



HAL
open science

Can Sex-Undifferentiated Teacher Expectations Mask an Influence of Sex Stereotypes? Alternative Forms of Sex Bias in Teacher Expectations

Aïna Chalabaev, Philippe Sarrazin, David Trouilloud, Lee Jussim

► **To cite this version:**

Aïna Chalabaev, Philippe Sarrazin, David Trouilloud, Lee Jussim. Can Sex-Undifferentiated Teacher Expectations Mask an Influence of Sex Stereotypes? Alternative Forms of Sex Bias in Teacher Expectations. 2008. hal-00390122

HAL Id: hal-00390122

<https://hal.science/hal-00390122>

Preprint submitted on 3 Jun 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Running head: ALTERNATIVE FORMS OF SEX BIAS IN TEACHER EXPECTATIONS

**Can Sex-Undifferentiated Teacher Expectations Mask an Influence of Sex
Stereotypes? Alternative Forms of Sex Bias in Teacher Expectations**

Aïna Chalabaev, Philippe Sarrazin, David Trouilloud

University J. Fourier of Grenoble , France

Lee Jussim

Rutgers University, USA

Journal of Applied Social Psychology (*in press*)



Submitted: June 22, 2007

Revised: October 11, 2007

Accepted: January 17, 2008

Correspondence concerning this article should be addressed to Philippe Sarrazin,
Laboratoire Sport et Environnement Social, Université Joseph Fourier - UFRAPS, BP53,
38041 Grenoble cedex 9, France.

+33 456.52.11.96 (voice); +33 456.52.11.90 (fax).

E-mail: philippe.sarrazin@ujf-grenoble.fr

Abstract

This research investigated different forms of sex bias in teacher expectations relative to gymnastics performance. First, a laboratory experiment including 163 physical education teachers confirmed that stereotypes favorable to boys may influence teacher expectations in gymnastics. Next, a naturalistic study involving 15 teachers and 422 students showed that teachers expected no sex differences even though girls performed higher than boys. However, this sex bias was due to the reliance on non-diagnostic student personal characteristics favorable to boys rather than on a stereotype per se. These results suggest that egalitarian beliefs may mask a bias in favor of a social group when group differences do really exist, and that sex-biased teacher expectations do not inevitably involve an influence of sex stereotypes.

Keywords: sex stereotypes; teacher expectations; accuracy; bias; performance; physical education.

Can Sex-Undifferentiated Teacher Expectations Mask an Influence of Sex Stereotypes?

Alternative Forms of Sex Bias in Teacher Expectations

Determining whether stereotypes are a source of teacher expectations is particularly important because such research goes to the heart of the question of whether teacher expectations contribute to social problems and inequalities. For example, expectations based on sex stereotypes may lead teachers to expect differences between boys and girls. These sex-biased expectations may result in a sex-biased treatment of students, which may in turn lead to sex inequalities (e.g., Nelson, 2002; Skrypnik & Snyder, 1982). This research will focus on the effects of sex stereotypes on teacher expectations in a domain subject to important sex differences (e.g., French Ministry of Education, 2000): physical education (PE) classes.

The Influence of Sex Stereotypes on Teacher Expectations

Sex stereotypes are people's beliefs about the attributes that characterize men and women (e.g., Deaux & Major, 1987). Although they are usually considered as inaccurate (e.g., Fiske & Neuberg, 1990), they may sometimes be accurate (e.g., Judd & Park, 1993; Jussim, Eccles, & Madon, 1996), that is, they may correspond to the actual group differences on a particular dimension. Research focusing on the effects of sex stereotypes on teacher expectations have been mostly experimental (e.g., Tom, Cooper, & McGraw, 1984; Van Matre, Valentine, & Cooper, 2000; for a meta-analysis, see Dusek & Joseph, 1983), and reported that teacher expectations regarding academic performance (e.g., Tom et al., 1984; Van Matre et al., 2000) favor girls. For example, Van Matre and colleagues (2000) presented teachers with a description of a fictitious student and manipulated student

sex. Teachers were asked to predict students' academic performance, and results showed higher teacher expectations for girls than for boys.

Although this experimental paradigm has a long tradition of use in studies examining the effects of stereotypes on teacher expectations (e.g., Dusek & Joseph, 1983) and is useful to investigate stereotypes effects with a relatively high degree of methodological control (e.g., Van Matre et al., 2000), it has its own limitations. Indeed, in such studies, students' diagnostic information is (1) either not taken into account (e.g., Tom et al., 1984; Van Matre et al., 2000), or (2) maintained constant across social groups (e.g., Darley & Gross, 1983; Duncan, 1976). This restricts the external validity of such designs for several reasons.

First, many studies showed that stereotypes are more powerful when perceivers have little information about targets, and less powerful when perceivers have a great deal of information about targets (see for a review Fiske, Lin, & Neuberg, 1999; see for a meta-analysis Kunda & Thagard, 1996). These results bring into question the generalization of experimental studies that do not provide teachers with relevant student personal information with regard to the dimension being judged (e.g., Tom et al., 1984; Van Matre et al., 2000). Indeed, in the context of the classroom teachers have more personal information about their students, and may thus be less susceptible to relying on stereotypes than in a typical experiment.

Second, the external validity of studies that maintain constant students' personal characteristics is limited because in reality, girls and boys may differ on the stereotyped dimension (e.g., Jussim et al., 1996). The fact that teacher expectations are sex-differentiated does not necessarily mean that they are biased by sex stereotypes, because

they may correspond to the actual sex differences on the dimension being judged (e.g., Jussim et al., 1996; Van Matre et al., 2000). It is important to note that this accuracy occurs at the group level. For example, expectations based on an accurate stereotype that boys are better than girls in sports could be accurate when boys perform better than girls on average. When applied to a particular individual group member, these stereotype-based expectations are likely to be inaccurate because group members are usually variable on a particular dimension.

Very few studies have taken into account the actual sex differences in a natural setting (e.g., Jussim, 1989; Jussim & Eccles, 1992; Jussim et al., 1996; Madon et al., 1998). Jussim and colleagues (1996) and Madon and colleagues (1998) measured several students' background variables like previous standardized test scores and previous year grades in addition to student sex. They reported that teachers expected girls to perform higher than boys in mathematics. But more importantly, these sex-differentiated expectations corresponded closely to the actual sex differences in mathematics. The authors then concluded that teacher expectations were not biased by student sex and were an accurate reflection of the reality.

Comparing the Expected and Real Sex Differences: A More Accurate Index of Bias

These studies highlight the necessity of comparing the expected sex differences to the actual sex differences in order to determine whether sex stereotypes bias teacher expectations (e.g., Jussim et al., 1996; Madon et al., 1998). Within this approach, teacher expectations are considered as biased only when the expected sex differences do not correspond to the actual sex differences. The implications of this approach are that sex-differentiated teacher expectations may be accurate when they correspond to the actual sex

differences, and sex-undifferentiated teacher expectations may be inaccurate when real sex differences do exist. This is clearly a more valid way of determining sex bias than merely examining if teacher expectations are sex-differentiated.

Moreover, this approach allows for the assessment of different forms of stereotype bias (Jussim, 1991; Jussim et al., 1996). On the one hand, when no group difference exists or when the group difference and the stereotype are in the same direction, the typical stereotype effect may be found, leading teachers to exaggerate the real sex difference. On the other hand, the group difference and the stereotype may push in opposite directions. For example, Beyer (1999) found that people generally believe that men have higher college grade point averages (GPA's) than do women, when, in fact, women generally have higher GPA's than do men. As a consequence, the stereotype may lead teachers to significantly underestimate the real sex differences, and in the extreme, teachers may expect no sex difference, even though there really is one. Thus, the lack of expected difference, in this context, may be interpreted as another form of stereotype effect.

