

Factors involved in making post-performance judgments in mathematics problem-solving

Trinidad García Fernández¹, Evelyn Kroesbergen², Celestino Rodríguez Pérez¹, Paloma González-Castro¹
and Julio Antonio Gonzalez-Pienda¹

¹ University of Oviedo and ² Utrecht University

Abstract

Background: This study examines the impact of executive functions, affective-motivational variables related to mathematics, mathematics achievement and task characteristics on fifth and sixth graders' calibration accuracy after completing two mathematical problems. **Method:** A sample of 188 students took part in the study. They were divided into two groups as function of their judgment accuracy after completing the two tasks (accurate= 79, inaccurate= 109). Differences between these groups were examined. The discriminative value of these variables to predict group membership was analyzed, as well as the effect of age, gender, and grade level. **Results:** The results indicated that accurate students showed better levels of executive functioning, and more positive feelings, beliefs, and motivation related to mathematics. They also spent more time on the tasks. Mathematics achievement, perceived usefulness of mathematics, and time spent on Task 1 significantly predicted group membership, classifying 71.3% of the sample correctly. **Conclusions:** These results support the relationship between academic achievement and calibration accuracy, suggesting the need to consider a wide range of factors when explaining performance judgments.

Keywords: achievement, calibration, elementary school, executive functions, mathematical problems.

Resumen

Factores implicados en la evaluación post-ejecución de la resolución de problemas matemáticos. Antecedentes: este estudio analiza la influencia del funcionamiento ejecutivo, componentes afectivo-motivacionales relacionados con las matemáticas, rendimiento académico y características de la tarea, en las habilidades de calibración de estudiantes de quinto y sexto curso tras la realización de dos problemas matemáticos. **Método:** participaron 188 estudiantes, divididos en dos grupos según la precisión de sus juicios de rendimiento tras la tarea (precisos = 79; imprecisos = 109). Se analizaron las diferencias entre los grupos en estas variables, y su capacidad discriminativa prediciendo la pertenencia al grupo. Los efectos de la edad, género y nivel educativo fueron analizados. **Resultados:** los resultados indicaron que los estudiantes precisos en sus juicios presentaron mejores niveles de funcionamiento ejecutivo, así como sentimientos, creencias y motivaciones más positivas hacia las matemáticas, y mejor rendimiento académico en la asignatura. También emplearon más tiempo realizando las tareas. El rendimiento académico, la utilidad percibida de las matemáticas, y el tiempo empleado en la primera tarea predijeron significativamente la pertenencia de los estudiantes al grupo, clasificando correctamente al 71.3%. **Conclusiones:** estos resultados apoyan la relación entre rendimiento académico y calibración, sugiriendo la necesidad de considerar un amplio rango de variables cuando estos juicios quieren ser explicados.

Palabras clave: calibración, escuela primaria, funciones ejecutivas, problemas matemáticos, rendimiento.

Mathematical problem-solving is a complex cognitive activity that involves multiple processes. Success in these tasks relies on the integration of cognitive, self-regulatory, and motivational components (Boonen, Van der Schoot, Van Wesel, De Vries, & Jolles, 2013; Montague, Enders, & Dietz, 2011). In order to perform efficiently, students need to regulate and monitor their learning processes. An important aspect of monitoring is formed by the judgments that students make about their performance and the extent to which these judgments match their actual performance.

The degree of correspondence between one's perception of performance and actual performance is referred to as *calibration* (Hacker, Bol, & Keener, 2008). These judgments can be made before (predictions), or after (postdictions) task completion. Postdictions tend to be more accurate and reliable than predictions, and inform about monitoring processes during the task (Bol, Hacker, Walck, & Nunnery, 2011; Hacker et al., 2008; Sheldrake, Mujtaba, & Reiss, 2014). The present study investigates which factors contribute to making accurate post-performance judgments.

As calibration informs about the status of one's knowledge or strategies at a cognitive level, it is involved in the regulation of further effort and strategy use (Van Loon, de Bruin, Van Gog, & Van Merriënboer, 2013). The relationship between calibration and achievement has been demonstrated in different subject areas, particularly in mathematics and problem-solving (Jacobse & Harskamp 2012; Özsoy 2012; Rinne & Mazzocco, 2014).

Specifically, calibration accuracy may explain 16-36% of the variance in mathematics achievement (Jacobse & Harskamp, 2012), while highly calibrated students tend to perform more successfully on problem-solving tasks because they exercise a more efficient control over their problem-solving processes (Özsoy, 2012). This is coherent with the characterization of calibration as an important metacognitive monitoring process (Stolp & Zabrocky, 2009).

