

Improving Teachers' Judgments: Accountability Affects Teachers' Tracking Decision

Sabine Glock, Sabine Krolak-Schwerdt, Florian Klapproth, Matthias Böhmer
University of Luxembourg

Abstract

In some European countries, teachers select students for entry into different secondary school tracks on the basis of students' achievement level. In Luxembourg, teachers join a council to select students. PISA provided evidence that students with an immigration background and/or low socioeconomic status are underrepresented in the highest school track. The question arises whether teachers' tracking decisions are biased towards non-performance related cues. Dual process theories of judgment formation suggest accountability to be a moderator of judgment accuracy. Judgments of highly accountable teachers should be less biased by non-performance related cues than those of teachers with low accountability. In groups such as the teacher councils, diffusion of responsibility may occur, thereby reducing accountability of the individual group members. We designed two experiments to investigate whether teachers' tracking decisions differ under different levels of accountability. In both studies, teachers in the high accountability condition did rely solely on performance related cues. Increasing accountability for teachers' decisions could reduce biases in tracking decisions for students with immigration background and low socioeconomic status.

1. Introduction

Educational systems differ within Europe and they differ between Europe and the USA. Nevertheless, they also share some aspects. One such aspect which is common in different European educational systems is school tracking. For instance, in Germany, Switzerland, Austria, and Luxembourg, the educational system consists of different secondary school tracks that are situated in different schools, and tracking is based on students' achievement levels. These tracks differ in the qualifications the students can acquire, ranging from a qualification for university entrance to a qualification for very limited job areas. Similarly, in the US, students are placed in academic or vocational

tracks, or in advanced, regular, or remedial tracks in secondary school. Tracking is used to ensure that students with similar/comparable academic abilities share the same classroom environments and instruction [1]. Empirical research on tracking in different countries has provided evidence that students with low socioeconomic status (SES) and minority students are disadvantaged. In Germany and Luxembourg for example, students with an immigration background and low SES are overrepresented in the lowest school track (where they can only acquire a qualification for limited vocational education) whereas they are underrepresented on the highest school track (where they can acquire a qualification for university entrance) [2]. In many European countries with tracking systems, primary school teachers decide upon the placement in the secondary school track.

In Luxembourg, students are tracked after six years of primary school into three clearly hierarchical school tracks. Students are oriented toward the *Enseignement Secondaire* (ES) when their academic achievement is above average. Students can acquire a qualification for university entrance when they successfully attend this school track. Students with average academic achievement are generally placed in the *Enseignement Secondaire Technique* (EST). Here they can also acquire a qualification for university entrance but the education is more vocational in nature, as students are prepared for different job areas. Students with major learning difficulties and below average achievement profiles are oriented toward the *Enseignement Préparatoire* (PREP). Students who attend this school track can acquire qualifications for very limited job areas. The tracking decision is made by a council consisting of the students' primary teacher, one secondary school teacher of each secondary school track, and the responsible school inspector. However, in groups such as the teacher councils, processes of diffusion of responsibility may occur [3], which may reduce the accountability of the individual group members.

In Luxembourg, the tracking decision is of particular importance for the students' future academic careers because the council's decision is mandatory,

i.e. parents must follow this decision and are not allowed to freely choose a secondary school track for their children. Because of the importance of the tracking decisions for students' future careers, our study asked the question whether increasing the accountability of decision makers would decrease any bias in the decisions made.

Large-scale studies try to answer the question of which student cues are taken into account when teachers make their decisions. There is ample evidence that school grades and academic achievement are the best predictors of teachers' tracking decisions [2]. In addition, social factors like immigration background and parents' socioeconomic and educational status also predicted teachers' decisions, even after controlling for academic ability [2]. Nevertheless, by employing multiple regression analyses, these studies suffer from high intercorrelations between the predictor variables analyzed. For instance, school grades and scores of achievement tests are highly correlated with the socioeconomic background of students [4]. This methodological problem affects the results of the multiple regression analyses and limits the predictive validity of the results [5]. Thus, the aim of our study was to investigate whether teachers' tracking decisions were biased against immigration background and SES when using an experimental setting with a set of independent predictors.