This is important because the typical experimental research on the influence of stereotypes on person perception assumes that believing groups differ constitutes stereotyping (see for a review Fiske et al., 1999), and implies that believing that individuals belonging to different groups do not differ is the only way to reach an accurate or valid judgment about those individuals. A few studies have started to question this implicit assumption (e.g., Wolsko, Park, Judd, & Wittenbrink, 2000), and the current paper is aimed at extending this line of research, by arguing ~~It is clear, however,~~ that: 1) sometimes, individuals from groups really do differ; so that 2) in such cases perceiving a difference will be most accurate, and failing to perceive a difference may reflect bias. It is thus

important to understand different ways in which stereotypes bias person perception ~~forms~~ of ~~group bias~~, not only the typical bias leading to group-differentiated teacher expectations when there are no real differences between groups. ~~However, to our knowledge, no study yet has provided empirical support for the existence of other forms of group-biased expectations.~~

The Present Research

The goals of this research were to examine (1) whether sex stereotypes bias teacher expectations similarly in an experimental context (with little student personal information) *versus* a naturalistic context (with ample student personal information), and (2) alternative forms of stereotype bias depending on the degree of inconsistencies between stereotypes and real sex differences.

To address these issues, this research was conducted in the context of PE classes in France. We reasoned that PE is an appropriate context for several reasons. First, PE is subject to important sex differences in grades, boys obtaining generally from one to two more points – on a scale ranging from zero to twenty points – than girls in PE classes (French Ministry of Education, 2000). Moreover, sex stereotypes favorable to boys are still present in sports (e.g., Bowker, Gadbois, & Cornock, 2003; Jacobs & Eccles, 1992; Koivula, 1999). Given that sex stereotypes and sex differences exist in PE classes, this context seems to be appropriate to examine sex stereotypes effects on teacher expectations when real sex differences exist.

Second, although sex stereotypes and sex differences in PE achievement are consistent, they may be inconsistent in particular situations. Indeed, PE is a multifaceted school subject comprising a variety of physical activities that are compulsory for all

students (Wang, Papaioannou, Sarrazin, Jaakkola, & Solmon, 2006). Whereas boys are likely to perform higher than girls in many activities involving strength and speed (e.g., athletics, team sports) (e.g., Thomas & French, 1985), girls may perform as well as boys or better than them in activities that involve expressivity and fine motor skills (e.g., gymnastics, dancing) (e.g., Eccles & Harold, 1991). Therefore, real sex differences and the stereotype about male superiority in sports may be in the *same* direction in some physical activities and in *opposite* directions in other activities. In this research, gymnastics was chosen because this activity may generate sex differences in favor of girls (e.g., Eccles & Harold, 1991). Although some particular gymnastic activities may favor boys because they involve strength (e.g., pommel horse), gymnastics in general requires grace and expressivity and should thus favor girls. As an illustration, 78,4% of people practicing this activity in France are females (MJS Stat-Info, Ministry of Youth and Sports, 2004). Moreover, gymnastics taught in PE classes requires expressivity and fineness because PE teachers generally choose to teach floor gymnastics and acrobatic elements more than exercises requiring strength. In this situation, girls are likely to perform better than boys.

Consequently, real sex differences in gymnastics performance and sex stereotypes associated with sports may be in opposite directions. This allowed us to investigate forms of sex bias different from the typical group-differentiated teacher expectations examined in the literature. In this context, do teachers rely on sex stereotypes when elaborating their performance expectations? Or are they unaffected by a sex stereotype that is contradictory with the actual sex differences?

Before investigating these questions, it seemed necessary to demonstrate that teacher expectations may be affected by the stereotype about male superiority in sports in

the particular context of gymnastics teaching. Indeed, teachers could be aware that sex differences in this activity differ from other activities, and may therefore think that this stereotype is not applicable. To answer this question, Study 1 used the classic experimental paradigm by manipulating student sex while holding constant students' personal characteristics. This methodology was chosen because as noted earlier, its strength is to isolate stereotypes effects by rendering social category orthogonal to personal characteristics. This rigorous manipulation of the variables is useful for identifying whether stereotypes may influence teacher expectations (e.g., Madon et al., 1998).

Next, the second step of this research (study 2) was to investigate the effect of this stereotype on teacher expectations in a naturalistic study that took into account several student personal characteristics (e.g., standardized test scores, previous grades, motivation) in addition to student sex. This methodology was chosen because first, teachers in this context have more personal information and could be less susceptible to stereotype effects; second, counter-stereotypic sex differences may exist in this context, resulting in unusual forms of stereotype bias such as the underestimation of sex differences or even the absence of expected sex differences.

Study 1

The first step of this research was to conduct a laboratory experiment examining whether teacher expectations about students' performance in gymnastics were based on sex stereotypes. Teachers were provided with information about student sex and performance on a gymnastics task. Given that boys' and girls' performance was held constant in this experiment, expectations that differed according to student sex were interpreted as being influenced by sex stereotypes. Stereotypes were expected to influence teacher expectations

in the typical way assessed in experiments, by leading teachers to expect imaginary sex differences. Although the lack of real differences between groups was artificially constructed, the experiment also had two aspects that substantially strengthened its ecological validity: perceivers viewed videos of actual students performing actual gymnastics routines, and genuine PE teachers participated in this study (in contrast, for example, to the prevalent use of college students as perceivers in studies of stereotypes).

A second goal was to assess whether stereotype effects on teacher expectations were moderated by teachers' sex. Given that teachers are members of a sex category (i.e., male or female), they may be differently affected by sex stereotypes. According to the *social identity* theory (e.g., Tajfel & Turner, 1979), individuals may adopt stereotypes that are favorable to their ingroup, in order to have a positive social identity. It was thus hypothesized that stereotypes favorable to boys would be more powerful among male teachers than among female teachers.

Method

Participants

One hundred and sixty three PE teachers (81 women and 82 men; M age = 40.6 years, $SD = 11.4$) teaching in 35 schools voluntarily participated in this experiment.

Design

Student sex was manipulated in a between-group design, so that one boy *or* one girl was presented to the teacher. Sex was indicated by the video and the name of the student (Antoine or Elodie) which was verbally presented by the experimenter and written on the questionnaire.

Criteria

In many situations, including gymnastics performance, there is no gold standard for assessing the quality of a performance. Gymnastics is not like the 100-yard dash or a basketball game, which have objective metrics (speed and points, respectively). The only criterion that exists for determining the quality of gymnastics performance is the judgment of trained experts. Using trained experts is a well-known method well-established in the accuracy literature for maximizing the quality of criteria, when no objective criteria exist (e.g., Jussim, 2005; Ryan, 2002). The present study, therefore, used the judgments of trained experts to select male and female students whose quality of gymnastics performances were identical.

The videos were selected for the experiment from among a set of videos sequences that were evaluated by three gymnastics experts, who had been practicing competitive gymnastics for at least five years. Although experts are not always correct, they are typically far more knowledgeable, experienced, and accurate than are naive judges (Jussim, 2005; Ryan, 2002). A particular concern in this research, however, was the possibility that the experts themselves may have been biased by sex stereotypes. Therefore, precautions were taken to further reduce the possibility of the experts themselves being biased by sex stereotypes: (1) The experts were asked to evaluate the students as accurately as possible. Several studies indicated that when perceivers are motivated to elaborate an accurate impression of the target, they allocate more attention to target's individuating information (e.g., Madon, Gyll, Hilbert, Kyriakatos, & Vogel, 2006), and their judgments are not affected by stereotypes (e.g., Judice & Neuberg, 1998; for a review see Snyder & Stukas, 1999); and (2) the ratings were made in reference to standardized criteria relative to the execution of the task. More particularly, expert judges rated each student on a scale from 1

to 7, each grade referring to a particular position of the body. For example, a grade of “1” referred to a complete absence of straightness of the different parts of the body when passing the vertical line, whereas a grade of “7” referred to a perfect straightness of the body from the vertical line until arriving on the ground. The fact that the standards of judgment were identical for all of the experts limited the phenomenon of *shifting standards* (Biernat & Manis, 1994; Biernat, Manis, & Nelson, 1991)¹. In addition, the experts observed each students’ performance three times. In much the same way as assessing some construct with multiple questions increases the reliability of the questionnaire by reducing random error, this method also reduces random error in the final, overall, ratings.

