The Influence of Accountability on Teachers' Decision Making in a Simulated Classroom

Julia E. Herfordt-Stöpel, Thomas Hörstermann
University of Luxembourg

Abstract

The quality of teacher judgments about their students' achievements as a focus of research is connected with the question of how the quality of these judgments can be improved. A possibility of improving the quality is to make teachers aware of the great impact their judgments have. This means increasing accountability. To test the effects of accountability on teacher judgments, an experiment with the simulated classroom was conducted. In the simulated classroom, virtual students with different mathematical ability were presented on a computer screen. Teachers assessed these students' ability after two simulated lessons. With this approach, systematic biases could be shown. Girls were assessed as less able in solving mathematical problems than boys with the same ability. However, for the increased accountability condition, only marginal significant effects resulted. When accountability was high, no differences in ratings for boys and girls with the same ability were observed.

1. Introduction

One of the main assessment tasks of teachers is to grade their students. This is a very responsible task, because the grades which originate from these assessments determine professional opportunities of the student in the future [1]. However, former research showed that accuracy is not as high as it should be for such important judgments [2].

One problem which impairs the quality of such judgments is that certain student attributes are used to assess students' ability which are not related to performance in the domain of interest. One attribute which might be used in this way and which might lead to systematic biases is a student's gender [3] [4] [5]. Another student attribute which might have a similar effect as the student's gender is the migration background of a student. The influence of factors like gender or migration background on achievement judgments is especially severe in school systems with a strict selection procedure for different school tracks as it is the case in Luxembourg and Germany. Effects for the students' attributes socioeconomic and migration background could be shown for students in Germany and Luxembourg [6] [2]. The use of such information might influence the judgment in a systematic way. For example, when an attribute of the student is perceived to be associated with certain ability, like male gender and performance in mathematics, a girl should be assessed as less good in mathematics compared to a boy with the same ability, when the information of the gender category (girl) is activated [7].

The activation of a category depends on certain factors. First, the person who is assessed has to possess certain attributes which are in line with the category, so that the activation of the category is very likely. Second, the judging person has to judge the student in an effortless and efficient way [8] [9].

The continuum-model of judgment formation distinguishes two processes of judging, which can be seen as end points of a continuum [10]. One end of the continuum refers to the category-based strategy of judgment formation, which constitutes a rather effortless and efficient way of judging. By using this strategy of judgment, an activated category may bias the judgment, because judges may neglect individual person information which is not in accordance with the activated category. Thus, the activated category influences the perception, interpretation, and storage of new information [11]. This strategy of judgment is mainly used when the judging person is not very motivated to judge, does not feel very accountable for the judgment, or does not have the ability to make an elaborated judgment. Despite the disadvantages mentioned above, this strategy is used frequently because of the efficiency and the low cognitive effort needed [12].

The attribute-based strategy of judging is situated on the other end of the continuum. This strategy is used when the judging person is highly motivated to make an elaborated judgment, when the person feels accountable for the judgment, and when the judging person has the ability to make an elaborated judgment. If using this strategy, a judge integrates every attribute of the student into a judgment. Therefore, information processing and storage of person information involve data that are close to the student's individual attributes. A disadvantage of this strategy is that it is very effortful for the judging person [13] [14].

In the school context, we can assume that teachers have the opportunity the make elaborated judgments due to their professional practice [15]. However, the question is whether the situational circumstances allow for high quality judgments and if the feeling of accountability for the judgment plays a role [16].

2. Method

To test if accountability has an influence on the quality of teacher judgments, an experiment was conducted using a simulated classroom approach [17] [18]. A simulated classroom is a computer program. In this computer program, virtual students are presented by their names and pictures to simulate a class.

In the simulated classroom in this experiment, 12 students were presented. Half of the students were male and the other half were female. Furthermore, half of the male and half of the female students in the virtual class were of Luxembourgish origin, while the other half were of Portuguese origin and therefore members of the biggest immigrant group in Luxembourg.

To allow for the activation of a category, the names and pictures used in the study were pretested concerning their typicality. Only pictures and names were chosen which were clearly identifiable as male or female and Luxembourgish or Portuguese. Furthermore, the chosen pictures did not differ substantially on any other dimension (e.g. age, attractiveness).