2. Theoretical Background

Dual process theories of impression and judgment formation [6][7] suggest that people generally tend to rely on salient cues when processing information about a target person and when forming an impression or judgment of a target person. The more salient the cues are, the stronger is their influence on the resulting judgment. Each piece of information to which perceivers attend in a particular situational context might be considered a salient cue [7]. For instance, when teachers are giving written grades, not only the achievement score in a test might be a salient cue, but also a student's handwriting style. In another situation, for example, when giving oral grades, salient cues might include the correctness of a given answer or the manner in which the answer is given. Thus, the salience of cues differs from situation to situation, thereby also varying the extent to which they are used for impression and judgment formation.

Dual process models assume that there are two different ways of social information processing and judgment formation. One way is relatively automatic and relies on cues which activate social stereotypes. This category-based [7] or heuristic [8] strategy often results in judgments which are biased through the activated stereotype. This process has the advantage

that it occurs quickly and without much cognitive effort, thereby saving cognitive capacity. In this way of judgment formation, cues are salient when activating stereotypes. As a consequence, mainly stereotype-related cues become important when considering the given target person's attributes. The other, more controlled strategy is information integrating [7] or rule-based [8] processing, where cues are considered salient when they provide the perceiver with an accurate judgment, as they are unrelated to stereotypes but instead refer to the attributes of the target person (e.g., individual behaviours or performance related cues).

In many situations, the motivation of the perceiver determines which type of information processing comes into play [6][7]. In case of low motivation to form an accurate judgment, people tend to rely on the heuristic strategy using salient cues originating from social stereotypes, which often leads to a biased judgment. However, in case of high motivation to form an accurate judgment, people tend to use the rule-based strategy, which relies on attributes of the individual target person [6][7]. One factor contributing to variation in accuracy motivation is accountability [9]. People feeling highly accountable for their judgments engage more in rule-based thinking and judgment formation than people who feel less accountable. Less accountable persons engage more in heuristic thinking. Accountability may be increased by emphasizing the importance of the decision [10] and by highlighting personal responsibility for the decision. The responsibility for the decision may be increased by the need to justify the judgment to others [10].

Dual process models can also be applied to person judgment formation in professional domains [11], particularly in the educational domain [12]. For example, it was demonstrated that teachers who were made to feel accountable by emphasizing the importance of their decisions made less heuristic judgments than teachers who felt less accountable because their decisions had no further consequences [12].

3. Research Question

In the Luxembourgish school system, a council makes the tracking decision. Although the primary school teachers have an important say in this decision (their opinion/vote counts double), the council and not the individual teacher is held responsible for the decision. Hence, the individual teacher might attribute the responsibility to other group members and not to him- or herself. In group decisions, a diffusion of responsibility may occur [3], since the individual group members do not feel as accountable as an individual decision maker. Thus, it might be possible that primary

school teachers feel low accountability when making their tracking decisions in a group. Even when the decision is highly important for the students' future careers, teachers' perceived accountability for the decision might be low because they are not solely responsible for the judgment and it is up to the council to justify the decision to the parents.

According to dual process models of impression and judgment formation, which assume accountability as a moderator of judgment formation as has been outlined above, tracking decisions might be improved, i.e. less affected by stereotypes, when the perceived accountability of the decision makers is increased. Thus, we expected teachers who feel relatively less accountable for their decisions to tend to use the heuristic strategy and consider stereotypical person attributes. Within the domain of teachers' assessments, the stereotypes that teachers frequently use include socio-economic background (SES) and immigration status of students [13]. Consequently, the heuristic strategy should encourage teachers to not only use performance related cues, such as the school grades and the test scores, but also non-performance related cues like immigration background and/or SES to form a judgment. Accountability should decrease when teachers have no need to justify their decisions to others, for instance to the parents, and when the decision is of no particular importance for the teacher or for the student. In contrast, teachers who feel more accountable should use a rule-based strategy employing the rule that performance related cues of the individual student provide the best information for making an accurate decision. Accountability should increase when teachers have to justify their decisions and they are personally accountable for the judgment, and when the decisions are highly important for the students' lives. Furthermore, we expected teachers who make decisions as a member of the council to be more likely to employ the heuristic strategy to make a judgment compared to teachers who make decisions individually. Therefore, we expected teachers to consider immigration background or SES of the student while forming the judgment as a member of the council.

We designed two studies to investigate these hypotheses. In both studies, Luxembourgish primary school teachers participated. In Study 1, participants were required to make tracking decisions for 16 fictitious students. In Study 2, participants had the same task but were additionally required think aloud while gathering information and making decisions.

4. Study 1

Teachers received vignettes describing 16 fictitious students. Teachers were asked to decide which

secondary school track each student should attend. When participants had finished reading and judging, they completed a questionnaire assessing their demographic characteristics and their perceived accountability for the tracking decisions in the study.