In order to avoid ceiling or floor effects in teachers’ judgments, we selected students who received a score of 4 by the all three gymnastics experts. Finally, eight students (four girls and four boys) were chosen in order to control potential effects of students’ information other than sex (e.g., physical appearance, clothes) on teacher expectations.

Experimental Task and Procedure

Teachers were run individually in a classroom. After being told that the experiment consisted of predicting the performance of a student in gymnastics, they were given the following instruction: “You are going to watch a student completing a gymnastics task, and you will be asked to predict his/her future behavior (i.e., gymnastics performance), knowing that he/she completed this task at the beginning of a gymnastics course.” It was also said that the student was 13 years old, that he / she was in the 8th grade and had no school delay. Then the participants watched a videotape showing the student completing a gymnastics task (a stretched handstand back-drop) and lasting four seconds. This motor task was chosen because, for many gymnastics experts, stretched handstand is a

fundamental element underlying many gymnastics elements. Past studies reported a good predictive validity of this test, showing that it is positively correlated to gymnastics self-concept ($r_s > .55$) (Chanal, Marsh, Sarrazin, & Bois, 2005; Marsh, Chanal, & Sarrazin, 2006) and PE grade in gymnastics ($r > .64$) (Boiché, Sarrazin, Grouzet, Pelletier, & Chanal, in press). Accordingly, this test seems to be diagnostic of actual and perceived performance in gymnastics.

Teachers' Expectation

After watching the videotape, all participants filled out a questionnaire to indicate their expectations about that students' performance at the end of the 10-week period of gymnastics lessons, in addition to a manipulation check item regarding the students' sex. Teachers' performance expectations were measured by one item, similar to the one used by Jussim and colleagues (1996): "According to you, how good will this student be in gymnastics?". Teachers answered on a 7-point scale, ranging from (1) "very bad" to (7) "very good".

Results and Discussion

Preliminary Analyses

The analyses of the manipulation check item showed that all of the teachers correctly perceived the student sex. However, three teachers (one woman and two men) did not indicate their expectations about the student. They were not included in the following analyses. Before performing the primary analyses, we tested whether there were significant differences among the set of four male students and among the set of four female students. A one-way analysis of variance (ANOVA) examined whether there were differences in responses to the male targets, and another one-way ANOVA examined whether there were

differences in responses to the female targets. No significant main effects were found, $F_s < 1.37$, *ns*, showing that teachers did not expect differences between the four girls and between the four boys. Thus, we combined in a first group the four girls to reflect the “girl target” and in a second group the four boys to reflect the “boy target”.

Main Analyses

To assess the impact of student sex on teacher expectations, a 2×2 (student sex \times teacher sex) ANOVA was performed on performance expectations. The main effect of student sex was significant, $F(1, 156) = 11.66$, $p < .001$, $\eta^2 = .07$, indicating that teachers believed boys ($M = 4.82$, $SD = 0.90$) would perform higher than girls ($M = 4.32$, $SD = 0.99$) in gymnastics. Despite the identical ratings the male and female students received from the experts, teacher expectations were apparently biased by the sex stereotypes about male superiority in sports (e.g., Bowker et al., 2003; Koivula, 1999) and corroborates previous findings showing the influence of this stereotype on parents' perceptions (e.g., Fredricks & Eccles, 2005).

Moreover, neither the main effect of teacher sex [$F(1, 156) = 0.17$, $p > .60$, $\eta^2 = .001$], nor the two-way interaction student sex \times teacher sex [$F(1, 156) = 0.34$, $p > .50$, $\eta^2 = .002$] were significant, male and female teachers expecting the same differences between boys' and girls' performance. This finding does not corroborate our hypothesis based on social identity theory (e.g., Tajfel & Turner, 1979).

The first step of this research was to examine whether the stereotype about male superiority in sports affected teacher expectations regarding students' gymnastics performance, in a laboratory experiment providing a relatively high degree of

methodological control. We found that stereotypes did influence teacher expectations, suggesting that the stereotype about male superiority in sports may be influential even in gymnastics, a sport sex-typed as feminine (e.g., Koivula, 1999; Riemer & Visio, 2003). More particularly, stereotypes manifested their influence in the typical way found in experiments, leading teachers to expect gymnastics differences between boys and girls, even though there were no differences in their actual performance.

This experimental methodology, based on a rigorous manipulation of the variables, enabled us to determine whether sex stereotypes may influence teacher expectations. However, as noted earlier, this paradigm is limited because on the one hand, teachers have less relevant information about students than in a naturalistic context, and on the other hand, it does not take into account the real sex differences that may exist between girls and boys. To overcome this limitation, Study 2 was carried out in a natural educational setting and took into account several student personal characteristics. This allowed us to compare the expected sex differences to the actual sex differences, in order to evaluate the extent of sex bias in the actual educational context.

Study 2

The goal of this study was to investigate the effects of sex stereotypes on teacher expectations regarding students' performance in gymnastics in a naturalistic study. Few studies of the role of sex stereotypes in teacher expectations have been conducted in a natural setting (see Jussim & Eccles, 1995; Jussim et al., 1996; Madon et al., 1998, for exceptions). Contrary to a typical experiment, in this context, girls and boys may differ on dimensions relative to stereotypes. Consequently, in contrast to an experiment, simply demonstrating that teachers expect boys and girls to perform differently is not sufficient to

demonstrate a bias in a naturalistic study (e.g., Jussim et al., 1996). Instead, it is necessary to show that the *expected* sex differences do not correspond to the *existing* sex differences.

To answer this question, this study was based on the *reflection – construction* model (Jussim, 1991; Jussim et al., 1996), which addresses two broad issues related to stereotypes: content and process. The first step of the model consists of examining issues of stereotype content: Do teachers expect sex differences in performance? How accurate are these expected sex differences (or lack of differences)? This involves determining whether teacher expectations of individual students, aggregated across all students in each of two groups, correspond to the actual aggregated differences among the students in those groups. Teacher expectations could be inaccurate in either of two directions: 1) They might overestimate differences between groups (in the extreme, they might expect a difference when none existed); or 2) they might underestimate differences between groups (in the extreme, they might expect no difference, when one existed).

The second step of the model consists of examining issues of stereotype process, that is how teacher expectations became accurate or inaccurate: Did they rely on student sex and/or on students' personal characteristics when judging them? Did such reliance influence the extent to which teacher expectations of sex differences corresponded to the actual sex differences between the groups?

These content and process issues are addressed in term of a conceptual model describing relations between targets' group membership (e.g., student sex), targets' personal characteristics, and teacher expectations (see Figure 1).

In the model, r_1 is the correlation between targets' personal characteristics and their group membership (i.e., male *vs.* female). r_1 is the real group difference. r_2 is the

correlation between targets' group membership and teacher expectations. r_2 is the expected difference. Path A represents the influence of targets' personal characteristics on teacher expectations, and Path B represents the influence of targets' group membership on teacher expectations (i.e., teachers' stereotype).

First, the accuracy of teacher expectations of group differences is assessed with a simple comparison of r_2 to r_1 . When $r_2 > r_1$, the expected group difference exceeds the real group difference, indicating that teachers exaggerated real group differences. When $r_2 < r_1$, the expected group difference is less than the real group difference, indicating that teachers underestimated real group differences. Finally, when $r_2 = r_1$, the expected group difference corresponds to the real group difference, indicating that teacher expectations are accurate.