Mathematics was chosen as the subject to be taught in the virtual classroom setting. This choice has the advantage that clear assumptions are available for the subjectively assumed ability of male and female participants. The stereotype for achievement in mathematic is still that boys should have a higher performance in this subject than girls [19].

For each of the virtual students, a certain parameter of ability in mathematics and motivation can be programmed. This has the advantage that, compared with to real class settings, the true value for ability in mathematics and motivation is known. For the simulated classroom in this experiment, we programmed a good, an average and a poor student for each group of virtual students (male – Luxembourgian, male – Portuguese, female – Luxembourgian, and female – Portuguese). A good student solved 100% of the easy tasks, 80% of the medium tasks and 60% of the difficult tasks. An average student solved 80% of the easy tasks, 50% of the medium and 20% of the difficult tasks. Finally, the poor student performed well on 50% of the easy tasks, 20% of the medium tasks and 0% of the hard tasks.

With the simulated classroom, it is possible to record the following dependent variables: a) the order in which tasks were posed to each student, b) the total number of tasks posed to each student and c) the difficulty of tasks posed to each student.

After working on the simulated classroom, participants were asked to rate the mathematical ability and the motivation for each student.

3. Design and Procedure

The accountability for the decision task was varied between participants. Half of the participants were only informed that they should conduct a mathematical lesson in a virtual classroom. The other half of the participant was additionally reminded that grading of students is a very important task and that grading might have severe consequences for the professional development and the future of each student.

The induction of the categorical information about the students was manipulated in a within-group design, as was gender (see also the description in the method section).

Each participant was tested in an individual session. Therefore, a member of the research team visited the participant at her or his school. The experiment was conducted on a standard 15’’-screen laptop in a separate and quiet room at the participant’s school.

The experiment started with general information about the experiment and a short practice session to familiarize the participants with the use of the simulated classroom. After the participants were accustomed to the simulated classroom, the experimental classroom session was started. At the beginning of the classroom, an instruction was presented containing the manipulation of accountability (high vs. low). After instruction, the first lesson began. During the lesson, the participant had the opportunity to pose 18 tasks to the students in the simulated classroom. After the 18 tasks, the participants rated all of the students concerning their ability to solve the mathematical problems in the simulated classroom and their motivation concerning this subject. After the first lesson, a second lesson with 18 tasks and the same ratings at the end followed. At the beginning of the second lesson, the accountability instruction was presented again.

The course of each trial during the simulated classroom was as follows: First the participant had to choose whether he/she preferred to pose an easy, medium or difficult task to the class. Dependent on the participant’s choice, six easy, medium or difficult tasks appeared on the right side of the screen. In the next step, the participant chose which specific task he/she wanted to pose to the class. After confirmation by mouse click, this specific task was

posed to the class. After posing the task, markers (sun symbols) appeared next to some of the student's pictures, signalizing that these students volunteered to give a response to this mathematical problem.

The participant then chose a student to present his/her solution of the mathematical problem. The participant was free to choose one of the volunteering or non-volunteering students. Afterwards, the response of this student appeared on the right side of the screen.

In line with the ability of the virtual student, the presented answers could be right or wrong. To make it easier for the participant, right answer were framed by a green frame, while wrong answers where framed by a red frame. An example of the virtual classroom used in the experiment can be seen in Figure 1.

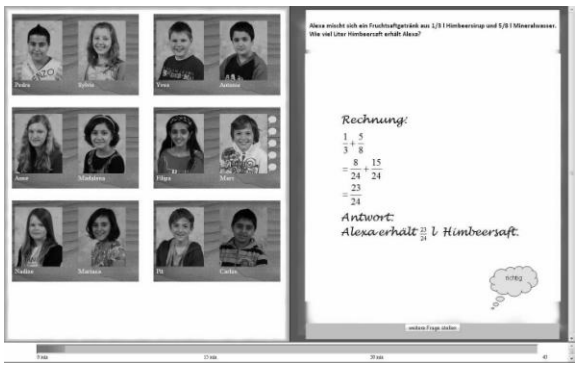

Figure 1. Simulated Classroom with correct response of a student

4. Hypotheses

For ability in mathematics we expected that boys showing the same performance in mathematics as girls would be assessed as more competent when accountability was low. If the accountability was high, no differences in the assessment between boys and girls were expected.