However, students tend to be inaccurate in their judgments, with a tendency towards over-confidence. This tendency has been confirmed in different educational stages and adult samples, suggesting that performance judgments (and biases) may be stable overtime (Bouffard & Narciss, 2011; Finn & Metcalfe, 2014; Hacker et al., 2008). Although most research has focused on examining the effects of over-confidence, under-confidence has also been demonstrated to potentially threaten learning (Dupeyrat, Escribe, Huet, & Régner, 2011; Sheldrake et al., 2014). Keeping in mind that this pattern of confidence bias may become stable, it is necessary to determine what influences student performance judgments.

One of the more consistent findings in this sense concerns the relationship between academic achievement and judgment accuracy. Specifically, high-achievers have been shown to be more calibrated, and somewhat under-confident, in comparison to low-achieving students (Bol et al., 2011; Hacker et al., 2008; Özsoy, 2012). In this same direction, task characteristics (mainly difficulty) have been widely studied, leading to mixed results. Howie and Roebbers (2007) reported poor calibration only for difficult questions, while other studies suggest that students would be more over-confident when the material is difficult and under-confident when it is easy (Hacker et al., 2008).

In the context of elementary school, Boekaerts and Rozendaal (2010) examined the effects of gender, type of mathematical problem (application versus computation), and time of measurement (predictions versus postdictions) on fifth graders' calibration accuracy. Results indicated that boys were more confident than girls judging their ability to successfully solve the problems. However, they over-estimated their performance in comparison to girls, revealing poor calibration. They also found that students performed worse on application problems, and were less calibrated in these problems, and that, surprisingly, performance judgments were more accurate before task completion (i.e., predictions).

Calibration accuracy has also been shown to be influenced by affective-motivational variables. Individual differences in self-efficacy beliefs, goal orientation, and susceptibility to social influences may account for differences in judgment accuracy, while the opposite relationship has also been demonstrated (Narciss, Koerndle, & Dresel, 2011; Sheldrake et al., 2014). Specifically, Narciss and colleagues found that fifth-grade students who made accurate judgments had greater increases in satisfaction with their performance compared to students who over-estimated their performance. Calibration accuracy may also influence students' academic itinerary in further educational stages (Sheldrake et al., 2014). In this sense, highly-calibrated students show higher levels of enjoyment and interest in mathematics than their inaccurate peers.

However, although there is a good deal of evidence suggesting the impact of the mentioned factors on calibration, less consideration has been given to domain-general factors that may also influence students' ability to make accurate performance judgments, such as Executive Functions (EF). Defined as "*the ability to maintain an appropriate problem-solving set for*

attainment of a future goal" (Welsh & Pennington, 1988, p. 201), EF refers to processes such as planning, goal setting, response inhibition, impulse control, attention, self-monitoring, and cognitive flexibility. These components have been shown to serve as powerful predictors of school readiness and achievement, specifically, of mathematics (Bull & Lee, 2014; Kolkman, Hoijtink, Kroesbergen, & Leseman, 2013). To date however, research on the relationship between calibration and EF is very scarce. While Rinne and Mazzocco (2014) suggest that calibration in mathematics would be linked to working memory, as well as other EF such as response maintenance, from a theoretical perspective, the components of inhibition (or impulse control), working memory, attention, planning, and cognitive flexibility may be related to the ability to evaluate one's performance (García, González-Pienda, Rodríguez, Álvarez-García, & Álvarez, 2014; Garner, 2009). Specifically, poor attention and inhibition may lead to paying attention to irrelevant information, making it difficult to maintain a goal in working memory. Without a clear goal in mind, one would be less able to switch strategies when necessary, and even less able to accurately evaluate one's performance. Roebbers, Cimeli, Röthlisberger and Neuenschwander (2012) argued that EF and children's awareness of the discrepancy between task demands and their own performance would progress in parallel during elementary school years. This suggests the advisability of properly examining the links between EF and calibration from a developmental perspective in childhood.

The present study

This study analyzed the impact of three sets of variables on students' performance judgment accuracy after completing two mathematical problems. A sample of fifth- and sixth-grade students took part in the study, divided into two groups as function of their post-performance judgment accuracy (accurate/inaccurate). The studied variables were:

- 1) Domain-general variables (EF): impulsivity control, activity regulation, emotional control, focus, concentration, planning, working memory, organization, and flexibility.
- 2) Domain-specific variables: affective-motivational components concerning mathematics (perceived usefulness of mathematics, self-efficacy beliefs, intrinsic motivation, anxiety, and enjoyment), and mathematics achievement.
- 3) Task-related variables: perceived difficulty and time spent on the tasks.

