Electronic Journal of Research in Educational Psychology

Children's Performance Estimation in Mathematics and Science Tests over a School Year: A Pilot Study

Thomas Roderer, Claudia M. Roebers

Department of Psychology, University of Bern

Switzerland

Correspondence: Thomas Roderer. Department of Psychology, University of Bern, Muesmattstr. 45 CH – 3000 Bern 9. Switzerland. E-mail: thomas.roderer@phsg.ch

Abstract

Introduction. The metacognitve ability to accurately estimate ones performance in a test, is assumed to be of central importance for initializing task-oriented effort. In addition activating adequate problem-solving strategies, and engaging in efficient error detection and correction. Although school children's' ability to estimate their own performance has been widely investigated, this was mostly done under highly-controlled, experimental set-ups including only one single test occasion.

Method. The aim of this study was to investigate this metacognitive ability in the context of real achievement tests in mathematics. Developed and applied by a teacher of a 5th grade class over the course of a school year these tests allowed the exploration of the variability of performance estimation accuracy as a function of test difficulty.

Results. Mean performance estimations were generally close to actual performance with somewhat less variability compared to test performance. When grouping the children into three achievement levels, results revealed higher accuracy of performance estimations in the high achievers compared to the low and average achievers. In order to explore the generalization of these findings, analyses were also conducted for the same children's tests in their science classes revealing a very similar pattern of results compared to the domain of mathematics.

Discussion and Conclusion. By and large, the present study, in a natural environment, confirmed previous laboratory findings but also offered additional insights into the generalisation and the test dependency of students' performances estimations.

Keywords: metacognition, primary school, performance estimation, mathematics, science

Received: 12/18/12 *Initial acceptance: 02/19/13* *Final acceptance: 03/12/13*

Estimación de Logro Infantil en Pruebas de Matemáticas y Ciencias Naturales a lo largo de un Curso. Un Estudio Piloto.

Resumen

Introducción. La habilidad para estimar metacognitivamente el rendimiento de uno/a mismo/a se considera que tiene una importancia crucial para la iniciación de un esfuerzo dirigido a la tarea, así como para activar estragias adecuadas de resolución de problemas y para participar en la detección y en la corrección eficiente del error. Mientras que la habilidad para estimar el propio rendimiento en niños en edad escolar ha sido estudiada ampliamente, esta investigación se ha realizado casi exclusivamente en contextos experimentales altamente controlados, los cuales han considerado unicamente un ensayo experimental.

Metodo. El objetivo de este estudio fue el de investigar esta habilidad metacognitiva en el contexto real de unas pruebas de aptitud en matemáticas. Estas pruebas fueron desarrolladas y administradas por un profesor de quinto grado (10 años de edad), a lo largo del año escolar. Las estimaciones del rendimiento permiten explorar el impacto que tiene la dificultad de las pruebas en la variabilidad de la estimación del rendimiento.

Resultados. Las estimaciones medias del rendimiento estaban generalmente más próximas a la ejecución real, presentando menos variabilidad, en relación con la ejecución de la prueba. Cuando los participantes se agruparon en distintos niveles de rendimiento, los resultados revelaron una mayor precisión en la estimación de la ejecución en aquellos participantes que mostraron un rendimiento alto, en comparación con los participantes con un rendimiento mediano y bajo. Con el fin de explorar si estos hallazgos se podrían generalizar, se analizaron además las pruebas en ciencia de los mismos participantes, los cuales revelaron un patrón de resultados muy similar al del dominio de las matemáticas.

Discusión y Conclusión. De modo general, este estudio naturalista confirma otros hallazgos previos del laboratorio, pero además ofrece información adicional por lo que se refiere a la validad ecológica de las pruebas y a la dependencia de las pruebas de las estimaciones del rendimiento de los estudiantes.

Palabras clave: metacognición, años primarios de educación, estimación del rendimiento, matemáticas, ciencia.

Recepción: 18/12/12 *Aceptación inicial: 19/02/13* *Aceptación inicial: 12/03/13*

Introduction

In the last decade, a growing theoretical and practical interest on aspects of students' self-regulating learning skills can be observed (Baumert et al., 2001; Hacker et al., 2009; Winne & Hadwin, 1998; Zimmerman & Moylan, 2009). Metacognitive processes have repeatedly been found to substantially influence students' academic success and learning progress in the long and short run (Schneider & Pressley, 1997), within shorter periods of time (Körkel & Schneider, 1991; Schneider et al., 1998), and their actual test performance (Roebers et al., 2009; Krebs & Roebers, 2010). Therefore, several authors have emphasized the need for transfer of the theoretical concepts, from basic research on metacognitive development, directly into the educational practice (Carr et al., 1989; Hacker et al., 2009, Kuhn & Dean, 2004; Williams et al., 2002). Such an approach allows assessing the possibility to generalize existing findings to complex, naturalistic educational settings. The current paper presents such an attempt in which 5th grade students' metacognitive ability to estimate their test performance in the domain of mathematics and science. The assessment was carried out within naturally occurring achievement tests over the course of one school year.