For the difficulty of tasks posed to the student, we expected that more difficult tasks would be posed to boys more frequently than to girls in the low accountability condition. In contrast, when the accountability was high, no differences in the difficulty of the tasks posed to boys and girls were expected.

For the motivation and the number of tasks asked to each student group, no clear hypotheses were formulated. These factors were only taken into account as additional factors.

Additionally, the migration background of the students was taken into account to control for level differences in the dependent variables (rating of students' mathematical ability, rating of students' mathematical motivation, number of tasks posed to each student, difficulty of the tasks posed to the students, and the difficulty of the first task) between boys and girls with or without a migration background. Although stereotypes of students with Portuguese origin do not imply clear advantages or disadvantages in mathematical ability, the migration background might influence teachers' ratings via indirect effects of the student stereotype.

5. Participants

31 Luxembourgish elementary school teachers took part in this study. The mean value of teaching experience was 14 years. Their mean age was 40 years.

6. Results

For the dependent variables (*mathematical ability*, *motivation*, *number of tasks posed*, and *difficulty of tasks posed*), the mean for the first and second lesson was calculated so that information from the first and second lesson could be analyzed. Then the values of all girls were averaged, as were those of all boys, so that the values for boys and girls could be tested against each other. For the mathematical ability ratings, the discrepancy to the programmed ability was calculated and used as dependent variable.

In an analysis of variance with the between-subject factor accountability, the repeated measurement factor gender (male vs. female), and the discrepancy between the programmed and the rated mathematical ability as dependent variable, a significant main effect of the factor gender was shown, $F(1,29) = 3.88$, $p = .001$, $\eta^2 = .32$. This main effect implies that boys are rated better than girls when they have the same ability in mathematics.

For the interaction between the accountability and the gender of the students, only a marginally significant effect could be shown, $F(1,29) = 3.05$, $p = .09$, $\eta^2 = .09$. On a descriptive level, the mathematical ability of boys was overestimated when accountability was low, while the mathematical ability of girls was underestimated. In contrast, when accountability was high, the mathematical ability of both boys and girls was overestimated, but the ability of boys was overestimated more than that of girls (Figure 2). No other effects could be shown, all further F's < 1.71, n.s.

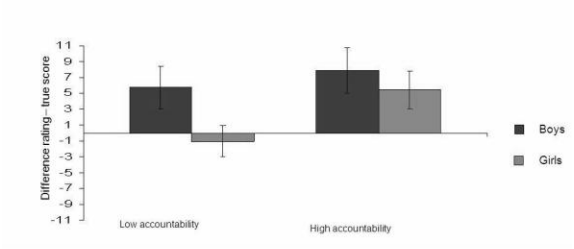

Figure 2. Participants' ratings of the mathematical ability of boys and girls in dependence of the accountability condition

Additional t-tests investigated whether boys and girls were assessed differently according to their migration status. In both accountability conditions, the mean rated mathematical ability did not differ significantly between migrants and non-migrants, all t's < ¦1.51¦, n.s. (see also Table 1 and Table 2).

Analogous to the analysis of rated mathematical ability an analysis of variance was conducted for the motivation ratings with the between-subject factor accountability, the repeated measurement factor gender. In this analysis, only a main effect for the gender variable appeared, $F(1,29) = 4.95$, $p = .03$, $\eta^2 = .15$. The motivation of boys for the subject mathematics was rated as significantly higher than for girls No further effects could be shown in this analysis, all further F's < 1.24, n.s. (Figure 3).

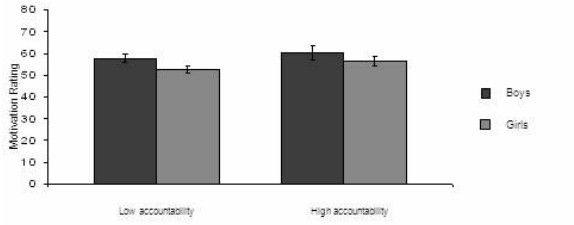

Figure 3. Participants' ratings of the motivation of boys and girls by accountability condition