Next, the model identifies the sources that contributed to the expectations of group differences: students' personal characteristics and/or stereotypes. In this study, students' personal characteristics included performance, perceived competence, motivation in gymnastics, past PE grades and academic achievement, and participation in sports. Students' motivation and perceived competence in gymnastics as well as previous grades in PE were taken into account because previous research clearly showed that this background information serves as a basis of teacher expectations (e.g., Jussim & Eccles, 1992).

In addition to these variables, student sex was included as a predictor of teacher expectations. In terms of the model, targets' personal characteristics influence teacher expectations when Path A $\neq 0$, and stereotypes influence teacher expectations when Path B $\neq 0$. Inaccurate teacher expectations of group differences could be caused by relying on stereotypes but also by relying on inaccurate students' characteristics. On the contrary,

accurate teacher expectations may be caused by relying on accurate students' characteristics and/or accurate beliefs about group differences (stereotypes).

Given that sex stereotypes and real sex differences were expected to push in opposite directions, it was hypothesized that teacher expectations would underestimate the real sex differences (content part of the model). This underestimation would be explained by a reliance on a sex stereotype favorable to boys, represented by a path coefficient (i.e., Path B) significant and in a direction opposite to the zero-order correlation between teacher expectations and student sex (process part of the model).

Method

Participants

Fifteen teachers (10 men and 5 women from 30 to 55 years old) and 422 students (234 boys and 188 girls; M age = 13.58 years; SD = 1.22) from 24 classes (from the sixth to the ninth grade) volunteered to participate in this study.

Procedure

This study was conducted at the first lesson of a 10-week period of gymnastics taught in PE lessons. Students filled out an anonymous questionnaire assessing several background characteristics. Then, they completed a standardized test of gymnastics that was videotaped as an assessment of their performance. Between the first and second gymnastics lessons, teachers filled out a questionnaire measuring their expectations regarding performance for each of their students. Although teachers rated their students after the first gymnastics lesson, teachers had been interacting with their students in PE classes for at least three months.

Measures

Teacher expectations. Teachers indicated their expectations relative to students' performance at the end of the 10-week period of gymnastics lessons using the same item as described in Study 1: "According to you, how good will this student be in gymnastics?". Teachers answered on a 7-point scale, ranging from (1) "very bad" to (7) "very good".

Performance in gymnastics. Three experts in gymnastics evaluated each student performing the same task as in Study 1 on the same 7-point scale ($\alpha = .92$). The same procedures used in Study 1 were used in Study 2 to maximize the validity of the experts' ratings.

Student self-determined motivation for gymnastics. Student motivation was assessed by self-determined motivation (Deci & Ryan, 1985) for gymnastics, which was measured by an adapted French version of the Sports Motivation Scale (SMS, Brière, Vallerand, Blais, & Pelletier, 1995). This questionnaire is composed of seven subscales of the intrinsic/extrinsic motivation continuum developed by Vallerand (1997) and begins by the following sentence: "I participate in gymnastics classes:...". Four items measured the "intrinsic motivation to knowledge" (IMK) and the "intrinsic motivation to accomplishment" (IMA) (e.g., "for the satisfaction I experience while I am mastering my abilities"), four items measured the "intrinsic motivation to stimulation" (IMS) (e.g., "for the excitement I feel when I am really involved in the activity"), four items measured the "identified extrinsic motivation" (IEM) (e.g., "because what I learn in this activity will be useful later"), four items measured the "introjected extrinsic motivation" (INEM) (e.g., "because I would feel guilty if I could not succeed in this activity"), five items measured the "external regulation motivation" (ERM) (e.g., "because that is what I am supposed to

do”), and five items measured “amotivation” (AM) (e.g., “I do not know why I participate in gymnastics, if I could, I would get exempted”). The students answered on a 7-point scale from (1) never to (7) always. In this study, each subscale presented an acceptable internal consistency ($\alpha s > .70$). The average of items of each subscale was then computed. Finally, we decided to use the self-determination index to reduce the number of the variables. This index rests on the existence of a quasi-simplex (ordered correlation) pattern between the seven subscales of motivation (e.g. Li & Harmer, 1996). The index was calculated by giving each subscale a specific weight according to its respective place on the self-determination continuum, multiplying this weight by the score of the subscale, and adding the scores of all subscales to derive a single score. Thus, the scores for the three types of intrinsic motivation were averaged and assigned the highest positive weight (+2) because intrinsic motivation is the highest self-determined form of motivation. Identified extrinsic motivation, a self-determined type of extrinsic motivation, was assigned a lower weight (+1). The scores for external regulation and introjection were averaged and assigned a negative weight (-1), and amotivation, which represents the absence of self-determination, was weighted highly negatively (-2). Support for the validity and reliability of this type of composite index has been obtained in several studies (see Ryan & Connell, 1989; Vallerand, 1997; Vallerand & Fortier, 1998; Vallerand & Losier, 1999).

Perceived competence is defined as the individual’s perception of his or her current competence (e.g., Wigfield & Eccles, 2000). Previous research (Boiché et al., in press; Chanal et al., 2005; Marsh et al, 2006) established the validity of the three items used in the present study to assess perceived gymnastics ability (for example: “I feel that my

competence in gymnastics is:”). Answers were on a 7-point scale ranging from (1) very bad, to (7) very good. In the present study, internal consistency was high ($\alpha = .90$). The items were combined by computing their mean.

Other student personal information. Time spent on participation in sports as an after-school activity, past PE grades and academic achievement were also expected to influence teacher expectations. Indeed participation in sports and PE grades may be interpreted as an indicator of competence in sports, and a “halo” effect of academic achievement on teacher expectations may be expected. *Participation in sports was measured by an item:* “How much time per week outside of school do you spend on sports activities?” and students were divided into six categories: those who did not spend time doing sports, those who spent from 1 to 2 hours doing sports, from 2 to 4, 4 to 6, 6 to 8, and more than 8 hours per week. *PE and academic achievements* were assessed by computing respectively the average of the previous grades in PE during the school year and the average of the grades received in the other school subjects.

Results and Discussion

Nonindependence of Teacher Expectations.

Because teachers rated all of the students in their classrooms, their expectations were not independent of one another. Traditional regression models are limited because they ignore the nesting of individuals within larger groupings, such as classrooms. Not attending to the multilevel nature of data introduces bias into the analyses, by assuming many more degrees of freedom than actually exist. Consequently, the standard errors of the estimated parameters are biased downward. Hierarchical linear models are a more

appropriate analytical approach than are traditional regression models because they take into account the hierarchical structure of data. Thus, in this study, all correlations were rendered independent from students' classrooms with multilevel modelling (Bryk & Raudenbush, 1992) using the MLwiN 1.10 software (Rasbash, Browne, Healy, Cameron, & Charlton, 2001).

Overview of Main Analyses

Two sets of analyses were performed addressing two questions: (a) did teachers expect performance differences between girls and boys? and (b) did boys and girls really differ in their performance?

Second, multilevel models examined the processes leading teachers to their expected sex differences and similarities. These analyses addressed whether teachers relied on stereotypes, students' personal characteristics, or both when arriving at their expectations.

The Accuracy of Teachers' Expectations for Boys and Girls

Table 1 presents descriptive statistics and correlations between teacher expectations, student sex and students' personal characteristics, that were standardized and rendered independent from students' classroom by the estimation of hierarchical linear models. Correlations showed that teachers did not expect sex differences in performance, $r(422) = .07, p > .10$. Next, we examined the actual differences between girls' and boys' performance. The analyses indicated that girls performed significantly higher than boys, $r(422) = -.18, p < .001$.