Useful theoretical frameworks for metacognitive processes were provided by Flavell and Wellman (1977) on the one side, and by Nelson and Narens (1990), on the other side. Flavell and Wellman suggested a fundamental distinction between declarative and procedural metacognitive knowledge, with declarative metacognition embracing factual knowledge about memory and learning processes as well as factors influencing performance (person characteristics, strategy knowledge, learning context, material, and interactions between these factors). Procedural metacognition relates to processes occurring simultaneously to cognitive operations and can be classified as *monitoring* and *controlling*. Nelson and Narens further categorized these monitoring and controlling processes with regard to the point in time when they potentially exert their influence on the learning process; during encoding, storage, or retrieval of information. The present investigation focused on monitoring as a procedural metacognitive process. Thus, the question was not how much students knew about their own cognitive processes, but how well they were able to monitor their own performance. There are a variety of occasions for monitoring cognitive processes: performance can be rated before, during, or after learning, or before, during, or after taking a test. In this study an investigation was made to distinguish whether primary school students are able to accurately estimate their performance after taking a test by expecting a lower score when their performance was indeed low (as marked by the teacher) compared to when they performed better. The discrepancy between estimated and teacher performance is assumed to be smaller for students with better monitoring abilities. In other words, their monitoring ability is well calibrated in relation to the actual test score, resulting in a high monitoring accuracy. The students' metacognitive calibration was analyzed in relation to actual performance, test difficulty, and school subject, as was done in earlier studies (Desoete & Roeyers, 2006; Howie & Roebers, 2007).

Both, declarative and procedural metacognitive processes have been shown to impact learning outcomes. Knowledge about potentially useful strategies, personal weaknesses, and interactions between task demands and strategies, directly influence strategy use hence indirectly affecting learning outcomes (Schneider et al., 1998; Van der Stel & Veenman, 2008; Veenman & Spaans, 2005). Precise procedural metacognitive skills, such as accurate monitoring of correct and incorrect answers, precede an efficient and successful search for committed errors (Roebers et al., 2009). This leads to control strategies, such as withdrawing previously given answers or correcting errors, thus increasing overall test performance (Krebs & Roebers, 2010). Importantly, student's achievement levels have consistently been found to affect the efficiency of metacognitive processes during the learning process: children with higher school achievement are better able to monitor their learning processes accurately, to evaluate their learning outcome realistically, and to operate efficiently on their own monitoring (Alexander et al., 1995; Coutinho, 2007; Garrett et al., 2006; Krebs & Roebers, 2011; Kruger & Dunning, 1999; Montague & Bos, 1990; Slife et al., 1985; Veenman & Spaans, 2005).

These studies, however, have been conducted under controlled and thus constrained conditions. The tests were designed by researchers, optimized for eliciting monitoring processes and collecting monitoring judgments. In these studies, extra attention is directed to the monitoring part of the testing, thus implicitly or explicitly emphasizing the importance of the adequacy of metacognitive monitoring processes. In addition, the tests were administered by unknown people (i.e., researchers or trained research assistants) and sometimes even without the teachers present in order to minimize students' inhibition. All this and more may have contributed to the students' perception of the testing as a special situation, with only minimal resemblance to the usual school achievement test. Such a setting can be compared with a more natural class room setting, where school achievement tests, in familiar formats are administered by the teacher, covering the actual curricula and relevant for actual grading. According changes in the setting may influence the students' behaviors. In such a natural and familiar setting, more time and resources may be invested in actual problem solving or retrieval, as these are the relevant contents for marking. This may in turn result in diminished monitoring effort, and thus in a less elaborate metacognitive judgment of performance. Similar effects can be expected, if metacognitive judgments are less emphasized than in a study setting and are not seen as relevant or helpful for performance by students. On the other hand, directly reporting metacognitive judgments to the teacher who is the socially and academically relevant reference person, may motivate students to accurately monitor their performance. Collecting performance estimations in a classroom setting without any indications of psychological testing, can yield important insight into the ecological validity of prior metacognitive monitoring research (Bronfenbrenner, 1979). This will help to clarify whether the positive relation between metacognitive monitoring and performance still holds in a natural context.

Researchers and practitioners alike highlight the necessity of establishing empirical investigations on metacognition in educational settings (Schneider & Artelt, 2010; Schoenfeld, 1992). Specifically, when reviewing the existing literature on metacognition in the domain of mathematics, the overall picture that emerges by and large confirms basic research: First, metacognitive processes generally influence mathematics achievement and vice versa (Cohors-Fresenborg, Kramer, Pundsack et al., 2010; Desoete & Roeyers, 2006; Legg & Locker, 2009; Van der Stel et al., 2010). Second, children with learning difficulties in mathematics are typically less accurate in their performance estimations (Garrett et al., 2006; Slife et al., 1985). Third, poor metacognitive awareness is associated with a higher number of errors which were not corrected (Lucangeli et al., 1997). Fourth, specifically training metacognitive monitoring and controlling in the context of mathematics not only increases the accuracy and efficiency of such processes, but may also have a direct benefit for performance on mathematics (Desoete, 2009; Desoete et al., 2003; Mevarech et al., 2010; Pennequin et al., 2010).