Again, it was tested separately for each gender group, whether the migration background of the students makes a difference in the motivation ratings compared to students with no migration background. Again the t-tests were run separately for each accountability instruction (high vs. low accountability). For the low accountability condition a significant difference between girls with a migration background and with no migration background could be shown, $t(15) = 2.40$, $p = .03$. The motivation for girls with a migration background was rated higher than for girls with no migration background. The motivation ratings for boys in the low accountability condition and for boys and girls in the high accountability condition did not differ significantly, all t's < ¦1.84¦, n.s. (see Table 1 and Table 2)

For the dependent variable *frequency of tasks posed*, an analysis of variance with the between subject factor accountability and the repeated measurement factor gender showed no significant effects, all F's < 1. This means that girls and boys received the same number of tasks in total and that the accountability manipulation had no effect on the number of tasks posed.

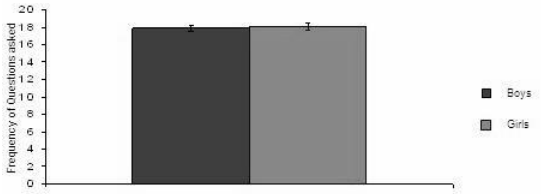

Figure 4. Frequency of tasks posed to boys and girls in the simulated classroom

Separate t-tests for each accountability condition revealed a significant difference between boys with a migration background and boys without migration background in the low accountability instruction, $t(15) = 2.25$, $p = .04$. For girls in the low accountability condition and for boys and girls in the high accountability condition, no significant differences in dependence of the migration background were indicated, all t's < ¦.97¦, n.s. (see also Table 1 and Table 2).

An additional analysis of variance was performed for the dependent variable *difficulty of tasks posed* with the between-subject factor accountability and the repeated measurement factor gender. For this analysis, only the main effect of accountability reached significance, $F(1,29) = 7.86$, $p = .01$, $\eta^2 = .21$. In the high accountability condition, the mean difficulty of tasks was higher than in the low accountability condition. No other effects reached significance, all further F's < 2.13.

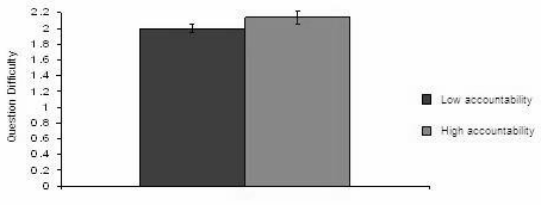

Figure 5. General difficulty of questions asked to boys and girls in the simulated classroom

Also for the dependent variable general difficulty it was tested if there were significant differences for

boys and girls with or without a migration background in dependence from the induced goal. Again, separate t-tests for each subgroup were conducted. Testing for differences in the mean difficulty of tasks posed between migrants and non-migrants revealed no significant differences, all t's < ¦1.87¦, n.s. (see also table 1 and table 2). Therefore, the difficulty of questions posed to each student in the simulated classroom was equal throughout all subgroups and conditions.

A final analysis was conducted for the difficulty of the first task posed to each student with the between subject factor accountability and the repeated measurement factor gender. The analysis revealed a significant main effect of gender, $F(1,29) = 4.21$, $p = .05$, $\eta^2 = .12$. The first task posed to girls was more difficult than for boys. No further effects reached significance, all further F's < 2.37.

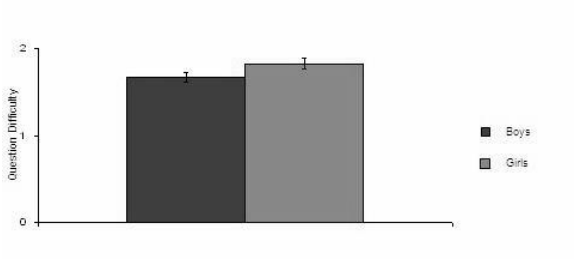

Figure 6. Difficulty of the first task posed to a student in the simulated classroom

For the first task, no difference could be shown between students with a migration background and with no migration background, in the high as well as in the low accountability condition, all t's < ¦1.38¦, n.s., (see also Table 1 und Table 2). Therefore, the difficulty of the first question is the same for children with a migration background and without a migration background.

