from two high schools, participated in the study. The schools are situated in a small town near Frankfurt, Germany, with a socioeconomically well-to-do population. In our 2 x 2-design we had four groups: the female stereotype group (*n* = 11), the female control group (*n* = 9), the male stereotype group (*n* = 9), and the male control group (*n* = 8). As this study is regarded as a preliminary study we used small groups. As a consequence we have only small statistical power and expect hardly significant results. To reflect the magnitude of our results we will provide Cohen's d (1992) as recommended by APA (2001). An effect size d > .20 is a small effect size, d > .50 is a medium effect size, and d > .80 is a large effect size.

Procedure

Participants had to study a computer-based physics program on torque for thirty minutes. The computer program contained five units with 12 interactive graphics (Wünscher & Ehmke, 2002). None of the students were familiar with the concept of torque, as this is not taught before the 11^{th} grade. The instruction for the stereotype groups included information about the stereotype threat. After instruction but before beginning to work with the program, prior knowledge and initial motivation (interest, challenge, probability of success, anxiety, QCM, Rheinberg, Vollmeyer, & Burns, 2001) were measured. We videotaped navigation behavior while using the program to find indicators for strategy. During learning, the students' flow-experience was measured after each unit (FKS, Rheinberg, Vollmeyer, & Engeser, 2003). After learning was completed, we administered a knowledge test.

Material

Stereotype Threat Instruction. The independent variable in our design was the stereotype threat instruction as used in Johns' et al. (2005). After the task instruction we added the following sentence for the stereotype group: "It is important to keep in mind that if you are feeling anxious while working with the program, this anxiety could be the result of these negative stereotypes that are widely known in society and have nothing to do with your actual ability to do well on the test." We added that the women should make a special effort to work against these stereotypes, but male students should, of course, give their best as well.

Prior Knowledge on Torque. To control for prior knowledge, we chose four items out of twelve from the knowledge test which we presented at the end of the learning session (see *performance* below). We measured prior knowledge after the instructions had been given. The final knowledge test was developed by Wünscher and Ehmke (2002). The four selected items differed in difficulty.

Initial Motivation. After reading the instructions, participants completed the QCM (*Q*uestionnaire of *C*urrent *M*otivation, by Rheinberg, et al., 2001). This questionnaire measures initial motivation on four factors. The answer format is a seven-point scale.

(1) *Probability of success* is an aspect of motivation that has been incorporated into models as early as Lewin, Dembo, Festinger, and Sears (1944) and Atkinson (1957), as well as being part of more recent theories such as Bandura's self-efficacy construct (1986), Anderson's ACT-R theory (1993) and Wigfield and Eccles's Expectancy-Value Model (2002). It is assumed that learners, at least implicitly, calculate the probability of success taking into account their ability and the perceived difficulty of the task (example items: "I think I am up to the difficulty of the task", "I probably won't manage to do this task").

(2) *Anxiety* can be partly interpreted as fear of failure in a specific situation (Atkinson, 1957). This aspect is not the opposite of high probability of success, as it can be high for learners who are in a social situation in which they do not want to fail even though they expect to succeed (example items: "It would be embarrassing to fail at this task", "I feel petrified by the demands of this task").

(3) *Interest* means that the content to be learned is important for a learner (e.g., Krapp, Hidi, & Renninger, 1992). If learners are interested they have positive affects and positive evaluations regarding the topic (example items: "After having read the instruction the task seems to be very interesting to me", "For tasks like this I don't need a reward, they are lots of fun anyhow.").

(4) *Challenge* assesses whether learners accept the situation as an achievement situation in which they want to succeed (example items: "This task is a real challenge for me", "If I can do this task, I will feel proud of myself").

Flow. As a motivational construct during learning, we chose *flow* (Csikszentmihalyi, 1975). Flow is a pleasant state, in which the following characteristics occur: (1) a challenge-skill balance, (2) merging of action and awareness, (3) unambiguous feedback, (4) concentration on the task at hand, (5) time transformation, and (6) fluency of action.

To measure flow, participants filled in the FKS (Flow Short Scale, by Rheinberg, et al., 2003) after every unit a student had completed (example items "I am totally absorbed in what I am doing", "I know what I have to do each step of the way."). The scale consists of 10 items on a seven-point scale. Thus, for students who had finished after Unit 3 we had collected three measures, for students who had finished after Unit 4 we had four measures, and so on. To compare students at the moment when they finished we chose the flow measure after each student's individual last unit.

Strategy. To retrieve information on how students worked with the interactive graphics we used the programs StatWin and Screen Virtuoso to videotape the students' learning. For example, we counted the *number of used graphics*, and we measured how long students spent with graphics, that is, *time spent with graphics*.