Furthermore, research on children's metacognitive monitoring in the domain of science, consequently showed accurate metacognitive monitoring of understanding of science texts in primary school children. In addition there were even positive effects of some interventions on monitoring accuracy (de Bruin, Thiede, Camp, & Redford, 2011; Dunlosky & Lipko, 2007; Dunlosky, Rawson, & Hacker, 2002; Renner, & Renner, 2001). A major drawback of this research, however, is its narrow focus on the monitoring of text comprehension. Although other studies investigated the reproduction of science content, this content was unfamiliar and unrelated to the students' curricula (Roebers et al., 2009; Krebs & Roebers, 2010). To our knowledge, no previous research investigated monitoring of the reproduction of curricula-based content in science. Therefore, in the present study students were asked to estimate their performance not only in math, but also in science tests in a classroom setting. Performance estimations and the correspondence between estimations and actual performance were explored as a function of achievement level in the corresponding school subjects.

Mathematics performance and performance estimations are easily and unambiguously quantifiable, and the detection and correction of errors can well be documented – two important advantages of mathematics for empirical research, as for example opposed to, essay writing. As a consequence, it is widely unknown whether the findings obtained in the domain of mathematics also apply to other school subjects. Science, in our view, can be considered as another school subject in which performance and monitoring processes in the form of performance estimations can be easily assessed (Roebers et al., 2009; Krebs & Roebers, 2010). Research on metacognitive processes including more than one school subject is underrepresented as of yet. Including two different school topics allows for comparing and therefore exploring the generalisation of the findings. Therefore, the present research is on both mathematics and science test performance and students' performance estimations in these tests occurring over a school year.

In addition, also investigated in the study was the influence of test difficulty on the variability of monitoring accuracy. Basic research provides ample evidence that monitoring accuracy highly depends on the difficulty of a particular task or item. Both, metacognitive discrimination (distinguishing between correct and incorrect answers) and calibration (pre- and post-dictions about one's own performance) being poorer for difficult as opposed to easy tasks or items (Allwood et al., 2005; Howie & Roebers, 2007; Olsson, 2000). In most of the above-mentioned studies in the domain of mathematics, only one single test has been used providing a snap-shot of student's monitoring skills. However, if task or test difficulty substantially impacts monitoring accuracy and efficiency, student's monitoring accuracy for mathematics problems should also vary considerably across different test situations. This is because efficient metacognitive processes allow a student to distinguish between easier and harder tasks, to identify problems that require more skills and effort to complete, to discriminate between real and apparent problems during task mastery, to identify the steps necessary to solve a problem, and to focus evaluation and revising on the more difficult tasks (Garrett et al., 2006). In the present paper therefore, students' performance estimations as measures of monitoring accuracy given for every mathematics and science test that occurred over the course of their 5th grade will be explored and related to task difficulty.

As a conclusion the present paper aims at extending the existing literature on meta-cognitive processes in the context of school achievement with respect to three aspects: First, student's performance estimations and the degree of deviation from their actual performance (i.e., miscalibration) will be addressed for all mathematics tests that our sample of 5th graders took over the course of the school year. This allows insights into the variability of individuals' calibration as a function of objective or subjective test difficulty. Because achievement level is known to consistently influence metacognitive monitoring accuracy, analyses will be conducted and presented for high, medium, and low achieving students, separately. Second, all analyses will be presented for mathematics and science similarly in order to investigate generalizability of findings stemming from the mathematics domain for other school domains.

Method

Participants

In total, data of N = 25 children were provided by the teacher. Data of N = 6 children were excluded because either not enough test and performance estimations were available, or because performance or estimations deviated more than 2 standard deviations from the mean. Therefore, data of N = 19 children (N = 9 female) will be included in the analyses reported below. Mean age at beginning of the school year was 11 years (SD = 9 months; range 114 – 150 months) and all children participating in the study were sampled from a 5th grade class in a private school in the region of Bern, Switzerland. The school's principal allowed the use of the data for the current purposes, provided, that no inferences on behalf of the involved school, teacher or children are possible.

Material and Procedure

The test scores presented in this paper stem from actual school exams students completed over the course of an entire school year (5th grade), which were designed and conducted by the teacher himself. The exams reported include two school subjects: mathematics and science. The exams covered the normal 5th grade curriculum encompassing analyzing, depicting and solving mathematical problems, rounding, fractions, multiplication, and division in mathematics, and various fields of knowledge in science (e.g. birds, map reading and orientation). Due to the differences in complexity and demands of the contents within the school subjects, the numbers of exams varied considerably. In total, data of 10 exams in mathematics, and 5 exams in science was provided. The contents of the mathematics exams tended to be less extensive and the maximum score tended to be lower ($M_{maxmathematics}$ = 20.3, Range: 9 – 43 credits) compared to the science exams ($M_{maxscience}$ = 35.4, Range: 21 – 52 credits). Importantly, in many tests with low maximum scores, partial solutions were possible, resulting in a finer grained scoring than may be expected at first sight.

The headline of each exam comprised four boxes including information about the scoring. The first two boxes indicated (1) the maximum score and (2) the score necessary to pass the exam. The third box ("performance estimation") was empty and the students were required to fill in their estimation of the credits they anticipated to achieve in the upcoming exam. After marking the exams, the teacher entered the number of credits achieved by the students in the fourth box ("performance"). In all exams, the credits that could be maximally achieved with a correct answer and/or an appropriate demonstration of the solution process were specified next to each question/each task.