Table 1. Means and Standard Deviation for Girls with and without Migration Background

	Low Accountability		High Accountability	
	Without Migration Background	With Migration Background	Without Migration Background	With Migration Background
	M (SD)	*M (SD)*	*M (SD)*	*M (SD)*
Mathematical Ability	-2.35 (10.63)	0.26 (7.69)	5.42 (10.91)	5.45 (10.10)
Motivation	48.76 (6.76)	56.66 (10.61)	55.62 (10.86)	57.57 (9.96)
Number of tasks	8.93 (1.53)	8.88 (2.06)	9.13 (1.77)	9.27 (1.58)
Difficulty of tasks	1.01 (0.32)	0.99 (0.27)	1.24 (0.29)	1.06 (0.31)
Difficulty first task	1.90 (0.42)	1.88 (0.42)	1.82 (0.45)	1.73 (0.49)

Table 2. Means and Standard Deviation for Boys with and without Migration Background

	Low Accountability		High Accountability	
	Without Migration Background	With Migration Background	Without Migration Background	With Migration Background
	M (SD)	*M (SD)*	*M (SD)*	*M (SD)*
Mathematical Ability	5.74 (14.23)	5.74 (12.20)	10.60 (13.25)	5.19 (13.00)
Motivation	58.22 (6.35)	54.82 (9.24)	62.81 (13.40)	56.24 (12.93)
Number of tasks	8.44 (1.31)	9.75 (1.77)	9.27 (1.87)	8.33 (2.44)
Difficulty of tasks	0.88 (0.17)	1.02 (0.23)	1.12 (0.32)	0.98 (0.33)
Difficulty of first task	1.53 (0.49)	1.77 (0.48)	1.78 (0.50)	1.56 (0.45)

7. Discussion

The findings of the experiment with the simulated classroom showed a marginally significant interaction between gender and accountability in the ratings of mathematical ability. On a descriptive level, the mathematical ability of girls was underestimated when the accountability was low, while the mathematical ability of boys was overestimated. In contrast, when the accountability was high, the mathematical ability for both boys and girls was overestimated, but to a lesser degree for girls. The only significant effect of accountability was that the teachers posed tasks with a higher difficulty in the high accountability condition.

Admittedly, some gender effects could be shown in the experiment. The mathematical ability and the

motivation in mathematics were rated higher for boys than for girls. However, referring to the first task posed to a student, girls were posed more difficult tasks than boys. The number of tasks posed to boys and girls was equal, without differences in the mean difficulty of the tasks posed to girls or boys.

Additional analyses revealed that not only the students' gender influenced the behavior of the participants towards the students in the simulated classroom, but also a migration background might have an influence. However, an influence of the migration background was only indicated for participants in the low accountability condition. If the accountability was low, the motivation of girls with a migration background was rated higher than for girls without a migration background. In addition, more tasks were posed to boys with a migration background than to boys without migration background.

The finding that girls with a migration background were rated as more motivated than girls without a migration background could be explained by contrast effects. If participants expect girls to be less motivated in mathematics than boys, they might also expect that girls with a migration background are the least motivated students in the classroom. However, in our experiment, the girls in the simulated classroom with a migration background showed the same motivation to take part in the math lesson as girls without migration background and boys (with and without migration background). Therefore, the behavior of girls with a migration background in a simulated classroom might have astonished the participants and therefore the participation of the girls with migration background might has been especially salient for them when they had to rate how motivated the students were. In contrast in the high accountable condition the participants might be more involved in the interaction with the students and were not so aware of the group to which the students belonged.

For the findings concerning the migration background, no explicit hypotheses were derived, because no explicit stereotype for the migration background was assumed, as it is the case for the gender stereotype and mathematical achievement (see above). However, these findings indicate that it might be useful to examine the effect of different migration background on the behavior towards students from different migration backgrounds in further simulated classroom experiments. Albeit, it might be useful to implement other subjects in the simulated classroom, e.g. native language lessons. Concerning languages, there are explicit stereotypes for the performance of boys and girls and for students with a migration background and with no migration background, namely that girls will perform better than boys, and students without a migration background will perform better than students with a migration background.