4.1. Participants and Design

Fifty-four Luxembourgish primary school teachers (34 female) with a mean age of 37.98 years ($SD = 11.06$) participated in Study 1. We recruited participants via the school presidents and the school inspectors. Teachers had a mean teaching experience in Luxembourgish primary schools of 14 years ($SD = 10.99$). Teachers received no payment for their participation.

We varied the degree of accountability between participants, which resulted in three conditions (high, low, council). Participants were randomly and in equal number allocated to the conditions. Moreover, cues in the student descriptions were varied within the participants. The profile of each student consisted of seven cues: school grades, test scores, nationality, SES, working and learning habits, social behaviors, and gender. Our dependent variable was the tracking decision.

4.2. Materials

In each vignette, the student description entailed seven cues which were (1) school grades in the main subjects. In Luxembourg, the main school subjects are German, French, and Mathematics. The grades were averaged across the subcomponents of the school subjects (e.g. in German: reading, writing, speaking) and across the last year of primary school; (2) Scores of standardized school performance tests conducted in the main subjects were provided as percentages of correctly answered questions averaged across the subcomponents (e.g. in French: reading comprehension, grammar, listening comprehension); (3) working and learning habits shown in descriptions of how the students did their homework and whether they fully used the time given for written work or had problems finishing their written work in a reasonable period of time; (4) social behaviours illustrated by information about the students' behaviour during instructions and school recess; (5) nationality information given indirectly through the language the students spoke at home; (6) socio-economic status indicated by the occupation of the students' fathers; (7) gender.

The cues were constructed as dichotomous variables. School grades were either above or below average, test scores were either high or low, working and learning habits as well as social behaviours were

either positive or negative, nationality was either with or without immigration background, socio-economic status was either high or low, and gender either male or female.

Cues were combined in a way that ensured that for any combination of the values of the cues the intercorrelations between the cues were minimal. In fact, school grades, test scores, nationality, and SES were completely independent from each other, which resulted in 16 case descriptions containing four cues with zero intercorrelations. To complete the case descriptions, the remaining three cues were added by distributing the values of these cues randomly to the 16 cases, yielding maximum intercorrelations of $r = .25$.

To manipulate accountability, we created three different instructions. In the high accountability instructions, teachers were asked to imagine that they were solely responsible for the tracking decisions and that these decisions would influence the future educational and occupational careers of the students. In the low accountability instruction, teachers were asked to imagine a situation in which a colleague would ask them for advice concerning the tracking decisions for the students of his class and that they were just required to provide their opinion without commitment. In the council instruction, teachers were asked to prepare the tracking decisions for the council and were informed that the final tracking decisions would be made by the council. This instruction corresponded to the actual tracking procedure in Luxembourg.

To validate the different instructions, six Luxembourgish primary school teachers read the different instructions and rated their perceived accountability for each set of instructions on a Likert-Scale ranging from 0 (no perceived accountability at all) to 5 (high perceived accountability). We computed simple effect tests to investigate whether the different instructions induced different levels of perceived accountability. Teachers felt more accountable after reading the high accountability instruction ($M = 5.00$, $SD = 0.00$) than after reading the low accountability instruction ($M = 2.67$, $SD = 1.37$), $t(5) = 4.18$, $d = 2.41$, $p < .01$ (one-tailed). The council instructions produced marginally lower perceived accountability ($M = 4.5$, $SD = 0.84$) than the high accountability instructions, $t(5) = 1.46$, $d = 0.84$, $p = .10$ (one-tailed), and significantly higher accountability than the low accountability instructions, $t(5) = 2.20$, $d = 1.61$, $p < .05$ (one-tailed).

4.3. Procedure

A test booklet containing the 16 vignettes with student descriptions was administered. We randomly allocated participants to one of the three different experimental accountability conditions. The

instructions were printed on the first page of the booklet. Then the 16 case descriptions were presented in random order. After each case description, participants were asked to decide upon the appropriate secondary school track for the described student, choosing between three options: ES (highest track); EST (middle track); and PREP (lowest track). Participants had the opportunity to look at all 16 descriptions before making their judgments. After finishing the 16 student descriptions, participants were required to fill out a questionnaire assessing personal data such as gender, age, and teaching experience. To check our accountability manipulation, teachers were additionally asked to indicate their perceived accountability during the task on a percentage scale ranging from 0 (no perceived accountability) to 100 (highest perceived accountability). In the end, participants were thanked and debriefed.