Two main objectives were established:

- 1) To analyze differences in domain-general, domain-specific or task-related variables between the groups with different calibration accuracy.
- 2) To delimit the predictive value of these variables on student's group membership (Accurate or Inaccurate group).

Method

Participants

A sample of 359 fifth and sixth grade Spanish students took part in the study, of which 188 were selected for further data-

analysis (ages 10 to 13, $M = 11.05$, $SD = .776$). Ninety-two students (48.9%) were female, and 73 (38.8%) attended fifth grade.

The sample was divided into two groups based on the correspondence between post-performance judgments and actual performance, both expressed in dichotomous terms (success = 1, failure = 0). Students were assigned to the accurate group when post-performance judgments and actual performance matched in both tasks, and to the inaccurate group when they mismatched. Students who showed a tendency towards over-confidence: 107 (56.91%) students in Task 1 and 102 (54.25%) in Task 2, judged their performance as correct while, in fact, it was incorrect. However, only 2 and 7 students respectively (less than 1%) showed the opposite pattern. Students who were accurate in only one task were excluded from the analyses to avoid haphazard responses ($n = 171$).

Table 1 shows groups characteristics. No statistical significant differences between groups in age, grade level and gender distribution were found.

Sample selection was made through accessibility procedures. Students volunteered for the study, and presented informed consent from their parents. Children with a diagnosis, or severe learning disabilities, were excluded from the analyses.

Instruments

Domain-general variables

The Executive Functioning Scale for Families (EFS-F; García, Álvarez-García, Cueli, González-Castro, & Álvarez, 2013) was used to evaluate Impulsivity Control, Activity Regulation, Emotional Control, Concentration, Focus, Planning, Organization, Working Memory, and Flexibility. It consists of 27 Likert-type items. Parents report the frequency with which children show different behaviors related to EF deficits (from 1 = *Never*, to 5 = *Always*). High scores indicate deficit. Reliability was high for the scale ($\alpha = .94$) and moderate for its factors (.67 to .74).

Domain-specific variables

The Inventory of Attitudes toward Mathematics- IAM (González-Pienda et al., 2012) was used to evaluate Perceived usefulness, Self-efficacy beliefs, Intrinsic motivation, Anxiety, and Enjoyment (20 items). Items are rated on a 5-point Likert scale (1 = *strongly disagree*, 5 = *strongly agree*). Perceived usefulness and Enjoyment are worded negatively (high scores indicate low perception of usefulness and low enjoyment, respectively). Cronbach's alphas ranged between .84 and .87.

Achievement in Mathematics was based on student final academic grades in the subject (from 0 to 10).

Tasks characteristics

Evidence from perceived difficulty was obtained after task completion. Students scored the degree of difficulty of each task (1 = *very easy*, 5 = *very difficult*). Time spent on the task (expressed in seconds) was registered for each task. Two tasks were used to guarantee that students performed without time pressure. No specific time restrictions were made.

Calibration variables

Post-performance judgments were obtained by asking students to evaluate whether they had solved the problem successfully, once they completed each task. Actual Performance was established based on the correctness of students' answers to the problems. Both variables were expressed in dichotomous terms (success = 1, failure = 0) given the characteristics of the problems, which required a unique response.

Procedure

The study was conducted according to The Helsinki Declaration of the World Medical Association (Williams, 2008).

The evaluation was collectively administered during a regular class, using computers. Students and parents had previously completed the questionnaires (IAM and EFS-F, respectively). Mathematical problems were given on paper and also displayed on the computer screen, through a module enabled on Moodle platform. After solving each problem, students evaluated the degree to which they had perceived the task as easy or difficult, and made a performance judgment. The questions "how easy or difficult was the problem for you?", and "do you think you have solved the problem correctly?" were presented on the computer screen, after students indicated they had finished the problems, by clicking on a button set up for this purpose. This module allowed tracking when students started and finished each task, providing a measure of performance time.

Actual performance was established based on students' written answers. Once evidence from post-performance judgments and actual performance was obtained, students were assigned to the groups (Accurate vs. Inaccurate).