The comparison of the students with migration background and the students without a migration background supports also assumption that accountability might be used to reduce the effect of stereotypes in school situations. The effects that girls with a migration background were rated as more motivated than girls without a migration background and that more tasks were posed to boys with a migration background than to boys without a migration background were found only in the low accountability condition. These findings are in line with the marginally significant interaction which was found for the mathematical ability ratings. Only in the low accountability condition, girls were rated as less able in mathematics than boys. For the high accountability condition, girls and boys where rated equally. Together these findings suggest that it might be useful conduct further research on the effect of accountability on the quality of assessments in schools. Preferably, further research should be done based on larger sample sizes, because the interaction effects in this study did not pass significance level, despite decent effect sizes.

In sum, the main results of this study showed that boys tend to be favored by teachers. Given the same mathematical ability and motivation, girls were rated as less able and motivated than boys. However, the first mathematical task posed to girls was more difficult than the tasks posed to boys.

8. Conclusion

In this simulated classroom study, we were able to test the effect of accountability and the effect of the gender category on teacher decisions in an experimental setting.

Positively we can register that the migration background has not a negative effect on the assessment of the students. Instead, girls with migration background were rated as more motivated in mathematics and boys with a migration background got more tasks to solve, indicating a special engagement with these boys.

A limitation of these findings is that the students with a migration background were chosen to represent a specific migration background, namely from Portugal. Although these students are the most common group of migrants in Luxembourg, it limits the generalization of the study's findings to other immigrant groups in Luxembourg.

Another limitation is the relatively short interaction time between the students and the teacher in a rather abstracted experimental setting. In real classroom settings, teachers have the possibility for more extensive interaction with the students as well

as a considerably larger variety of possible performance indicators, e.g. additional verbal communication with the students.

Despite all the methodical limitation of this study, their findings might indicate severe risks for real classroom settings.

If girls receive more difficult tasks in the beginning of a math lesson, it is more likely that they will respond incorrectly. On the one hand, this may lead girls to be less motivated to take part in math lesson. On the other hand, this may also shape the expectations of the teacher in a certain direction. The teacher may see girls as less competent than boys because of the higher rate of wrong answers, inducing a self-fulfilling prophecy for the expectation of mathematical ability.

Furthermore, the effect of a self-fulfilling prophecy is also problematic with respect to the preference of teachers, expressed in the mathematic ability and motivation ratings clearly favoring boys. In a study by Rosenthal and Jacobson (1968) [20], it could be shown that such expectancies may influence the real ability of children. When a certain student is seen as less able, he or she may become less able in solving subsequent tasks. This may be also true for girls. When they are repeatedly exposed to the stereotype that girls are worse in mathematics than boys, they may be impaired in their development of mathematical ability and actually become less able in solving mathematical problems than boys.

These findings may also function as an explanation for the stereotype threat effect, which is still today observed in the STEM-Subjects (science, technology, engineering, and mathematics) [19]. Girls are labeled as less proficient in the STEM domains and often do not receive a good introduction to this domain. After some time, girls label themselves as less proficient than boys and their strategies in preparing for the subjects suffer under these expectations, potentially leading to worse achievements in this area [19].

However, this experiment also showed a possible starting point for the improvement of this situation. Although the effect for the accountability manipulation was only marginally significant, the assessment of girls was better when accountability was high. Further research in this direction might explain how increased accountability may influence the assessment of girls in the STEM domain. Another possibility for the improvement of the assessment for girls in the STEM domain might be to inform teachers and practitioners about this effect and therefore help to improve girls' self-esteem and work against the demotivation of girls in this domain.

9. Acknowledgement

The research reported in this article was supported by a grant TR-PDR BFR08-044 to the first author by the Fonds National de la Recherche Luxembourg.

10. References

[1] Gresham, F. M., MacMillan, D. L., & Bocian, K. M., Teacher as tests: Differential validity of teacher judgments in identifying students at-risk for learning difficulties, *School Psychology Review*, NASP, Bethesda, 1997, Vol. 26, pp. 47-60.

[2] Maaz, K., Trautwein, U., Lüdke, O., & Baumert, J., Educational Transitions and Differential Learning Environments: How Explicit Between-School Tracking Contributes to Social Inequality in Educational Outcomes, *Child Development Perspectives*, Wiley, Oxford, 2008, Vol. 2, pp. 99-106.

[3] Bennet, R. E., Gottesmann, R. L., Rock, D. A., & Cerullo, F, Influences of Behavior Perception and Gender on Teachers' Judgments of Students' Academic Skills, *Journal of Educational Psychology*, APA, Washington, 1993, Vol. 85, pp. 347-358.