4.4. Results

4.4.1. Manipulation check. To check whether our instructions induced different levels of accountability, we conducted a one-way ANOVA with the ratings of perceived accountability as dependent variable. ANOVA yielded a significant main effect, $F(2,49) = 3.45$, $\eta_p^2 = 0.12$, $p < .05$. The high accountability instruction induced a higher degree of accountability ($M = 85.67$, $SD = 16.21$) than the low accountability instruction ($M = 68.35$, $SD = 25.57$), $t(33) = 2.41$, $d = 0.81$, $p < .05$ (one-tailed). The mean rating for the council instructions ($M = 78.71$, $SD = 16.05$) did not significantly differ from the high accountability instructions, $t(33) = 1.09$, $d = 0.43$, $p = .14$ (one-tailed). However, there was a marginally significant difference between the council instructions the low accountability instructions, $t(32) = 1.55$, $d = 0.49$, $p = .07$ (one-tailed).

4.4.2. Tracking decisions. First, we analysed how often each track was recommended; 42.47 % were ES-decisions, 51.39 % were EST-decisions, and 6.13 % were PREP-decisions. These data corresponded to the actual tracking decisions in Luxembourg [14], supporting the ecological validity of our experimental procedure. Because teachers chose the PREP track for only a few students, we dichotomized the tracking decisions into ES- and non-ES-decisions. To test our hypotheses regarding the influence of different levels of accountability, we conducted a multiple logistic regression analysis for each experimental condition using this dichotomous criterion (see Table 1 for all odds ratios).

Table 1. Odds ratios for the seven cues in each accountability condition in Study 1.

	low account-ability	council	high account-ability
school grades	85.18*	31.86*	50.67*
test scores	14.01*	12.32*	10.26*
Nationality	3.21*	1.84	1.68
SES	1.20	1.54	1.53
working and learning habits	2.42	2.01	2.81*
social behavior	1.53	1.38	1.50
gender	2.04	2.11	1.56

Note: * $p < .05$

For the high accountability condition, the performance related cues school grades, test scores, and working and learning habits predicted teachers' tracking decisions, whereby higher scores were associated with a higher chance of choosing the highest track. In the low accountability condition, working and learning habits was not considered as significant cue. However, only in the low accountability condition teachers' tracking decisions were additionally based on the nationality of the student as non-performance related cue. More specifically, in the low accountability condition and given similar academic performances of pupils, students without immigration background had approximately a three times higher chance of being recommended to the highest track than students with immigration background.

4.5. Discussion

Drawing on dual process models [6][7] and the notion of accountability as a moderator of judgment formation [9], we designed Study 1 to test whether different levels of accountability could influence teachers' tracking decisions.

We expected that teachers in the low accountability condition or deciding as members of a council would employ a heuristic judgment formation strategy whereas teachers in the high accountability condition would use a rule-based strategy. Heuristic judgment formation would be indicated by considering non-performance related cues such as immigration background or the SES of the student for the decision while the rule-based strategy was indicated by taking solely the performance related information into consideration. As expected, decisions of teachers who felt a low degree of accountability were influenced by immigration background of the student, suggesting the use of a heuristic strategy. Even though we took care that performance related and nationality related information were uncorrelated, we found an influence of the immigration background of the student in the low accountability condition. This finding corresponds

with the results of large-scale studies, which have provided evidence that immigration background is a predictor of teachers' tracking decisions [2]. Thus, in the low accountability condition, teachers' tracking decisions seem to be affected by immigration background even though students share the same level of academic achievement. In contrast, teachers who felt highly accountable for their tracking decisions relied solely on performance related cues, adopting a rule-based strategy. Thus, one factor which could turn heuristic judgment formation into rule-based judgment formation is accountability. This implies that making teachers highly accountable for their tracking decisions by giving them sole responsibility for making their decisions and communicating these to students and parents, could improve the objectiveness (and possibly the accuracy) of their judgments [9] in the sense that only performance related cues would be weighted to form a judgment.

For teachers who made their decision as members of the council, Study 1 provided inconsistent results. We expected teachers who made their decisions in the council and who had no need to justify their decisions to others to exhibit heuristic processing. The regression analysis for the tracking decisions indicated a rule-based strategy, as teachers solely relied on performance related cues. It is possible that teachers felt more accountable in the council condition because they had two votes and, therefore, their judgment weighed more in the council's decision. The manipulation check confirms a medium level of perceived accountability between high and low accountability, as the council condition did not differ from the high accountability condition. This unexpectedly high induced level of accountability may have produced this unexpected result.