A test of comparison of correlations (Cohen & Cohen, 1983, p.57) showed that the expected sex differences were significantly different from the real sex differences, $t(419) =$

4.84, $p < .001$. These content analyses indicated that teacher expectations were biased: teachers did not expect sex differences in performance even though girls performed better than boys. Given that initial performance is a strong predictor of performance at the end of the period of lessons (e.g., Trouilloud, Sarrazin, Martinek, & Guillet, 2002), we can conclude that teacher expectations were inaccurate. Interestingly, they were inaccurate not because they created imaginary differences when none existed, nor because they exaggerated real differences. Instead, teacher expectations were inaccurate because they perceived boys and girls as similar, when, in fact, girls outperformed boys.

Sources of Teacher Expectations

The follow-up analyses more deeply probed the processes by which teachers arrived at their inaccurate sex-undifferentiated expectations of performance. To do so, we examined which variables predicted teacher expectations (i.e., students' personal characteristics and student sex).

Predictors of performance. Before investigating the sources of teacher expectations, preliminary analyses determined, among the students' characteristics, which ones could be considered as reliable predictors of students' performance, by examining which student characteristics significantly predicted student performance. Two two-level random intercept models were formulated to examine the predictors of students' performance, with the individual student on level 1, and the class on level 2. As a first step, a Model 0a was estimated, which models variation only in the intercept. This model is useful because it yields a decomposition of the total variability of performance into two parts: one due to the individual, and one due to the class level. Next the model 1a was estimated, including the

potential predictors of performance (i.e., perceived competence, self-determined motivation, participation in sports, PE and academic achievement, sex).

Results showed that the reduction of the deviance of Model 1a, compared with Model 0a, was significant ($\Delta = 158.3$ with 6 degrees of freedom; $p < .001$), indicating the better fit of this overall model compared with Model 0a. Model 1a revealed a significant effect of perceived competence ($\beta = .46$, $p < .001$), student sex ($\beta = -.24$, $p < .001$), and a marginally significant effect of self-determined motivation ($\beta = .09$, $p < .10$). We concluded from these analyses that these variables were diagnostic of students' performance in gymnastics, whereas participation in sports, PE and academic achievement were not.

Predictors of teacher expectations. Next, teacher expectations were regressed on students' personal characteristics and student sex. These predictors were included progressively in three two-level random intercept models: the empty Model 0b, the Model 1b in which diagnostic student personal characteristics (i.e., performance, perceived competence, self-determined motivation) and student sex were added as predictors, and the Model 2b in which non diagnostic student personal characteristics (i.e., PE and academic achievement, participation in sports) were added as predictors.

When personal characteristics diagnostic of students' performance in gymnastics were added as predictors (Model 1b), the reduction of the deviance of this model, compared with the Model 0b, was significant ($\Delta = 140.9$ with 4 degrees of freedom; $p < .001$), indicating the better fit of this overall model compared with Model 0b. Model 1b revealed a significant effect of gymnastics performance ($\beta = .27$, $p < .001$), perceived competence ($\beta = .23$, $p < .001$), and self-determined motivation ($\beta = .14$, $p < .01$) on performance expectations. In other words, teacher expectations were higher for students with high

perceived competence, performance and self-determined motivation, than for students with low perceived competence, performance and self-determined motivation.

However, independently of these characteristics, student sex also predicted teacher expectations ($\beta = .10, p < .05$). Teachers expected boys to perform higher than girls, when controlling for the previous variables. The direction of this path coefficient is opposite to the zero-order correlation between student sex and performance, and suggests that the reliance on a sex stereotype favorable to boys explained why teacher expectations underestimated the real sex differences. This, however, turned out not to be the case, as shown in the next set of analyses.

Model 2b examined whether teacher expectations were based on students' characteristics that were not diagnostic of gymnastics performance (i.e., PE and academic achievement, participation in sports), in addition to the previous variables. Results showed that the reduction of the deviance of this model, compared with Model 1b, was significant ($\Delta = 75.5$ with 3 degrees of freedom; $p < .001$), indicating the better fit of this overall model compared with Model 1b. The multilevel analysis indicated that the diagnostic student personal characteristics (included in model 1b) still predicted significantly teacher expectations – gymnastics performance ($\beta = .24, p < .001$), perceived competence ($\beta = .18, p < .001$), and self-determined motivation ($\beta = .11, p < .05$).

More important, Model 2b showed that student sex did not significantly predict teacher expectations ($\beta = -.00, p > .50$). In other words, the sex stereotype effect that appeared in Model 1b was spurious. Why did it occur? Because teachers also relied on nondiagnostic student personal information that favored boys. Specifically, Model 2b also showed that the nondiagnostic student personal characteristics also predicted teacher

expectations. PE achievement ($\beta = .33, p < .001$) and participation in sports ($\beta = .11, p < .01$) significantly predicted teacher expectations, even though these characteristics were unrelated to actual performance in Model 1a.

These analyses indicated that teacher expectations were based on students' personal characteristics: performance, perceived competence and self-determined motivation in gymnastics, past experience in PE and participation in sports. When controlling for these variables, student sex had no influence on teacher expectations. In other words, they did not rely on a sex stereotype for their expectations. This result is consistent with abundant laboratory and field research showing that perceivers evaluate targets far more on the basis of targets' personal characteristics than on targets' group membership when provided with targets' individuating information (e.g., Krueger & Rothbart, 1988; see Fiske & Neuberg, 1990; Jussim, 1991, for reviews).

However, these results show that relying on students' personal characteristics did not enhance the accuracy of teacher expectations. Instead, teachers based their expectations on nondiagnostic characteristics favorable to boys, namely past experience in PE and participation in sports. In turn, these characteristics accounted for teachers failing to see a sex difference that actually existed.

General Discussion

The purpose of the current research was to examine (1) whether sex stereotypes bias teacher expectations similarly in an experimental (with little student personal information) *versus* a naturalistic context (with ample student personal information), and (2) alternative forms of stereotype bias depending on the degree of inconsistencies between stereotypes and real sex differences. Results suggest that sex stereotype biased teacher expectations in

the experiment, but not in the naturalistic study. Nonetheless, even though there was no sex stereotype effect in the naturalistic study, teacher expectations favored boys, in the sense that teachers saw boys and girls as about equal, when, in fact, girls performed better in gymnastics.

The question of stereotype bias in teacher expectations has been already investigated, but much of the research has been experimental (e.g., Tom et al., 1984; Van Matre et al., 2000) and assumed that, if teachers perceived sex differences, they were biased by invalid stereotypes. However, in a naturalistic context, perceiving sex differences is not sufficient to conclude that stereotype effects occurred. The present research used a more appropriate index of stereotype bias by comparing the sex differences expected by teachers to the actual sex differences in a natural PE setting (e.g., Jussim et al., 1996).

Before doing so, it was necessary to understand whether sports stereotypes favorable to boys may be influential in the context of gymnastics. Study 1 used a classic experimental paradigm to examine this question and the results showed that teacher expectations of performance were higher for boys than for girls. This finding suggested that the stereotype about male superiority in sports may influence teacher expectations (e.g., Koivula, 1999; Messner, 1988, 1990) even in the context of gymnastics, a sport sex-typed as feminine.

The second step of the research was to investigate whether and how stereotypes affect teacher expectations when real sex differences may exist. More particularly, we examined whether stereotypes would lead teachers to underestimate sex differences in gymnastics performance. The findings showed that teachers failed to detect the sex differences in favor of girls that actually existed. Although this result is a bias that favors

boys, there was no sex stereotype effect. Instead, teacher expectations were biased because they were based on inaccurate student personal characteristics (i.e., PE achievement and participation in sports). In the context of the classroom, one could argue that sex-undifferentiated expectations are not as problematic as sex-differentiated teacher expectations, because they are less likely to result in sex differences in achievement. However, if Group A members perform lower than Group B members in a particular domain, and teachers expect Group B members to perform as low as Group A members in this domain, these group-undifferentiated expectations could prevent Group B members from performing as high as they possibly could.