The instruction for the estimation was limited to the request of "not to forget to fill in the estimation of the anticipated score" achieved in the exam. The students could fill in their estimation at any time during test-taking. They could for example rate the solution of each task separately and add these estimations up for a total estimation. In general however, they waited with their total estimation up to the very end of the test session and filled it in after a reminder by the teacher. Thus the estimations are post-dictions of the actual test-performance and at the same time predictions of the teacher-rated scores as in other studies assessing metacognitive monitoring (e.g. Desoete & Roeyers, 2006; Krebs & Roebers, 2011, Renner & Renner, 2001).

The teacher made educational use of the estimation values if there was an obvious difference between the credits estimated and achieved. The intervention consisted in general of calling the child's attention to obvious and marked differences between performance estimation and actual performance and in discussing possible reasons.

All exams were completed in the usual school environment during the obligatory school lessons, meaning that all children were tested at the same time under the same conditions. The tests were identical for all students. The time of completion differed between exams and children who finished the test earlier stayed in the classroom occupying themselves with some other school assignment. After scoring and marking the exams, the teacher handed them to the children so that they were able to check for mistakes and make corrections, compare the achieved and estimated credits, and provide their parents' signature.

Statistical Analysis

For the analyses reported below, test scores and performance estimations were converted into percentages of the maximum number of credits in that particular test, to account for the different numbers of to-be-obtained credits across tests. For each test and for each individual, the degree of deviation between performance estimation and actual test performance was quantified by the absolute value of the difference between these two indices (in percent), mirroring a score of "miscalibration".

Results

The available data set of test scores and performance estimations spanning over an entire school year allows assessing variability of the students' performance estimations across different test occasions. As outlined, we assume that performance estimations will vary (a) as a function of a student's achievement level within that school subject and (b) as a function of subjective and/or objective test difficulty. For the following analyses, miscalibration scores will be used exclusively because they allow calculating means across students and tests. Initial analyses revealed a significant sex difference for the mean performance estimation in mathematics, with more positive estimations in boys (M = 87.9%, SD = 7.6%) than girls (M = 79.3%, SD = 9.9%), t(17) = 2.14, p= .05. This result seemed to be fuelled by marginally significant higher mean test scores in mathematics among the boys (M = 89.3%, SD = 7.1%) compared to the girls (girls: M = 82.7%, SD = 7.7%), t(18) = 1.99, p = .06. There were no other (marginally) significant effects, and as the focus of the analyses is not on sex differences, boys' and girls' data were collapsed.

In addition to significance values, eta²- values are reported as estimators of effect sizes. For both school subjects (i.e., mathematics and science) the students were subdivided into three achievement groups. In both school topics, each of the three achievement level groups consisted of 6 to 7 students.

Mathematics

First, the question of how miscalibration varied across the school year and the different tests was addressed. Both the mean number of credits achieved and students' performance estimations are depicted in Figure 1. Inspection of that figure suggests that the graph of the estimations is close to that of the achieved credits. The variations seem somewhat smaller for the estimations, thus differences in test difficulty do not translate directly into estimations from one test to the other. A series of t-tests was conducted in order to test for systematic differences between achieved and estimated credits. Results revealed that for test numbers 5, 8 and 10 mean performance estimations differed significantly from mean test performance, $t_{test\ 5}(17) = 3.247$, $p < .01$, $t_{test\ 8}(16) = 2.20$, $p < .05$, $t_{test\ 10}(17) = 2.52$, $p = .02$; the difference for test number 3 was only marginally significant, $t_{test\ 3}(16) = -1.822$, p = .09, and all remaining differences were not significant ($p > .10$).

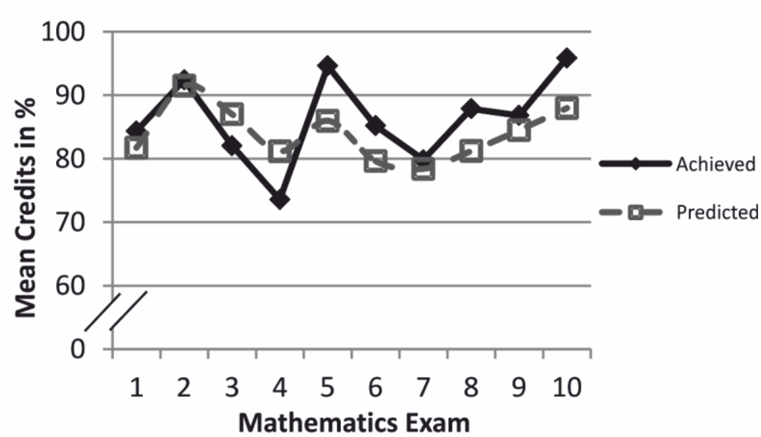

***Figure 1*. Achieved and estimated scores in % for the ten mathematics exams**

In Figure 2, the variability of miscalibration in mathematics is shown, with each line representing one of the three different achievement groups. When descriptively comparing the groups, it is apparent that miscalibration varies considerably across tests. In some tests more pronounced differences in miscalibration between the three achievement groups are evident, whereas in other tests, the amount of miscalibration seems to be comparable between groups. Overall, however, inspection of Figure 1a nevertheless reveals that miscalibration scores tended to be higher in the low achieving mathematics group.