In sum, the results of Study 1 imply that teachers' judgments may be biased through non-performance related cues, particularly immigration background, under low accountability conditions. Minority students do frequently show lower academic performance than students without immigration background [15]. Because immigration background was completely independent of academic performance in our vignettes, however, we can rule out the possibility that the relationship between nationality and tracking decision was a side-effect of an empirical relationship between performance and nationality. To our knowledge, this is the first demonstration of the influence of immigration information on teachers' judgments which is not also confounded by performance related information.

Nevertheless, Study 1 is limited in several ways. First, the vignettes were administered as a booklet and the participants could make their decisions after reading all student descriptions. This may have induced a frame of reference effect, on which teachers

may have relied [16]. This fact could have contributed to the results, since teachers could compare the students' cues and adjust their decisions accordingly. Teachers' judgments often depend on the achievement level of the class [16]. Second, due to the booklet format, we were not able to gain insight into information processing during judgment formation.

5. Study 2

To address the limitations of Study 1, we presented each vignette via computer in Study 2, ensuring that teachers had to work on the student descriptions one at a time. Additionally, by using a computer, we were able to assess reading times and judgment latencies to get a deeper insight into cognitive processes during judgment formation. Because heuristic information processing and judgment formation occurs quickly while rule-based processing is more time-consuming [17] reading times and judgment latencies could be additional indicators for the different processing strategies. Thus, we expected that reading times and judgment latencies should be decreased in the low accountability and council condition compared to the high accountability condition.

Moreover, complexity of thoughts increases as accountability increases [10]. Therefore, we used a think aloud method to shed more light on cognitive processes. We expected that thoughts should be more complex in the high accountability condition than in the low accountability or council conditions. Regarding the tracking decisions, we tested the same hypotheses as in Study 1.

5.1. Participants and Design

Sixty Luxembourgish (34 female) primary school teachers participated in Study 2. We recruited participants via the school inspectors. Teachers had an average of 15.98 ($SD = 10.77$) years' experience working in primary schools and were on average 40.08 ($SD = 10.66$) years old. Teachers received no payment for their participation. As in Study 1, we varied the accountability level between participants and included the same seven cues (predictor variables). Dependent variables were the tracking decisions, the reading times, and judgment latencies.

5.2. Materials

We used the same vignettes as in Study 1 with three minor modifications. One modification concerned the school grades, which were averaged across the school year but divided into the subcomponents of the school subjects. More specifically, the German and French school grades were split into: (1) one grade for writing,

and (2) one grade for speaking. The second modification concerned the information about test scores, where we now provided the score percentiles for the students. This had the advantage that the information was independent of the task difficulty and that teachers could see how the students' test scores were ranked within the whole cohort. The third modification was that we also provided the test scores for the individual subtests. The modifications derived from an advice of some school inspectors who validated the materials.

5.3. Procedure

We visited participants in their schools. We randomly allocated teachers to one of the three accountability conditions. Participants were seated in front of the computer screen and then the first set of instructions appeared, informing participants about the procedure of the study and the Think aloud requirement. After this, the accountability instructions appeared on the computer screen. Subsequently, the 16 student descriptions were presented. Teachers had to make the tracking decision for each of the 16 students by pressing "1" to indicate ES, "2" to indicate EST, and "3" to indicate PREP. During teachers' work on the student descriptions, the experimenter recorded teachers' thoughts with a voice recorder. After finishing the last student description, teachers were asked to fill out the same demographic questionnaire and rank their perceived accountability as in Study 1. Finally, participants were debriefed and thanked.

5.4. Results

5.4.1. Manipulation check. To check whether our instructions induced different levels of perceived accountability, we conducted a one-way ANOVA. The ANOVA yielded no significant results, $F < 1$. Teachers in the high accountability condition ($M = 77.50$, $SD = 17.81$) felt as accountable as teachers in the council ($M = 76.00$, $SD = 20.81$) and in the low accountability condition ($M = 74.44$, $SD = 28.59$).

5.4.2. Tracking decisions. First, we analysed how often teachers recommended each secondary school type; 42.08 % decisions were made for the highest track, 52.29 % for the middle track whilst the lowest track was recommended in 5.63 % of all cases. As in Study 1, these results correspond to the actual tracking recommendations in Luxembourg. As in Study 1, we dichotomized the tracking decisions into ES- and non-ES decisions. In our further analyses, we used this dichotomous variable as criterion. To investigate the influence of accountability on teachers' tracking decisions, we analysed the tracking decisions using

multiple logistic regression analysis for each experimental (accountability level) condition (see Table 2 for all odds ratios).