As these two measures are merely quantitative, Püttmann (2008) developed a category system to assess the quality of the graphic use. She rated each use of an interactive graphic in one of the following categories:
A = not used
B = careful, little use
C = experimental, but meaningful, extensive
D = playful, meaningless.

Püttmann (2008) demonstrated that this category system was reliable (Cohen's κ = .88). She also found that the more students used interactive graphics in terms of category C, the better their overall performance (r = .51). Therefore, to determine the *quality of graphic use* we counted how often students used an interactive graphic extensively and in a meaningful way (Category C).

Performance. Students had to fill in a knowledge test with 12 tasks. These tasks had to be solved in different formats. For some tasks students had to calculate, some needed drawings. As the questions consisted of several parts, the performance maximum score was 70.

Results

Table 1 displays the descriptive statistics for all four groups in our design. Before testing our hypotheses, we checked whether the control group and the stereotype group had the same prior knowledge, because prior knowledge could be a confounding variable. Prior knowledge was measured after the stereotype instruction and, surprisingly, there was no instruction effect, F (1, 36) = .10, p = .75, but a significant instruction by gender interaction effect, F (1, 36) = 13.16, p = .001. Whereas females who were told to work against the stereotype produced more prior knowledge than females in the control group, $t(18) = 3.47$, $p < 0.01$, $d = 1.59$, the opposite was true for males. Males in the stereotype group had fewer correct answers in the prior knowledge test than males in the control group, $t(15) = 1.94$, p = 0,72, $d = 0.95$.

Hypothesis 1. In our first hypothesis we assumed that females in the stereotype group compared with females in the control group would have a more positive initial motivation, especially probability of success should increase. Table 1 presents the descriptive statistics for all four factors of initial motivation. As can be seen, the stereotype group believed more in their success, $t(18) = 2.13$, $p = 0.74$, $d = .95$, and was more challenged, $t(18) = 1.77$, $p = 0.093$, $d = .78$, although not significantly.

With a second test we checked whether our manipulation reduced the gender effect on motivation. Therefore, we tested whether the females in the stereotype group differed from the males in the control group. Results show that women in the stereotype group believed even more in their success than men in the control group did, $t(17) = 2.98$, $p = 0.008$, $d = 1.39$.

According to the stereotype lift assumption, males in the stereotype group have a more positive probability of success than males in the control group. The data, however, revealed a weak but not significant difference, $t(15) = 0.68$, $p = 0.40$, $d = .43$.

Hypothesis 2. In Hypothesis 2 we assumed that females in the stereotype group would have more flow experience (after last unit) than females in the control group. The means in Table 1 and also the statistical test support the hypothesis, $t(18) = 4.30$, $p < 0.001$, $d = 1.91$.

Our stereotype instruction helped women to experience even more flow than the male control group, $t(17) = 2.10$, $p = 0.051$, $d = .99$.

Although initial motivation, that is probability of success, was more positive for the male stereotype group compared with the male control group, flow, was more positive, but not significantly for the male control group, $t(15) = 1.25$, $p = 0.23$, $d = .62$. This result contradicts the stereotype lift assumption.

Hypothesis 3. For strategy use we chose three indicators: number of used graphics, time spent with graphics and quality of graphic use. We first checked, whether they are correlated (see Table 2).

Table 1: Descriptive statistics (M, SD) for all variables separated for gender and instruction (f = female, m = male, con= control group, st = stereotype group).

	Gender	Instruction	M	SD
Prior knowledge	f	con	3.89	1.97
	f	st	7.82	2.89
	m	con	9.63	3.34
	m	st	6.33	3.64
Interest	f	con	4.13	0.86
	f	st	4.44	0.82
	m	con	3.95	0.76
	m	st	3.58	1.25
Challenge	f	con	4.64	1.21
	f	st	5.41	0.71
	m	con	5.19	0.83
	m	st	4.64	0.86
Probability of success	f	con	3.69	1.74
	f	st	5.20	1.43
	m	con	3.31	1.28
	m	st	3.93	1.57
Anxiety	f	con	2.11	1.05
	f	st	2.76	1.03
	m	con	2.63	1.26
	m	st	3.11	1.17
Flow (after last unit)	f	con	3.31	1.31
	f	st	5.54	1.01
	m	con	4.63	0.82
	m	st	4.03	1.09
Number of used graphics	f	con	3.67	4.39
	f	st	5.73	3.55
	m	con	7.88	0.99
	m	st	8.33	1.00
Time spent with graphics	f	con	94	121
	f	st	210	143
	m	con	290	88
	m	st	280	74
Quality of graphic use	f	con	2.22	2.64
	f	st	5.09	3.42
	m	con	5.88	2.23
	m	st	6.11	1.54
Performance	f	con	26.22	6.28
	f	st	35.91	7.80
	m	con	33.88	7.12
	m	st	31.22	6.96

Table 2 demonstrates that all indicators for strategy use are intercorrelated, that means they measure the same construct. According to our hypothesis, we expected that females in the stereotype group used better strategies than females in the control group.