The problems were taken from the book "Problem solving and comprehension" (Whimbey & Lochhead, 1999). They were based on daily and practical situations. Teachers evaluated their adequacy prior to the evaluation.

Data analysis

(M)ANCOVA's were conducted to analyze group differences in the dependent variables. Variables distribution was examined, paying special attention to skewness and kurtosis (Table 2). Effect sizes were analyzed, using Cohen's (1988) criterion. A small association was defined as $\eta p^2 = .010$ ($d = .20$), a medium association as $\eta p^2 = .059$ ($d = .50$), and a large association as $\eta p^2 = .138$ ($d = .80$). A discriminant analysis was carried out to determine the significance of these variables predicting group membership, following a stepwise method. As this analysis requires equal

Table 1
Sample characteristics

	Group 1	Group 2
N	79/42.2%	109/57.8%
Age (M/SD)	11.04/0.808	11.06/0.756
Female	41/61.9%	51/46.8%
Male	38/48.1%	58/53.2%
Fifth grade	29/36.7%	44/40.4%
Sixth grade	50/63.3%	65/59.6%
Note: Group 1 = accurate; Group 2 = Inaccurate		

covariance matrices in the groups, Box's test was conducted. Age, gender, and grade level were included in the analyses as covariates, and as possible discriminative variables. Gender and grade level were coded as dummy variables for the discriminant analysis.

Results

Group differences in the studied variables

Domain-general variables

Statistically significant differences between groups in EF were found, Wilks' $\lambda = .901, F(8, 145) = 5.251, p = .026, \eta p^2 = .099$. Only the covariate gender was statistically significant, Wilks' $\lambda = .913, F(8, 154) = 2.094, p = .039, \eta p^2 = .087$. ANCOVAs indicated small group differences in Impulsivity Control, Focus, Concentration, Working Memory, Planning, and Organization levels (Table 2). Although no important EF deficits were found

(scores in each component may range from 3 to 15), inaccurate students were reported by their parent as having more difficulties in all the EF components.

Domain-specific variables

Statistically significant differences between groups were found in the affective-motivational components, Wilks' $\lambda = .872, F(5, 178) = 5.242, p < .001, \eta p^2 = .128$. The covariates age, gender, and grade level, were not statistically significant. Accurate students showed higher motivation, and more positive perceptions and emotions related to mathematics, than their inaccurate pairs (Table 2). ANCOVA's indicated the existence of significant differences in all these variables. Effect sizes were moderate, but considerably higher in Perceived usefulness. Accurate students also showed significantly higher levels of mathematics achievement (expressed as academic grades). This variable generated the highest effect size. The covariates were not statistically significant.

Table 2
Differences between groups, and descriptive statistic for the dependent variables

	Group 1	Group 2	Differences			Total		
	M(SD)	M(SD)	F	p	ηp^2	M(SD)	K	S
Domain-general variables (EF)								
Impulsivity control	5.72 (2.20)	6.54 (2.52)	5.375	.022	.028	7.16 (2.70)	-.411	.443
Activity regulation	7.00 (2.29)	7.29 (2.71)	.535	.465	-	6.19 (2.42)	.663	.919
Emotional Control	6.51 (2.41)	6.72 (2.55)	.299	.585	-	6.63 (2.48)	-.151	.576
Focus	6.65 (2.35)	7.73 (2.60)	8.452	.004	.044	7.26 (2.56)	-.226	.614
Concentration	6.40 (2.47)	7.50 (2.50)	9.108	.003	.047	7.03 (2.53)	-.471	.412
Working Memory	5.23 (2.14)	6.22 (2.33)	8.754	.003	.045	5.81 (2.30)	.407	.856
Planning	6.38 (2.47)	7.68 (2.77)	10.820	.001	.055	7.14 (2.71)	-.245	.558
Organization	5.87 (2.66)	6.61 (2.47)	3.847	.050	.020	6.29 (2.57)	.303	.887
Flexibility	6.18 (2.28)	6.62 (2.00)	1.989	.160	-	6.43 (2.13)	.298	.703
Domain-specific variables								
Affective-motivational components								
Perceived usefulness	6.43 (3.39)	8.81 (3.92)	17.699	<.001	.089	7.81 (3.88)	.268	1.023
Self-efficacy	17.90 (1.87)	16.56 (3.36)	10.691	.001	.055	17.12 (2.90)	1.007	-.002
Intrinsic motivation	15.43 (3.18)	14.42 (3.04)	5.392	.021	.029	14.84 (3.13)	-.339	-.378
Anxiety	7.67 (3.08)	9.40 (3.81)	11.577	.001	.060	8.67 (3.62)	.006	.721
Enjoyment	5.89 (2.86)	7.11 (3.46)	6.113	.014	.032	6.60 (3.27)	.584	.675
Mathematics achievement								
Academic grades	7.56 (2.40)	5.80 (2.32)	24.640	<.001	.119	6.54 (2.51)	-1.078	-.069
Task variables								
Perceived difficulty								
Task 1	2.78 (1.39)	2.70 (1.65)	.163	.687	-	2.74 (1.54)	-1.075	.083
Task 2	2.75 (1.41)	3.06 (1.27)	2.723	.101	-	2.94 (1.33)	-1.289	.067
Time spent on the task								
Task 1(sec.)	418.46 (132.69)	354.49 (148.79)	8.60	.004	.045	382.12 (145.56)	.026	.509
Task 2 (sec.)	325.12 (98.05)	321.10 (122.62)	.084	.773	-	323.61 (113.04)	.286	.869
Note: Group 1 = accurate; Group 2 = Inaccurate; Sec. = seconds; EF = Executive Functions; K = Kurtosis; S = Skewness High scores in EF indicate deficit. Perceived Usefulness and Enjoyment are negatively worded (high scores indicate low perception of usefulness and enjoyment, respectively)								