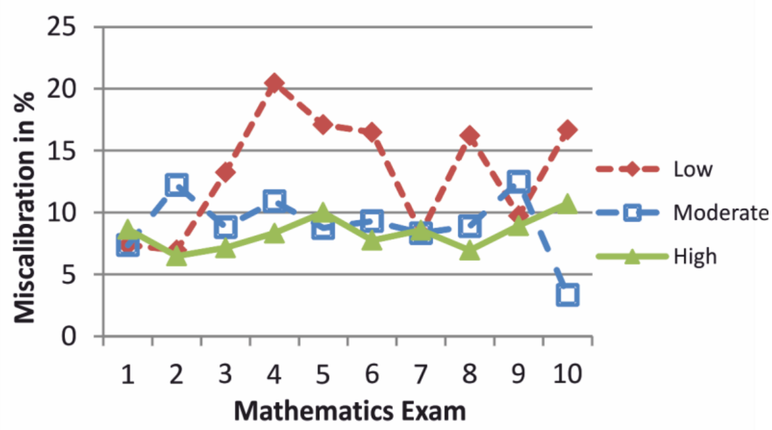

***Figure 2*. Miscalibration scores in % for the ten mathematics exams as a function of achievement group**

Figure 3 presents the miscalibration scores averaged over the entire school year as a function of achievement level grouping. As can be seen, the three groups differ substantially from one another with the low achievement group demonstrating the most pronounced miscalibration. A one-way ANOVA was conducted to test for these differences in the degree of miscalibration for mathematics exams between the three achievement groups. Miscalibration was found to differ significantly between groups, $F(2, 16) = 4.31$, $p = .03$, $\eta_p^2 = .35$, confirming the above-mentioned descriptive observations for miscalibration scores for each test separately. Tukey post-hoc comparisons between the three groups indicated that the high achieving group was significantly less miscalibrated than the low performance group, $p = .03$. The differences between the moderate achievement group and the other two groups were not statistically reliable at $p < .05$.

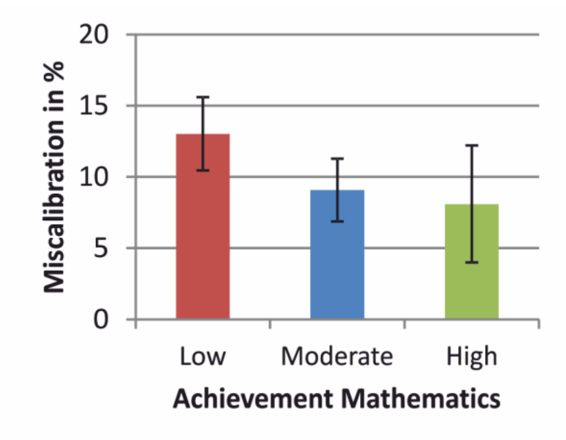

***Figure 3*. Miscalibration Scores in % as a Function of the three achievement levels in the domain of mathematics**

Science

Analogous to mathematics, Figure 4 presents the number of credits achieved and students' performance estimations. Again, the courses of the two lines across the 5 tests appear to be very close. Thus, on average the students seem to be well capable of estimating their test performance. It was only in the first test, that difference between mean performance estimation and mean actual performance reached significance, $t(14) = 2.20$, $p < .05$.

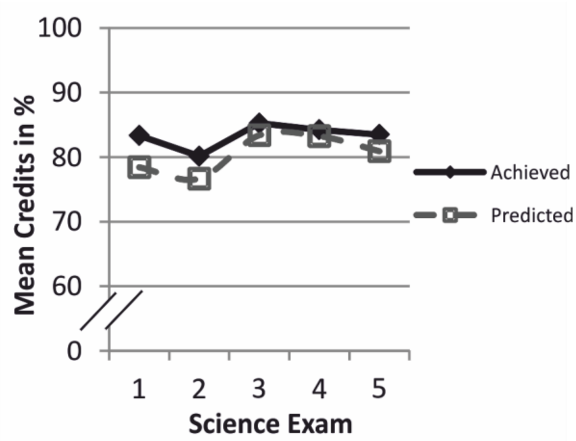

***Figure 4*. Achieved and estimated scores in % for the five science exams**

The miscalibration for each of the conducted tests in the domain of science over the course of students' 5th grade is presented in Figure 5. Again, the degree of miscalibration was found to vary considerably across exams. At the same time, however, as explored in the preceding paragraphs on a descriptive level, there seems to be a stable difference in the magnitude of miscalibration between the three achievement groups, the high achieving group showing the smallest miscalibration in most of the tests.