Limitations

Accuracy criterion. In order to assess the accuracy of the expected sex differences, we used the assessments of trained gymnastics experts regarding students' behavioral performance. The accuracy of these assessments was strengthened by asking the experts to evaluate the students as accurately as possible, referring the ratings to standardized criteria relative to the execution of the task, and by asking the experts to observe each performance three times.

However, there is almost never a perfect criterion for assessing either the accuracy of laypeople's beliefs. It could be argued that there is a particular type of imperfection in the expert assessments of students' performance (i.e., the measure of the actual performance) - perhaps they were biased by sex stereotypes. If this were true, our research might underestimate the effects of sex stereotypes on teachers' expectations (if the experts held and applied the stereotype that boys were better at gymnastics than girls), or it might overestimate the effects of sex stereotypes on teachers' expectations (if the experts held and

applied the stereotype that girls were better at gymnastics than boys). Next, therefore, we consider the plausibility of bias in expert judgments.

Experts are, in general, less biased than naive observers (Jussim, 2005; Ryan 2002). More important, we instructed our experts to use standardized behavioral standards when evaluating gymnastics performances. Use of this type of clearly standardized metric typically reduces bias in judgments and evaluations (Biernat & Manis, 1994). Furthermore, experts were instructed to be as accurate as possible in their ratings, instructions which are well-established at eliminating stereotype bias (Judice & Neuberg, 1998; Snyder & Stukas, 1999). In addition, experts devoted their full attention to their judgments, and attention to personal characteristics is a well-established condition under which stereotype effects are routinely eliminated (Fiske & Neuberg, 1990).

Last, one must consider what the experts were actually doing. They were not providing vague or general evaluations of overall gymnastics ability or performance. They were evaluating relatively objective characteristics of specific gymnastics performances, such as the straightness of the body. Whereas it is well-established that stereotypes sometimes influence perceptions of vague or general attributes that are open to interpretation (e.g., personality characteristics, assertiveness, skill at math, etc.), it is equally well-established that they typically have no effect whatsoever on objective behaviors to which perceivers pay close attention (e.g., Fiske & Neuberg, 1990; Jussim, 1991; Kunda & Thagard, 1996). We conclude, therefore, that, given the way in which we assessed expert judgments, the idea that our experts' judgments were biased very much by sex stereotypes is highly implausible.

A related issue is that the evaluations made by the teachers (i.e., expectation relative to a future performance) differed from the ones made by the experts (i.e., perception of a current performance). Perhaps it would have been preferable that the two judgments were in the same temporal perspective. However, we believe that this discrepancy did not affect the results because previous research showed that predictions and evaluations are virtually identical and highly correlated (Kahneman & Tversky, 1973; Smith et al., 1998).

Correlation and causality. Study 1 was an experiment, which allows one to draw causal conclusions about the effect of target sex on teachers' expectations of performance, at least in the artificial context of the experiment. Study 2, however, was a naturalistic study, which raises issues regarding the direction of causal effects. Could teacher expectations have caused the actual differences between boys and girls? This seems unlikely. Teachers expected boys and girls to be equal, yet they were not equal. If these beliefs were self-fulfilling, boys and girls should have performed equivalently, or nearly so. This did not happen. Nonetheless, perhaps teacher expectations created small self-fulfilling prophecies, which reduced but did not eliminate girls' superiority at gymnastics. The design of the two studies was transversal so we cannot address this issue, but this is a possibility that deserves further investigation. A longitudinal study would be necessary to examine to what extent teacher expectations may create self-fulfilling prophecies. Also, the particular variances and covariances among the data also always affect the size of regression coefficients. Whether our particular coefficients generalize to new data is a question that will have to be addressed by future research.

There is also the omitted variable problem. Even if we can rule out reverse causality, there is always the possibility that variables not assessed accounted for some or

all of the relations between the variables that were assessed. It is important to note that no matter how many control variables are assessed in a particular study, it is impossible to know that one assessed all relevant variables.

What is possible, however, is to assess plausible and well-known and well-established predictors of the variables in one's model. Study 2 did just this. A slew of prior achievement and motivational variables were assessed, because such variables are well-known to be influences on both performance and teacher expectations were plausibly related to student sex. Indeed, a strength of Study 2 was that it identified variables leading to a spurious relation between student sex and teacher expectations.

Comparison with Prior Research

These limitations notwithstanding, our results are similar to a vast accumulated literature on the role of sex stereotypes in particular and stereotypes more generally. They typically have modest effects on person perception when little individuating information is available (as we found in Study 1), and (as we found in Study 2) typically have little or no effect on judgments in the presence of clear individuating information (e.g., Jussim et al., 1996; Locksley, Borgida, Brekke, & Hepburn, 1980; Swim, Borgida, Maruyama, & Myers, 1989). Thus, although it is possible that different results would be found in different social contexts from the one we studied, this, too, seems unlikely.

Theoretical and Applied Contributions of the Present Research

These results are important for several theoretical and applied reasons. First, whereas a major concern about the effect of sex on teacher expectations has been that it unjustifiably leads people to see differences that do not exist (e.g., Fiske & Neuberg, 1990), Study 2, however, showed a different type of bias: People may sometimes, and perhaps

often, fail to detect differences that really do exist. Thus, even egalitarian beliefs may be biased in favor of a social group if differences do really exist. In the particular case of Study 2, teachers were completely egalitarian in the sense that they saw boys and girls as performing equally well in gymnastics. They were, however, wrong – girls had performed better. This egalitarian belief, therefore, was actually biased against girls.

Second, relying on individuating information is usually considered as some sort of gold standard of unbiased judgment. Study 2, however, shows that is not necessarily true and that it depends on the basis for individuation. In the particular case of Study 2, teachers individuated on the basis of characteristics that were unrelated to actual gymnastics performance and which favored boys. This produced a bias in favor of boys that was just as unjustified as any bias produced by stereotypes.

Conclusion

Our research showed that teachers' expectations of students' gymnastics performance were biased in both studies in favor of boys, such that boys were erroneously seen as performing better than girls in Study 1, and as erroneously seen as performing as good as girls in Study 2. This pattern was quite robust, occurring in both the experiment and the naturalistic study.

This research, however, also demonstrated limitations to the power of sex stereotypes in several ways. First, sex stereotypes did not influence expectations of performance in Study 2, when teachers had ample individuating information about students. Second, the effects of individuating information in Study 2 were much larger than the effects of stereotypes. In other words, modest bias existed right alongside teachers being nicely in touch with the real ways in which boys and girls were similar and different.

Nonetheless, any bias, especially in a potentially socially charged domain such as sex stereotypes and discrimination, is too much. Western societies have worked quite hard over the last several decades to eliminate barriers to girls' athletic opportunities and achievements. Despite considerable success in these efforts (there are far more athletic opportunities for girls now than in 1960), our research has also revealed a small, but continuing, obstacle: teachers' small but systematic tendency to see girls' athletic performance as less good (compared to boys) than it really is. Even if teacher expectations have – on average – only modest impact on student achievement, from .10 to .30 in terms of standardized regression coefficients (e.g., Brophy, 1983; Jussim, Smith, Madon, & Palumbo, 1998; Trouilloud et al., 2002), students belonging to stigmatized social groups (e.g., girls in masculine domains, African-Americans, students from low social class) are more vulnerable to self-fulfilling prophecies than students from non stigmatized social groups (see for a review Jussim & Harber, 2005). Then, future research should examine whether these sex-biased teacher expectations prevent girls from performing less well than they could, and whether the results of this research could be replicated in other school subjects.