[4] Lenkholm, A. K., & Cliffordson, C., Discrepancies between school grades and test scores at individual and school level: effects of gender and family background, *Educational Research and Evaluation*, Taylor & Francis, London, 2008, Vol. 14, pp. 181-199.

[5] De Boer, H., Bosker, R. J., & van der Werf, M. P. C., Sustainability of Teacher Expectation Bias Effects on Long-Term Student Performance, *Journal of Educational Psychology*, APA, Washington, 2010, Vol. 102, pp. 168-179.

[6] Burton, R., Reichert, M., Brunner, M., Keller, U., Boehm, B., & Martin, R. (2007). Befunde zum Luxemburger Regelschulwesen [Results for the Luxembourgish regular school system]. In Ministerium für Erziehung und Berufsbildung, SCRIPT, Universität Luxemburg, Forschungseinheit EMACS (Hrsg.), PISA 2006 - Nationaler Bericht Luxemburg, Luxembourg: Author, pp. 31-45.

[7] Aronson, J. & Steele, C. M., Stereotypes and the fragility of human competence, motivation and self-concept. In C. Dweck & E. Elliot (Eds.), *Handbook of competence and motivation*, (pp. 436-456) Guilford, New York, 1989.

[8] Kahneman, D. & Tversky, A., On the psychology of prediction, *Psychological Review*, APA, Washington, 1973, Vol. 80, pp. 237-251.

[9] Bodenhausen, G. V., Macrae, C. N., & Sherman, J. W., On the dialectics of discrimination: Dual process theories in social stereotyping. In S. Chaiken & Y. Trope (Eds.),

Dual process theories in social psychology, (pp. 271-290) Guilford, New York, 1999.

[10] Fiske, S. T., Lin, M. H., & Neuberg, S. L., The Continuum Model: Ten Years Later. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology*, (pp. 231-254) Guilford, New York, 1999.

[11] Castel, A. D., McCabe, D. P., Roediger, H. L., & Heitman, J. L., The Dark Side of Expertise, *Psychological Science*, SAGE, Thousand Oaks, 2007, Vol. 18, pp. 3-5.

[12] Gigerenzer, G., Why Heuristics Work, *Perspectives on Psychological Science*, SAGE, Thousand Oaks, 2008, Vol. 3, pp. 20-29.

[13] Chaiken, S. Liberman, A., & Eagly, A. H., Heuristic an systematic processing within and beyond the persuasion context. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought*, (pp. 212-251) Guilford, New York, 1989.

[14] Fiske, S. T., Kinder, D. R., & Later, W. M., The Novices and the expert: Knowledge-based strategies in political cognition, *Journal of Experimental Social Psychology*, Elsevier, Amsterdam, 1983, Vol. 19, pp. 381-400.

[15] Hoge, R. S., Gottesmann, R. L., Rock, D. A., & Cerullo, F, How invalid are marks assigned by teachers, *Journal of Educational Psychology*, APA, Washington, 1952, Vol. 43, pp. 218-228.

[16] Lerner, J. S., & Tetlock, P. E., Accounting for the Effects of Accountability, *Psychological Bulletin*, APA, Washington, 1999, Vol. 125, pp. 255-275.

[17] Fiedler, K., Walther, E., Freytag, P., & Plessner, H., Judgment Biases in a Simulated Classroom – A Cognitive –Environmental Approach, *Organizational Behavior and Human Decision Processes*, Elsevier, Amsterdam, 2002, Vol. 88, pp. 527-561.

[18] Fiedler, K., Freytag, P., & Unkelbach, C., Pseudocontingencies in a Simulated Classroom, *Journal of Personality and Social Psychology*, APA, Washington, 2007, Vol 92, pp. 665-677.

[19] Appel, M., Kronberger, N., & Aronson, J, Stereotype threat impairs ability building: Effects on test preparation among woman in science and technology, *European Journal of Social Psychology*, Wiley-Blackwell, West Sussex, 2011, Vol. 41, pp. 904-913.

[20] Rosenthal, R. & Jacobson, E., *Pygmalion in the classroom*, Holt, Rinehart & Winston, New York, 1968.