Table 2. Odds ratios for the seven cues in each accountability condition in Study 2.

	low account-ability	council	high account-ability
school grades	237.51*	75.93*	46.19*
test scores	153.60*	27.38*	63.23*
nationality	1.62	1.93	0.93
SES	0.69	0.77	0.66
working and learning habits	1.83	5.17*	2.90*
social behavior	1.05	1.70	1.22
gender	1.08	1.98	1.19

Note: * $p < .05$

Irrespective of accountability level, teachers' tracking decisions were based on school grades and test scores. Working and learning habits were additionally taken into account by participants in the council and the high accountability condition.

5.4.3. Reading Times. The time participants needed to read each vignette was assessed. To ensure that reading times not depend on length of student descriptions, we divided the reading time for the whole description of a student by the number of characters. We submitted these relative reading times to a one-way ANOVA with *accountability* (low vs. high vs. council) as factor. The main effect did not reach significance, $F < 1$.

5.4.4. Judgment latencies. The time participants needed to make the tracking decisions for each student was assessed. To test our hypotheses regarding judgment latencies in ms, we conducted a one-way ANOVA with *accountability* (low vs. high vs. council) as factor. ANOVA results revealed no significant main effect of accountability condition, $F < 1.10$.

5.4.5. Reference to cues. We consulted think aloud protocols to investigate whether teachers referred to the different cues and how often they considered the different cues in their thoughts, as an indicator of thought complexity. One independent judge rated the protocols corresponding to the reference frequency of each cue. Because each cue consisted of a different amount of information (i.e. school grades consisted of five pieces of information; working and learning habits consisted of three pieces of information), we divided these frequencies by the number of pieces of information to ensure that the references of the different cues were comparable. The resulting relative numbers of references were submitted to a 3 x 7 mixed

ANOVA with *accountability* (low vs. high vs. council) as between-subjects factor and *cue* (school grades vs. test scores vs. nationality vs. SES vs. working and learning habits vs. social behavior vs. gender) as within-subjects factor. The ANOVA yielded a main effect for *cue*, $F(6, 336) = 57.57$, $\eta_p^2 = 0.51$, $p < .001$ (see Table 3 for all means and standard deviations).

Table 3. Means and standard deviations for the relative numbers of references of the cues in Study 2.

	mean	SD
school grades	5.71	3.03
test scores	2.97	1.60
nationality	3.22	4.25
SES	2.88	4.62
working and learning habits	12.95	9.80
social behavior	5.94	4.62
gender	0.63	1.63

In order to investigate the question whether the number of references had an effect on the importance of cues in the accountability conditions, we compared the number of references between conditions for those cues which differed in their weight for the tracking decisions. These were immigration background as well as working and learning habits. We found no differences between the accountability conditions, all $F_s < 1$.

5.5. Discussion

Study 2 provided new results on the effects of accountability: High accountability induced rule-based processing, as teachers solely relied on performance related cues, but the council instruction had the same effect. Teachers in the low accountability condition did not rely on all performance related cues, as they did not consider working and learning habits when making their decisions. This finding contradicts our assumptions. Results regarding the frequency of cue reference showed that participants referred to almost all cues but they did not consider all cues in decision making. This suggests that even though teachers thought about for instance, SES, they did not rely on SES while making the decision. This might be due to the special think aloud situation, in which teachers were always monitored by the experimenter. The presence of others often induces the desire to be socially correct [19]. Perhaps teachers thought they were required to be socially and politically correct, thus considering non-performance related cues in thoughts but not in decisions. Finally, in contrast to our expectations, accountability had no influence on reading times and judgment latencies. These results

also imply an influence of the experimental think aloud situation.

6. General Discussion

In Luxembourg, students are selected for entry into different secondary school tracks depending on their achievement level. Research has provided evidence that students with immigration background and low SES are underrepresented in the highest track. Hence, we designed two studies to investigate whether teachers' tracking decisions are biased through non-performance related cues. Drawing on dual process theories [6][7] and the idea that accountability functions as a moderator of information processing and judgment formation [9], we examined whether the decision making process could be improved (i.e. become less biased) by making teachers highly accountable for their decisions. The results of our two studies support the assumption that teachers who feel highly accountable for their decisions employ the rule-based strategy, resulting in more accurate (less biased) performance related judgments.