Table 2: Correlations of indicators for strategy use (*r*, *p*).

	2	3
Number of used graphics (1)	.82 <.001	.88 .001
Time spent with graphics (2)		.78 <.001
Quality of graphic use (3)		

In the context of learning a physics program they should use the interactive graphics sooner and more often in a meaningful way. Out of the twelve interactive graphics females in the stereotype group used about six graphics (see Table 1), the control group only four, $t(18) = 1.16$, $p = 0.26$, $d = .52$. Therefore, they spent more time with graphics, $t(18) = 1.93$, $p = 0.070$, $d = .88$. Females in the stereotype group used more graphics in an experimental, but meaningful way, $t(18) = 2.06$, $p = 0.054$, $d = .94$. Despite medium and large effect sizes these differences are not significant. Thus there is some evidence that the stereotype instruction supported females to find and use helpful tools of the program.

The next question was whether the stereotype instruction also helped females to learn in a similar way than the male control group. This question has to be answered negatively (see Table 1): Females in the stereotype group used fewer graphics, $t(17) = 1.65$, $p = 0.12$, $d = -.83$, they spent less time with the graphics, $t(17) = 1.40$, $p = 0.18$, $d = -.67$, however, qualitatively, their use was similar, $t(17) = 0.57$, $p = 0.58$, $d = -.27$.

The stereotype lift hypothesis assumes that males in the stereotype group would use the graphics slightly more effectively than the male control group. As expected, the three indicators showed only small and incoherent effects. Males in the stereotype group used more graphics ($t(15) = 0.95$, $p = 0.36$, $d = .45$) with higher quality ($t(15) = 0.26$, $p = 0.80$, $d = .12$), but they spent less time with the graphics ($t(15) = 0.27$, $p = 0.79$, $d = -.12$) than the control group. All differences were small and not significant.

Hypothesis 4. Finally, the stereotype threat instruction should improve females' performance in the stereotype group compared to the control group. In line with the motivation and strategy results, the data support the hypothesis, $t(18) = 3.01$, $p < .01$, $d = 1.37$. Performance of females in the stereotype group was as high as that of males in the control group, $t(17) = 0.58$, $p = 0.57$, $d = .27$.

The stereotype lift hypothesis claiming that males in the stereotype group performed slightly better than the male control group, was not confirmed, $t(15) = 0.78$, $p = 0.45$, $d = -.38$.

Discussion

The aim of our research was to find evidence whether the observed gender effect when learning with the computer could be reduced through a stereotype threat instruction. The manipulation consisted of informing female and male students that there exists a stereotype that females performed worse than males in physics. They were also told that there was no solid evidence for this stereotype, so that if they felt worried or confused during work with the physics program this had nothing to do with their ability, but was probably due to the stereotype. In addition, we challenged the female participants to actively work against this stereotype. This small manipulation had large effects, however, as there was not enough statistical power, the effects were often not significant.

As a treatment check we tested if initial motivation had improved, especially probability of success. Females in the stereotype group should now believe in their ability and thus they should be confident to understand information about physics which would be presented in the computer program. The results supported the hypothesis that females in the stereotype group were more confident than females in the control group and believed even more in their success than the male students in the control group. This is very encouraging, however, it needs a replication.

The stereotype instruction should not only help females' initial motivation, it should also increase flow-experience during work with the physics computer program. We found the same pattern as for initial motivation: Females in the stereotype group experienced more flow than the females in the control group and even more flow than the males in the control group. Therefore, our manipulation not only reduced the males' initial advantage, it even could reverse the effect, that is, males were less positively motivated.

With regard to the question whether our manipulation could also affect the way in which females work with the program, we considered their usage of adequate strategies when working with the interactive graphics. For the difference between females in the stereotype vs. control group we found weaker and not significant effects than for motivation, but females in the stereotype group still improved their strategies. However, compared to the males in the control group they still did not use the interactive graphics that often.

As females' motivation and their strategies improved through stereotype threat instruction, their performance also increased compared with the females in the control group. The effect was sufficiently large to bring the performance of the female participants at a level with the male participants.

Overall, we found that the stereotype threat instruction had a significant effect on relevant variables contained in the cognitive-motivational process model. However, as we had a small sample size the study needs a replication.

References

APA (2001). *Publication manual*. Washington, DC: American Psychological Association.

Anderson, J. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.

Atkinson, J. W. (1957). Motivational determinants of risk taking behavior. *Psychological Review, 64*, 359-372.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.

Csikszentmihalyi, M. (1975). *Beyond boredom and anxiety*. San Francisco: Jossey-Bass.