Task characteristics

Perceived difficulty did not generate any significant differences. The covariates were not significant in Task 1. Age was statistically significant in Task 2, $F(1,182) = 4.367, p = .038, \eta^2 = .022$. Accurate students spent more time on the tasks than their inaccurate peers. These differences were significant in Task 1. No differences due to age, gender, and grade level were found.

Discriminant value of the variables predicting group membership

Box’s test confirmed equal covariance matrixes (Box’s $M = 3.127, p = .378$). Table 3 shows results from the discriminant analysis. Mathematics achievement (i.e., academic grades), perceived usefulness, and time spent on Task 1 significantly predicted group membership. Although significant, the inter-group variability explained by the discriminant function (“eigenvalue”) was low. The high values of Wilks’ λ , and the moderate canonical correlation corroborated this result.

Standardized and structure coefficients represent the correlations between the discriminant function and the variables, revealing academic grades as the most influential variable.

Function coefficients provided the resulting discriminant function. This function correctly classified 71.3% of subjects (79.9% from the inaccurate group, and 59.6% from the accurate group). Students with higher levels of mathematics achievement, who perceived mathematics as more useful, and spent more time performing the first task, were better calibrated.

Discussion

This study analyzed the impact of cognitive, affective-motivation and task-related variables on students’ post-performance judgment accuracy in two mathematical problems. It is important to note in this sense, that calibration was low in this sample, which is confirmed by the higher number of students in the inaccurate group. This is coherent with previous literature, which highlights the relationship between inaccurate judgments and the absence of effective metacognitive strategies (Bol et al., 2011; García, Rodríguez, González-Castro, González-Pianda, & Torrance, 2015; Hacker et al., 2008).

Regarding the factors studied in the present study, results indicated that:

First, accurate students were reported as having better levels of executive functioning by their parents, mainly in Impulsivity

Control, Working Memory, Focus, Concentration, Planning, and Organization. Thus, an association between calibration accuracy and EF was found. Inhibitory control seemed to be important in this relationship, maybe preventing students from making impulsive and erroneous judgments. These results resemble the classical distinction between the so-called “hot” and “cool” EF components (Zelazo & Carlson, 2012), suggesting a more significant relationship between the latter and calibration. In this sense, “cool” EF involve a rational processing of the information, and are intimately linked to learning processes, while “hot” EF are more related to emotional information processing and social adaptation. García, Rodríguez, González-Castro, Álvarez-García, & González-Pianda (in press) reported similar results using the EFS scale in its two versions (families and teachers), finding a significant relationship between mathematics metacognitive knowledge and the “cool” EF components.

Second, accurate students showed higher motivation, and more positive feelings and beliefs related to mathematics. Perceived usefulness was the most significant variable. These results are coherent with those found by Sheldrake et al. (2014), suggesting that students at upper elementary school already show motivations, beliefs and attitudes towards mathematics, including whether they enjoy the subject, how capable they believe they are of learning mathematics, or the value that they attribute to this subject (Adelson & McCoach, 2011). These components showed their impact on students’ ability to judge their performance in the present study. Additionally, an important relationship between mathematics achievement and calibration accuracy was found. This variable has substantially demonstrated to be an important determining factor in predicting calibration accuracy, but also intervention outcomes in this area (Bol et al., 2011; Hacker et al., 2008).