References

- Beyer, S. (1999). The accuracy of academic gender stereotypes. *Sex Roles, 40*, 787-813.
- Biernat, M., & Manis, M. (1994). Shifting standards and stereotype-based judgments. *Journal of Personality and Social Psychology, 66*, 5-20.
- Biernat, M., Manis, M., & Nelson, T. E. (1991). Stereotypes and standards of judgements. *Journal of Personality and Social Psychology, 60*, 485-499.
- Boiché, J., Sarrazin, P., Grouzet, F.M.E., Pelletier, L.G., & Chanal, J. (*in press*). Students' Motivational Profiles and Achievement Outcomes in Physical Education: A Self-Determination Perspective. *Journal of Educational Psychology*.
- Bowker, A., Gadbois, S., & Cornock, B. (2003). Sports participation and self-esteem: Variations as a function of gender and gender role orientation. *Sex Roles, 49*, 47-58.
- Brière, N. M., Vallerand, R. J., Blais, M. R. et Pelletier, L. G. (1995). Développement et validation d'une mesure de validation intrinsèque, extrinsèque et d'amotivation en contexte sportif : l'Echelle de motivation dans les sports (EMS). *International Journal of Sport Psychology, 26*, 465-489.
- Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Sage Publications. Newburg Park London.
- Brophy, J. (1983). Research on the self-fulfilling prophecy and teacher expectations. *Journal of Educational Psychology, 75*, 631-661.
- Chanal J., Marsh H., Sarrazin P., & Bois J (2005) Big-Fish-Little-Pond Effects on Gymnastics Self-Concept: Social Comparison Processes in a Physical setting. *Journal of Sport and Exercise Psychology, 27*, 53-70

- Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioural sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Darley, J. M. & Gross, P. H. (1983). A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology*, 44, 20-33.
- Deaux, K., & Major, B. (1987). Putting gender into context: An interactive model of gender-related behavior. *Psychological Review*, 94, 369-389.
- Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-regulation in human behavior*. New York : Plenum.
- Duncan, B. L. (1976). Differential social perception and attribution of intergroup violence : Testing the lower limits of stereotyping of Blacks. *Journal of Personality and Social Psychology*, 34, 590-598.
- Dusek, J. B., & Joseph, G. (1983). The bases of teacher expectancies: A meta-analysis. *Journal of Educational Psychology*, 75, 327-346.
- Eccles, J. S., & Harold, R. D. (1991). Gender differences in sport involvement: Applying the Eccles' expectancy-value model. *Journal of Applied Sport Psychology*, 3, 7-35.
- Fiske, S. T., Lin, M., & Neuberg, S. L. (1999). The continuum model. Ten years later. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 231-254). New York London : The Guilford Press.
- Fiske, S. T. & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 1-74). New York : Academic Press.

- Fredricks, J. A., & Eccles, J. S. (2005). Family socialization, gender, and sport motivation and involvement. *Journal of Sport and Exercise Psychology, 27*, 3-31.
- French Ministry of Education (2000). Les épreuves du baccalauréat en éducation physique et sportive. Rapport de l'Inspection Générale. [High school diploma in physical education. Report from the General Inspection. Retrieved November, 25th, 2006 from <ftp://trf.education.gouv.fr/pub/edutel/syst/igen/baceps.pdf>.
- Funder, D. C. (1987). Errors and mistakes: Evaluating the accuracy of social judgment. *Psychological Bulletin, 101*, 75-90.
- 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.
- Judd, C. M., & McClelland, G. H. (1989). *Data analysis: A model comparison approach*. New York: Harcourt Brace Jovanovich.
- Judd, C. M., & Park, B. (1993). Definition and assessment of accuracy in social stereotypes. *Psychological Review, 100*, 109-128.
- Judice, T. N., & Neuberg, S. L. (1998). When interviewers desire to confirm negative expectations : Self-fulfilling prophecies and inflated applicant self-perceptions. *Basic and Applied Social Psychology, 20*, 175-190.
- Jussim, L. (1989). Teacher expectations: Self-fulfilling prophecies, perceptual biases, and accuracy. *Journal of Personality and Social Psychology, 57*, 469-480.
- Jussim, L. (1991). Social perception and social reality: A reflection-construction model. *Psychological Review, 98*, 54-73.

Jussim, L. (2005) Accuracy: Criticisms, controversies, criteria, components, and cognitive processes. *Advances in Experimental Social Psychology*, 37, 1-93

Jussim, L. & Eccles, J. (1992). Teacher expectations 2: Construction and Reflection of student Achievement. *Journal of Personality and Social Psychology*, 63, 947-961.

Jussim, L. & Eccles, J. (1995). Naturally occurring interpersonal expectancies. *Review of Personality and Social Psychology*, 15, 74-108.

Jussim, L., Eccles, J. & Madon, S. (1996). Social perception, social stereotypes, and teacher expectations: Accuracy and the quest for the powerful self-fulfilling prophecy. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol.28, pp. 281-388). San Diego, CA : Academic Press.

Jussim, L., & Harber, K. D. (2005). Teacher expectations and self-fulfilling prophecies: Knowns and unknowns, resolved and unresolved controversies. *Personality and Social Psychology Review*, 9, 131-155.

Jussim, L., Smith, A., Madon, S. & Palumbo, P. (1998). Teacher expectations. *Advances in Research on Teaching*, 7, 1-48.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237-251.

Koivula, N. (1999). Sport participation: Differences in motivation and actual participation due to gender typing. *Journal of Sport Behavior*, 22, 360-380.

Krueger, J., & Rothbart, M. (1988). Use of categorical and individuating information in making inferences about personality. *Journal of Personality and Social Psychology*, 55, 187-195.

- Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits and behaviors: A parallel-constraint satisfaction theory. *Psychological Review*, *103*, 284-308.
- Li, F., & Harmer, P. (1996). Testing the simplex assumption underlying the Sport Motivation Scale: a structural equation modeling analysis. *Research Quarterly for Exercise and Sport*, *67*, 396-405.
- Locksley, A., Borgida, E., Brekke, N., & Hepburn, C. (1980). Sex stereotypes and social judgment. *Journal of Personality and Social Psychology*, *39*, 821-831.
- Madon, S., Guyll, M., Hilbert, S. J., Kyriakatos, E., & Vogel, D.L. (2006). Stereotyping the stereotypic: When individuals match social stereotypes. *Journal of Applied Social Psychology*, *36*, 178-205.
- Madon, S. J., Jussim, L., Keiper, S., Eccles, J., Smith, A., & Palumbo, P. (1998). The accuracy and power of sex, social class and ethnic stereotypes: Naturalistic studies in person perception. *Personality and Social Psychology Bulletin*, *24*, 1304-1318.
- Marsh, H., Chanal, J., & Sarrazin, P. (2006). Self-belief does make a difference: A reciprocal effects model of the causal ordering of physical self-concept and gymnastics performance. *Journal of Sports Sciences*, *24*, 101-111.
- Merton, R. (1948). The self-fulfilling prophecy. *Antioch Review*, *8*, 193-210.
- Messner, M. A. (1988). Sports and male domination: The female athlete as contested ideological terrain. *Sociology of Sport Journal*, *5*, 197-211.
- Messner, M. A. (1990). Men studying masculinity: Some epistemological issues in sport sociology. *Sociology and Sport Journal*, *7*, 136-153.

MJS Stat-Info (2004). Les femmes dans la pratique sportive licenciée en 2002. *MJS Stat-Info no 04-07*. November 2004. [Women in institutionalized sports participation].

Retrieved January 16, 2007, from

<http://www.jeunessesports.gouv.fr/IMG/pdf/statinfo04-07.pdf>

Nelson, T. (2002). *The psychology of prejudice*. Boston: Allyn & Bacon.