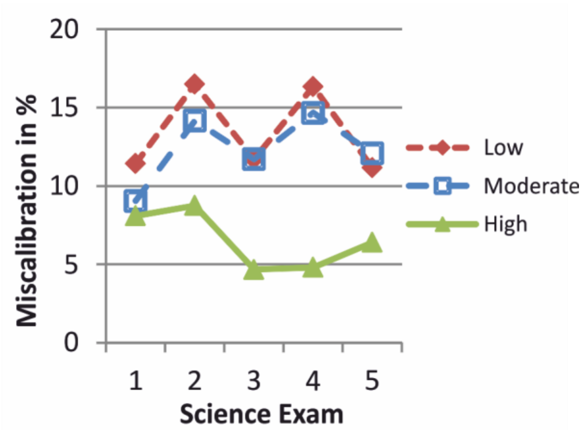

***Figure 5*. Miscalibration scores in % for the five science exams as a function of achievement group**

Figure 6 presents students' miscalibration in science, averaged across the entire school year's tests as a function of achievement level in that subject. As can be seen in the figure, miscalibration was lowest in the high achieving science group. When systematically comparing the achievement groups' miscalibration scores for science exams, a one-way ANOVA indicated a significant main effect of achievement group, $F(2, 15) = 3.71$, $p > .05$, $\eta_p^2 = .33$. However, when comparing two groups each with Tukey post-hoc comparisons, there was only a marginally significant difference between the low and high achieving groups, $p = .06$, but no significant difference between the moderate achievers and the two other groups.

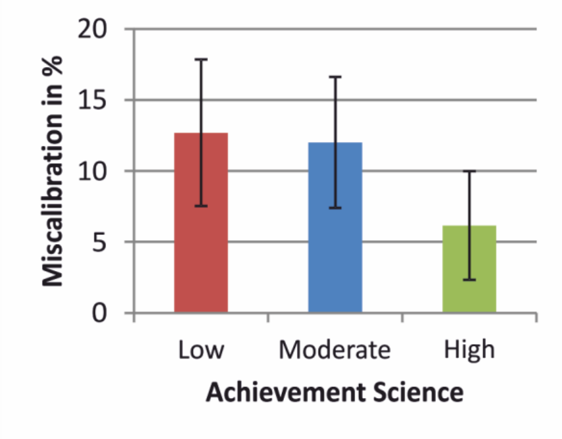

***Figure 6*. Miscalibration Scores in % as a function of the three achievement levels in the domain of science**

Discussion

The present study provided some first and explorative insights into metacognitive monitoring in a naturalistic context, where performance estimations were assessed in curricula based tests constructed and instructed by the teacher. Although the source of the data underlying the present analyses is somewhat special as this study was never specifically planned but accumulated naturally by one specific teacher's didactical methods. The study was only analyzed post-hoc, the results obtained nevertheless provide interesting aspects of children's metacognitive monitoring skills in two core school subjects. In fact, one might argue that the lack of experimental control of many (naturally occurring) influences constitutes an advantage of the present approach. Generalising ability does not need to be questioned as primary school students monitoring accuracy was explored under ecologically valid conditions. The students in our study were tested under conditions, in which the content of the tests was meaningful to them. At the same time they tried to achieve a good performance as the results had an impact on their grades. The tests were not only highly relevant, but also administrated by the usual teacher. Therefore the situation diverged markedly from testing with new and sometimes unknown material put forward by researchers unfamiliar to the students. It is therefore likely, that the students' monitoring performance was not impeded or enhanced by especially low or high interest in test content, external motivation (praise, small gifts brought forward by researchers), social desirability or other factors that might come along with more controlled assessments and are typical for experimental studies.

With this information in mind, it seems especially noteworthy that the well-documented influence of achievement level on monitoring accuracy was confirmed with the present data, not only for mathematics but also for science (Desoete & Roeyers, 2006; Garrett et al., 2006; Thiede et al., 2003). That is, children who perform better in a school subject are also better able to estimate their performance. Of course and unfortunately, the interesting question of the exact nature of the relationship between performance and monitoring still remains open. It is possible that better monitoring skills lead to superior academic performance, especially in test situations. However, assuming the reversed direction, high achievement enabling more precise monitoring is also sensible. Most likely, a reciprocal relationship exists between these two constructs and future longitudinal studies should continue its pursuit to clarify this important issue.

As to 5th graders calibration, the overall pattern that emerged from the analyses presented here is that students were relatively well able to estimate their performance, that is, the average performance estimation did not deviate more than 15% from the actual performance even in the low achieving group. Based on a consistent body of evidence, we had expected that 11- to 12- year olds can estimate their performance fairly well, given the pronounced developmental increases in monitoring accuracy between the ages of 7 to 10 year (Schneider, 2010). Strikingly, students' performance across the different tests spread over the school year and covering different contents varied considerably. Miscalibration, nevertheless, appeared to be relatively unaffected by averaged test performance. The present findings can thus be interpreted as indicating that performance estimations' accuracy is less influenced by test difficulty compared to actual performance. This demonstrates that students took the perceived test difficulty into account when making their estimations. In addition, when estimating their performances, students may have capitalized on the teacher feedback from earlier tests or some inner, more stable confidence in their test solving abilities (self-confidence, Kleitman &

Stankov, 2007). Even though such influences did not increase monitoring accuracy across a school year, they may help to compensate for performance differences due to variations in test difficulty.

To complicate matters, it is likely that high achieving students perceived the more difficult tests as relatively less difficult compared to low achieving students and this confound between objective and subjective difficulty makes interpretations difficult in general. Unfortunately, we were not able to systematically manipulate the objective and subjective degree of test difficulty on performance estimations for investigating its influence, given the nature and quality of the underlying data. Our explorative approach, however, may motivate a larger and more controlled investigation of performance estimations as a function of test difficulty.