Study 1 showed that high accountable teachers solely relied on performance related cues whereas teachers in the low accountability condition additionally took non-performance related cues into account. Teachers who made their decisions as a member of the council did not consider non-performance related cues in their tracking decisions. However, the results of Study 1 are inconsistent with the results of Study 2. Contrary to Study 1, teachers who felt low accountability took only performance related cues into account, but they did not rely on working and learning habits. In Study 2, our manipulation check showed that the accountability manipulation did not work as well as in Study 1. The inconsistency between the results of Study 1 and Study 2 might therefore be due to the differences in accountability elicited in the two experiments. In Study 1, participants worked on the vignettes and were required to make tracking decisions. In Study 2, teachers were asked to express their thoughts verbally during the decision making process. Thus, the different experimental situations might have evoked different types of accountability per se, as Study 1 only induced outcome accountability whereas Study 2 elicited process accountability. Outcome accountability matches our accountability manipulation: participants had to account for their judgments [20], particularly for their tracking decisions. Therefore, our accountability manipulation worked well in Study 1, as the experimental situation and the accountability instruction only affected participants' outcome accountability. In contrast, the think aloud procedure might have provoked process accountability; that is,

participants might have felt the need to account for the ways they processed the given information [20], particularly as participants were required to explain the way they made their decisions. As a result, in Study 2, the experimental situation and the accountability instructions might have induced both process and outcome accountability. The results imply that process accountability could have overruled the effects of outcome accountability, as process accountability is suggested to elicit rule-based decision making [20]. One may conclude that introducing a strategy that induces process accountability could also be a valuable tool in improving teachers' decisions. Asking teachers, for instance, to explain how they review the student information and use it to come to a decision could induce process accountability and thus lead to unbiased decisions.

Our experimental results add to the results of large-scale studies. They also show that students with immigration background are disadvantaged when it comes to school tracking decisions. Considering the fact that Luxembourgish teachers make tracking decisions as members of a council and that under these conditions accountability for the decision might be low, our results may provide one possible explanation of the disadvantage of students with an immigration background in Luxembourg. Nevertheless, further research is needed, as our student descriptions were fictitious and teachers were provided with minimal information about the students. Future research should rely on authentic student information to see whether these cases solicit similar results. However, relying on authentic student information always implies high correlations between different student cues [15], which makes it harder to separate the effects of different predictors. In contrast, experimentally created student descriptions provide the exclusive procedure to disentangle influences of performance and non-performance related cues to estimate the actual weights of non-performance related cues in teachers' tracking decisions. Future research should combine both authentic student information and experimental procedures to investigate the validity of experimental results. One could also argue that our findings stem from a non-naturalistic setting and that the results of our studies do not hold for teachers' tracking practice in daily life. However, the data concerning the recommendation frequency for each track corresponded to the distribution of actual tracking decisions. Accountability improved tracking decisions in an experimental situation which had no real consequences for students or teachers. One could expect even stronger effects when increasing accountability among teachers in real-life settings, because in this case actual consequences for teachers

and students will follow, which should further induce perceptions of accountability.

Our results have implications for tracking decision procedures. Considering the main finding of our studies (i.e. high accountability leads to rule-based judgments), introducing a procedure that ensures that teachers are accountable for tracking decisions and that permanently makes teachers aware of their responsibility for the future of students may be valuable in reducing the disadvantage of students with immigration background and low SES. In addition, the question arises whether tracking decisions should be made as individual or as group decisions. Further research should explore whether increased accountability in groups can improve group decisions as well as increased accountability in individual teachers. However, increasing accountability in teachers' judgment procedures irrespective of individual or group decisions could be a valuable tool in improving teachers' judgments about students.

7. Acknowledgements

The experiments reported in this paper were funded by Grant C08/LM/02 from the Fonds National de la Recherche Luxembourg (FNR).