Rasbash, J., Browne, W., Healy, M., Cameron, B., & Charlton, C. (2001). MLwiN Software 1.1.

Riemer, B. A., & Visio, M. E. (2003). Gender typing of sports: An investigation of Metheny's classification. *Research Quarterly for Exercise and Sport*, 74, 193-204.

Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom: Teacher expectation and student intellectual development*. New York : Holt, Rinehart et Winston.

Ryan, C. (2002). Stereotype accuracy. *European Review of Social Psychology*, 13, 75-109.

Ryan, R. M., & Connell, J. P. (1989). Perceived locus of causality and internalization: Examining reasons for acting in two domains. *Journal of Personality and Social Psychology*, 57, 749-761.

Skrypnek, B., & Snyder, M. (1982). On the self-perpetuating nature of stereotypes about women and men. *Journal of Experimental Social Psychology*, 18, 277-291.

Smith, A. E., Jussim, L., Eccles, J., VanNoy, M., Madon, S., & Palumbo, P. (1998). Self-fulfilling prophecies, perceptual biases, and accuracy at the individual and group levels. *Journal of Experimental Social Psychology*, 34, 530-561.

Snyder, M., & Stukas, A. (1999). Interpersonal processes: The interplay of cognitive, motivational, and behavioural activities in social interaction. *Annual Review of Psychology*, 50, 273-303.

- Swim, J. K. (1994). Perceived versus meta-analytic effect sizes: An assessment of the accuracy of gender stereotypes. *Journal of Personality and Social Psychology*, *66*, 21-36.
- Swim, J., Borgida, E., Maruyama, G., & Myers, D. G. (1989). Joan McKay vs. John McKay: Do gender stereotypes bias evaluations? *Psychological Bulletin*, *105*, 409-429.
- Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin, & S. Worchel (Eds.), *The social psychology of intergroup relations*. Belmont, CA: Wadsworth.
- Tom, D. Y., Cooper, H., & McGraw, M. (1984). Influence of student background and teacher authoritarianism on teacher expectations. *Journal of Educational Psychology*, *76*, 259-265.
- Thomas, J. R., & French, K. E. (1985). Gender differences across age in motor performance: A meta-analysis. *Psychological Bulletin*, *98*, 260-282.
- Trouilloud, D., Sarrazin, P., Martinek, T. J., & Guillet, E. (2002). The influence of teacher expectations on student achievement in physical education classes: Pygmalion revisited. *European Journal of Social Psychology*, *32*, 591-607.
- Vallerand, R. J. (1997). Toward a Hierarchical model of intrinsic and extrinsic motivation. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 29, pp.271-359). New York : Academic Press.
- Vallerand, R. J., & Fortier, M. N. (1998). Measures of intrinsic and extrinsic motivation in sport and physical activity : A review and critique. In J. L. Duda (Ed.), *Advances in*

Sport and Exercise Psychology Measurement (pp. 81-101). Morgantown: Fitness Information Technology.

Vallerand, R.J., & Losier, G.F. (1999). An integrative analysis of intrinsic and extrinsic motivation in sport. *Journal of Applied Sport Psychology*, 11, 142-169.

Van Matre, J. C., Valentine, J. C., & Cooper, H. (2000). Effect of students' after-school activities on teachers' academic expectancies. *Contemporary Educational Psychology*, 25, 167-183.

Wang, C. K., Papaioannou, A. G., Sarrazin, P. G., Jaakola, T., & Solmon, M. A. (2006). A brief description of physical education and school children's sport involvement in Singapore, Greece, France, Finland, and the United States. *International Journal of Sport and Exercise Psychology*, 4, 220-226.

Wigfield, A. & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, 25, 68-81.

Wolsko, C., Park, B., & Judd, C. (2000). Framing interethnic ideology: Effects of multicultural and color-blind perspectives on judgments of groups and individuals. *Journal of Personality and Social Psychology*, 78, 635-654.

Author Note

Acknowledgements: We would like to thank Julien Chanal and Damien Tessier for their help in collecting the data, and Brenda Major for her helpful comments on previous drafts of this manuscript.

Manuscript in press in JASP

Notes

1. Research on shifting standards shows that, when using *subjective* criteria (e.g., “ratings of how tall or how good someone is on scales with undefined endpoints such as “very tall” or “very good”), people will appear to be unbiased, but this is only because perceivers do not use the same standard when judging a man or a woman (e.g, tall for a woman does not mean the same thing as tall for a man). However, when person perception is assessed using *objective* measures (such as height in inches), bias often emerges even when it did not emerge using subjective measures. The present research, requiring the expert judges to evaluate the students using standardized criteria, maximizes therefore our ability to discover whether or not the expert judges perceived a difference between particular male and female students.

Table 1.

Descriptive Statistics and Intercorrelations Among Variables (N = 422)

Variable	Mean	SD	1	2	3	4	5	6	7	8
1. Teacher performance expectation	4.30	1.35	-	.41***	.47***	.35***	.47***	.28***	.12*	.07
2. Gymnastics performance	3.49	1.33		-	.49***	.37***	.14**	.08	.09 [†]	-.18***
3. Perceived competence	4.35	1.39			-	.52***	.27***	.23***	-.11*	.12*
4. Self-determined motivation	3.88	7.38				-	.17***	.11*	.06	-.11*
5. PE achievement	13.15	2.62					-	.30***	.21***	.26***
6. Participation in sports ^b	2.13	1.38						-	.05	.26***
7. Academic achievement	12.54	2.34							-	-.29***
8. Student sex ^c	1.55	0.50								-

Note: All correlations were rendered independent of students' classrooms with multilevel modeling.

^bNo participation in sports was coded as 0, from 1 to 2 hours per week spent on doing sports was coded as 1, from 2 to 4 hours was coded as 2, from 4 to 6 hours was coded as 3, from 6 to 8 hours was coded as 4, and more than 8 hours per week was coded as 5. ^cGirls were coded as 1 and boys as 2.

[†] $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$.

Table 2.

Hierarchical Linear Models With Students' Performance As Dependent Variable

Variables	Model 0a		Model 1a	
	β or σ^2		β or σ^2	t
Fixed Effects				
Constant	.00	.05	.01	0.07
Perceived competence			.46	9.26***
Self-determined motivation			.09	1.80 [†]
PE achievement			.05	.94
Participation in sports			.02	.48
Academic achievement			.07	1.52
Sex			-.24	5.11***
Random Effects				
Class level	.07	2*	.164	2.88**
Individual level	.92	4.22***	.60	14.05***
-2 log L	1185.03		1026.71	

[†] $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$.

Table 3.

Hierarchical Linear Models With Performance Expectations As Dependent Variable

Variables	Model 0b		Model 1b		Model 2b	
	β or σ^2	<i>t</i>	β or σ^2	<i>t</i>	β or σ^2	<i>t</i>
Fixed Effects						
Constant	.02	0.28	.03	0.31	.02	0.25
Performance			.27	5.67***	.24	5.20***
Perceived competence			.23	4.40***	.18	3.54***
Self-determined motivation			.14	2.80**	.11	2.40*
PE achievement					.33	7.22***
Participation in sports					.11	2.78**
Academic achievement					.03	0.81
Sex			.10	2.52*	-0.00	0.02
Random effects						
Class level	.13	2.46*	.16	2.82**	.14	2.88**
Individual level	.87	14.20***	.61	14.07***	.51	14.03***
-2 log L	1167.72		1026.84		951.36	

* $p < .05$; ** $p < .01$; *** $p < .001$.

β refers to the standardized regression coefficients of the fixed parameters; σ^2 refers to the variance of the random parameters.

Figure Captions

Figure 1. An adapted version of the reflection – construction model (e.g., Jussim et al., 1996)

Manuscript in press in JASP