Both the significant impact of the achievement level on monitoring accuracy and the relatively independence of monitoring accuracy from test difficulty was found for mathematics and science. Intriguingly, the groups were created based on students' achievement in that particular domain, implying that different students were assigned to the three achievement groups for mathematics and science, respectively. These similarities can thus be interpreted as indicating considerable generalisation of conclusions for this form of metacognitive monitoring from one school subject onto another.

Obviously, there are some methodological problems attached to the present data and consequently to the analyses performed. Mostly due to the limited number of participants, and the relatively small variations in test performance, interesting research questions could only be explored but not systematically addressed. Despite these obvious limitations, the present data provide some interesting insights into monitoring accuracy, test difficulty, and the role of achievement level. And finally, the assumption that more realistic estimations enable students to adjust their behavior during the test such as investing more cognitive resources in difficult tasks compared to easy tasks, better distinguishing between actual and assumed difficulties, and focusing more efficiently on error detection and correction may set a good starting point future directions for research. Against this background, intervention studies aiming to increase students' accuracy of performance estimation seem promising for the educational practice (Coutinho, 2007; Desoete, 2009; Dunlosky & Lipko, 2007).

References

González-Alexander, J. M., Carr, M., & Schwanenflugel, P. J. (1995). Development of metacognition in gifted children: Directions for future research. *Developmental Review, 15,* 1-37. doi: http://dx.doi.org/10.1006/drev.1995.1001

Allwood, C. M., Jonsson, A.-C., & Granhag, P. A. (2005). The effects of source and type of feedback on child witnesses' metamemory accuracy. *Applied Cognitive Psychology, 19,* 331-344. doi: http://dx.doi.org/10.1002/acp.1071

Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K., & Weiss, M. (Eds.). (2001). PISA 2000. Basiskompetenzen von Schuelerinnen und Schuelern im internationalen Vergleich. PISA 2000. *An international comparison of basic competences in school students*: Opladen: Leske + Budrich.

Urie Bronfenbrenner. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, MA: Harvard University Press.

Carr, M., Kurtz, B. E., Schneider, W., Turner, L. A., & Borkowski, J. G. (1989). Strategy acquisition and transfer among American and German children: environmental influences on metacognitive development. *Developmental Psychology, 25,* 765-771.

Cohors-Fresenborg, E., Kramer, S., Pundsack, F., Sjuts, J., & Sommer, N. (2010). The role of metacognitive monitoring in explaining differences in mathematics achievement. *ZDM, 42,* 231-244. doi: 10.1007/s11858-010-0237-x

Coutinho, S._A. (2007).The relationship between goals, metacognition, and academic success. *Educate Journal, 7,* 39-47. Retrieved August 12, 2011, from http:iArww.educatejoumal.org/index.php? journal:educate&page=article&op= viewFile&path%5B%5F I I 6 &path%58%5D= I 34

Desoete, A. (2009). Metacognitive prediction and evaluation skills and mathematical learning in third-grade students. *Educational Research and Evaluation, 15,* 435-446. doi: http://dx.doi.org/10.1080/13803610903444485

Desoete, A., & Roeyers, H. (2006). Metacognitive macroevaluations in mathematical problem solving. *Learning and Instruction, 16,* 12-25. doi: http://dx.doi.org/10.1016/j.learninstruc.2005.12.003

Desoete, A., Roeyers, H., & De Clercq, A. (2003). Can offline metacognition enhance mathematical problem solving? *Journal of Educational Psychology, 95,* 188-200.

Dunlosky, J., & Lipko, A. R. (2007). Metacomprehension: A brief history and how to improve its accuracy. *Current Directions in Psychological Science, 16*, 228-232. doi: http://dx.doi.org/10.1111/j.1467-8721.2007.00509.x

Flavell, J. H. & Wellman, H. M. (1977). Metamemory. In R. V. Kail & J. W. Hagin (Eds.), *Perspectives on the development of memory and cognition* (pp. 3-33). Hillsdale, NJ: Lawrence Erlbaum Associates.

Garrett, A. J., Mazzocco, M. M., & Baker, L. (2006). Development of the metacognitive skills of prediction and evaluation in children with or without math disability. *Learning Disabilities Research & Practice, 21*, 77-88. doi: http://dx.doi.org/10.1111/j.1540-5826.2006.00208.x

Hacker, D. J., Dunlosky, J., & Graesser, A. C. (2009). *Handbook of metacognition in education*. New York, NY: Routledge/Taylor & Francis Group.

Howie, P., & Roebers, C. M. (2007). Developmental progression in the confidence-accuracy relationship in event recall: Insights provided by a calibration perspective. *Applied Cognitive Psychology*, *21*, 871-893. doi: http://dx.doi.org/10.1002/acp.1302

Krebs, S. S., & Roebers, C. M. (2010). Children's strategic regulation, metacognitive monitoring, and control processes during test taking. *British Journal of Educational Psychology, 80,* 325-340. doi: http://dx.doi.org/10.1348/000709910X485719

Krebs, S., & Roebers, C. (2011). The impact of retrieval processes, age, general achievement level, and test scoring scheme for children's metacognitive monitoring and controlling. *Metacognition and Learning,* 1-16. doi: 10.1007/s11409-011-9079-3

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality & Social Psychology, 77*, 1121-1134.