8. References

- [1] G. Ansalone, Poverty, tracking, and the social construction of failure: International perspectives on tracking, *Journal of Children and Poverty*, Routledge, London, 2003, Vol. 9, pp. 3–20.
- [2] K.-H. Arnold, W. Bos, P. Richert, and T. C. Stubbe. Schullaufbahnpräferenzen am Ende der vierten Klassenstufe [Preferences in school career at the end of the fourth grade]. In: W. Bos, S. Hornberg, K.-H. Arnold, G. Faust, L. Fried, E.-M. Lankes, K. Schwippert and R. Valtin (Ed.). IGLU 2006. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich. Waxmann, Münster, 2007, pp. 271–297.
- [3] B. Latané and J. M. Darley, Bystander intervention in emergencies: Diffusion of responsibility, *Journal of Personality and Social Psychology*, APA, Washington, 1968, Vol. 8, pp. 377–383.
- [4] G. Kao and J. S. Thompson, Racial and ethnic stratification in educational achievement and attainment, *Annual Review of Sociology*, Annual Reviews, Palo Alto, 2003, Vol. 29, pp. 417–442.
- [5] J. Cohen, P. Cohen, S. G. West, and L. S. Aiken, *Applied multiple regression/correlation analysis for the behavioral sciences*, Lawrence Erlbaum Associates, Mahwah NJ, 2003.
- [6] M. B. Brewer. A dual process model of impression formation, in: R. S. Wyer and T. K. Srull (Ed.), *Advances in Social Cognition*, Hillsdale, Erlbaum, 1988, pp. 1–36.
- [7] S. T. Fiske and S. L. Neuberg, 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*, Academic Press, New York, 1990. Vol. 23, pp. 1–74.
- [8] M. B. Ferreira, L. Garcia-Marques, S. J. Sherman and J. W. Sherman, Automatic and controlled components of judgment and decision making, *Journal of Personality and Social Psychology*, APA, Washington, 2006, Vol. 91, pp. 797–813.
- [9] J. Lerner and P. E. Tetlock, Accounting for effects of accountability, *Psychological Bulletin*, APA, Washington, 1999, Vol. 125, pp. 255–275.
- [10] P. E. Tetlock, Accountability and complexity of thought, *Journal of Personality and Social Psychology*, APA, Washington, 1983, Vol. 45, pp. 74–83.
- [11] S. T. Fiske, D. R. Kinder and W. M. Larter, The novice and the expert: Knowledge based strategies in political cognition, *Journal of Experimental Social Psychology*, Elsevier, Amsterdam, 1983, Vol. 19, pp. 381–400.
- [12] K. Dünnebier, C. Gräsel, and S. Krolak-Schwerdt, Urteilsverzerrungen in der schulischen Leistungsbeurteilung. Eine experimentelle Studie zu Ankereffekten [Biases in teachers' assessment of student performance: An experimental study of anchoring effects]. *Zeitschrift für Pädagogische Psychologie*, Verlag Hans Huber, Bern, 2009, Vol. 23, pp. 187–195.
- [13] J. E. Ormrod, *Educational psychology: developing learners*, Prentice Hall, Pearson/Merrill, 2006.
- [14] F. Klapproth, S. Krolak-Schwerdt, S. Glock, R. Martin and M. Böhmer, Prädiktoren der Sekundarschulempfehlung in Luxemburg: Ergebnisse einer Large-Scale-Untersuchung [Predictors of the Luxembourgish secondary school recommendations: Findings from a large scale study]. Manuscript submitted for publication. 2012.
- [15] T. S. Dee, A teacher like me: Does race, ethnicity, or gender matter? *The American Economic Review*, American Economic Association, Nashville, 2005, Vol. 95, pp. 158–165.
- [16] H. W. Marsh, The big-fish-little-pond effect on academic self-concept. *Journal of Educational Psychology*, APA, Washington, 1987, Vol. 79, pp. 280–295.
- [17] G. D. Hemsley and H. H. C. Marmurek, Person memory: The processing of consistent and inconsistent person information. *Personality and Social Psychology Bulletin*, Sage Publications, Thousand Oaks, 1982, Vol. 8, pp. 433–438.

[18] J. W. Sherman, A. Y. Lee, G. R. Bessenoff and L. A. Frost, Stereotype efficiency reconsidered: Encoding flexibility under cognitive load. *Journal of Personality and Social Psychology*, APA, Washington, 1998, Vol. 75, pp. 589–606.

[19] P. E. Tetlock and A. S. Manstead, Impression management versus intrapsychic explanations in social psychology: A useful dichotomy? *Psychological Review*, APA, Washington, 1985, Vol. 92, pp. 59–77.

[20] L. Scholten, D. van Knippenberg D, B. A. Nijstad and C. K. W. De Dreu, Motivated information processing and group decision-making: The effects of process accountability on information processing and decision quality. *Journal of Experimental Social Psychology*, Elsevier, Amsterdam, 2007, Vol. 43, pp. 539–552.