Fredricks, J. A., Eccles, J. S. (2002). Children's competence and value beliefs from childhood through adolescence: Growth trajectories in two male-sex-typed domains. *Developmental Psychology*. *38,* 519-533.

Frey, A., Asseburg, R., Carstensen, C.H., Ehmke, T., & Blum, W. (2007). Mathematische Kompetenz [Mathematical competence]. In PISA Konsortium Deutschland (Ed.), *PISA 2006. Ergebnisse der dritten internationalen Vergleichsstudie* (pp. 249-276). Münster: Waxmann.

Heider, F. (1958). *The psychology of interpersonal relations.* New York Wiley.

Höffler, T. N., & Leutner, D. (2005). Instructional animation versus static pictures: A meta-analysis. *Learning and Instruction*, *17*, 722-738.

Imhof, M., Vollmeyer, R., & Beierlein, C. (2007). Computer use and the gender gap: The issue of access, use, motivation, and performance. *Computers in Human Behavior, 23,* 2823-2837.

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

Krapp, A., Hidi, S., & Renninger, K. A. (1992). Interest, learning and development. In K. A. Renninger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 3-25). Hillsdale, NJ: Erlbaum.

Lewin, K., Dembo, T., Festinger, L., & Sears, P. S. (1944). Level of aspiration. In J. McHunt (Ed.), *Personality and the behavior disorders* (Vol. 1, pp. 333-378). New York: Ronald Press.

Püttmann, A. (2008). *Selbst reguliertes Lernen mit Multimedia im Physikunterricht* [Self regulated learning with multimedia in physics]. Unpublished doctoral dissertation, University of Frankfurt, Germany.

Rheinberg, F., Vollmeyer, R., & Burns, B. D. (2001). FAM: Ein Fragebogen zur Erfassung aktueller Motivation in Lern- und Leistungssituationen [A questionnaire to assess current motivation in learning situations]. *Diagnostica, 47*, 57-66.

Rheinberg, F., Vollmeyer, R., & Engeser, S. (2003). Die Erfassung des Flow-Erlebens [The assessment of flow]. In J. Stiensmeier-Pelster, & F. Rheinberg (Eds.), *Diagnostik von Motivation und Selbstkonzept* [Diagnosis of motivation and self-concept] (pp. 261-279). Göttingen, Germany: Hogrefe.

Roy, M., Taylor, R., & Chi, M. T. (2004). Searching for information on-line and off-line: Gender differences among middle school students. *Journal of Educational Computing Research*, *29*, 229-252.

Schaumburg, H. (2004). Laptops in der Schule – ein Weg zur Überwindung des Digital Divide zwischen Jungen und Mädchen [Laptop computers in the classroom – A way to overcome the technological gender gap among students]? *Zeitschrift für Medienpsychologie, 16*, 142-154.

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

Schütte, K., Frenzel, A. C., Asseburg, R., & Pekrun, R. (2007). Schülermerkmale, naturwissenschaftliche Kompetenz und Berufserwartung [Student characteristics, competence in science, and professional expectations]. In PISA Konsortium Deutschland (Ed.), *PISA 2006. Ergebnisse der dritten internationalen Vergleichsstudie* (pp. 125-146). Münster: Waxmann.

Shapka**,** J. D., & Ferrari, M. (2003). Computer-related attitudes and actions of teacher candidates. *Computers in Human Behavior, 19*, 319-334.

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.

Vollmeyer, R., & Imhof, M. (2007). Are there gender differences in computer performance? If so, can motivation explain them? *Zeitschrift für Pädagogische Psychologie, 21*, 251-261.

Vollmeyer, R., & Rheinberg, F. (1999). Motivation and metacognition when learning a complex system. *European Journal of Psychology of Education, 14*, 541-554.

Vollmeyer, R., & Rheinberg, F. (2000). Does motivation affect performance via persistence? *Learning and Instruction*, *10,* 293-309.

Walton, G. M., & Cohen, G. L. (2003). Stereotype lift. *Journal of Experimental Social Psychology*, *39*, 456-467.

Wigfield, A., & Eccles, J. S. (2002). The development of competence beliefs. In A. Wigfield, & J. S. Eccles (Eds.), *Development of achievement motivation* (pp. 91-122). San Diego: Academic Press.

Wünscher, T., & Ehmke, T. (2002): *Drehmomente sehen. Eine Lerneinheit mit interaktiven Geometrie-Modulen für den Physikunterricht* [How to see torque: A teaching unit with interactive geometry modules for Physics instruction]. Universität Kiel: Leibniz-Institut für die Pädagogik der Naturwissenschaften (IPN). URL: http://www.ipn.uni-kiel.de/abt_physik/drehmomente [15.06.2008]