Körkel, J., & Schneider, W. (1991). Domain-specific versus metacognitive knowledge effects on text recall and comprehension. In M. Carretero, M. L. Pope, P. R.-J. Simons & J. I. Pozo (Eds.), *Learning and instruction: European research in an international context, Vol. 3* (pp. 311-323). Elmsford, NY: Pergamon Press.

Legg, A. M., & Locker, L., Jr. (2009). Math performance and its relationship to math anxiety and metacognition. *North American Journal of Psychology, 11*, 471-486.

Lucangeli, D., Coi, G., & Bosco, P. (1997). Metacognitive awareness in good and poor math problem solvers. *Learning Disabilities Research & Practice, 12,* 209-212.

Mevarech, Z., Terkieltaub, S., Vinberger, T., & Nevet, V. (2010). The effects of meta-cognitive instruction on third and sixth graders solving word problems. *ZDM, 42*, 195-203. doi: 10.1007/s11858-010-0244-y

Montague, M., & Bos, C. S. (1990). Cognitive and metacognitive characteristics of eighth grade students' mathematical problem solving. *Learning and Individual Differences, 2*, 371-388. doi: http://dx.doi.org/10.1016/1041-6080%2890%2990012-6

Nelson, T.O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *The Psychology of Learning and Motivation, 26*, 125-173.

Kleitman, S., & Stankov, L. (2007). Self-confidence and metacognitive processes. *Learning and Individual Differences, 17*, 161–73.

Kuhn, D., & Dean, D., Jr. (2004). Connecting scientific reasoning and causal inference. *Journal of Cognition and Development, 5*, 261-288. doi: http://dx.doi.org/10.1207/s15327647jcd0502_5

Olsson, N. (2000). A comparison of correlation, calibration, and diagnosticity as measures of the confidence-accuracy relationship in witness identification. *Journal of Applied Psychology, 85*, 504-511.

Pennequin, V., Sorel, O., Nanty, I., & Fontaine, R. (2010). Metacognition and low achievement in mathematics: The effect of training in the use of metacognitive skills to solve mathematical word problems. *Thinking & Reasoning, 16*, 198-220. doi: http://dx.doi.org/10.1080/13546783.2010.509052

Renner, C. H., & Renner, M. J. (2001). But I thought I knew that: Using confidence estimation as a debiasing technique to improve classroom performance. *Applied Cognitive Psychology, 15*, 23-32.

Roebers, C., Schmid, C., & Roderer, T. (2009). Metacognitive monitoring and control processes involved in primary school children's test performance. *British Journal of Educational Psychology, 79*, 749-767. doi: http://dx.doi.org/10.1348/978185409X429842

Schneider, W. (2010). Metacognition and memory development in childhood and adolescence. In H. S. Waters & W. Schneider (Eds.), *Metacognition, strategy use, and instruction* (pp. 54-81). New York: Guilford Press.

Schneider, W., & Artelt, C. (2010). *Metacognition and mathematics education. ZDM*, 42, 149-161. doi: 10.1007/s11858-010-0240-2

Schneider, W., & Pressley, M. (1997). *Memory development between two and twenty (2nd ed.).* Mahwah, NJ: Erlbaum.

Schneider, W., Schlagmüller, M., & Visé, M. (1998). The impact of metamemory and domain-specific knowledge on memory performance. *European Journal of Psychology of Education, 13*, 91-103.

Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning: A project of the National Council of Teachers of Mathematics* (pp. 334-370). New York, NY: Macmillan Publishing Co, Inc; England.

Slife, B. D., Weiss, J., & Bell, T. (1985). Separability of Metacognition and Cognition: Problem Solving in Learning Disabled and Regular Students. *Journal of Educational Psychology, 77*, 437-445.

Thiede, K. W., Anderson, M. C., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, 95*, 66-73. doi: http://dx.doi.org/10.1037/0022-0663.95.1.66

Van der Stel, M., & Veenman, M. V. (2008). Relation between intellectual ability and metacognitive skillfulness as predictors of learning performance of young students performing tasks in different domains. *Learning and Individual Differences, 18*, 128-134. doi: http://dx.doi.org/10.1016/j.lindif.2007.08.003

Van der Stel, M., Veenman, M., Deelen, K., & Haenen, J. (2010). The increasing role of metacognitive skills in math: a cross-sectional study from a developmental perspective. *ZDM, 42*, 219-229. doi: 10.1007/s11858-009-0224-2

Veenman, M. V., & Spaans, M. A. (2005). Relation between intellectual and metacognitive skills: Age and task differences. *Learning and Individual Differences, 15*, 159-176. doi: http://dx.doi.org/10.1016/j.lindif.2004.12.001

Williams, W. M., Blythe, T., White, N., Li, J., Gardner, H., & Sternberg, R. J. (2002). Practical intelligence for school: Developing metacognitive sources of achievement in adolescence. *Developmental Review, 22*, 162-210. doi: http://dx.doi.org/10.1006/drev.2002.0544

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Mahwah, NJ: Erlbaum.

Zimmerman, B. J., & Moylan, A. R. (2009). Self-regulation: Where metacognition and motivation intersect. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 299-315). New York, NY: Routledge/Taylor & Francis Group.