Third, a non-consistent pattern of results emerged as function of perceived difficulty. However, accurate students spent more time on the tasks, mainly on Task 1. This is coherent with the claim that working on the task would provide information about aspects such as task specific features, feelings about the correctness of the solution process and the extent to which it has been smoothly or plenty of obstacles, which may favor calibration accuracy (Boekaerts & Rozendaal, 2010). Thus, it stands to reason that calibration accuracy improves as the time spent on the task increases. This result could also be related to previous differences in impulsivity control. Higher levels of self-control may have refrained accurate students from making impulsive (and erroneous) judgments. However, this effect was mainly found in Task 1, suggests a greater impact of this variable when task is novel.

Table 3
Results of discriminant analysis, using stepwise method

	Standardized coefficients	Structure coefficients	Function coefficients	Wilks' Lambda	F	p
1. Academic grades	.716	.767	.692	.880	25.195	<.001
2. Perceived usefulness	-.561	-.655	-.128	.837	17.899	<.001
3. Time spent on task 1	.402	.463	.003	.812	14.136	<.001
Constant			-1.884			
Eigenvalue	.232					
Wilks' Lambda	.812					
($\chi^2; p$)	($\chi^2 = 38.246; p < .001$)					
Canonical correlation	.434					

Finally, the predictive value of some of the studied variables on calibration was confirmed, with mathematics achievement being the most important predictor. Students with higher levels of achievement, who perceived mathematics as more useful and spent more time performing the first task, showed higher calibration. Although significant group differences in EF were found, they did not significantly predict calibration accuracy. As stepwise method was used in the discriminant analysis, it is possible that previous differences in EF were partially explained by other variables, mainly mathematics achievement. This association has been widely reported, using both performance and observation EF measures (Bull & Lee, 2014; Presentación, Siegenthaler, Pinto, Mercader, & Miranda, 2015; Thorell, Veleiro, Siu, & Mohammad, 2013).

These findings advocate examining a wide range of factors when differences in calibration accuracy need to be explained. Not surprisingly, achievement was the most important variable. Nevertheless, the impact of affective-motivational variables was also demonstrated, suggesting the need to extend their study to early educational stages, in which references on this topic remain quite scarce (Adelson & McCoach, 2011; Sheldrake et al., 2014). The potential impact of performance time on post-performance judgment accuracy was also highlighted.

However, there are some limitations in the present study. First, the use of only two mathematical tasks and the dichotomous nature of the variables must be considered. Increasing the number

of tasks would allow obtaining more continuous measures, as well as exploring the influence of task characteristics at a deeper level. Additionally, different types of mathematical problems should be introduced, taking into consideration Boekaerts and Rozendaal's (2010) findings. Second, only one measure of executive functions was used, based on parents' reports. A more detailed evaluation of this construct, including other sorts of measures, based on students' performance, would provide additional information about the impact of these components on calibration. Third, it must be noted that the studied variables classified 71% of the sample correctly. This indicates that additional variables may be explaining group differences, which should be properly examined. Finally, although the impact of the analyzed variables on calibration accuracy has been confirmed, the opposite pattern should be explored. Thus, further research must be conducted on this issue. A comprehensive analysis of the factors involved in making accurate judgments would set the basis for more adapted and effective interventions to improve calibration, based on students' characteristics and needs. This is necessary, taking into consideration the effects of calibration on learning and task performance.

Acknowledgments

This work is funded by the I+D+i project with reference EDU2010-19798, and the support of a grant from the Program Severo Ochoa of the Principality of Asturias (BP 11-067).

References

- Adelson, J. L., & McCoach, D. B. (2011). Development and psychometric properties of the Math and Me Survey: Measuring third through sixth graders' attitudes towards mathematics. *Measurement and Evaluation in Counseling and Development*, 44, 225-247.
- Boekaerts, M., & Rozendaal, J. S. (2010). Using multiple calibration measures in order to capture the complex picture of what affects students' accuracy of feeling of confidence. *Learning and Instruction*, 20(4), 372-382.
- Bol, L., Hacker, D. J., Walck, C. C., & Nunnery, J. A. (2011). The effects of individual or group guidelines on the calibration accuracy and achievement of high school biology students. *Contemporary Educational Psychology*, 37, 280-287.
- Boonen, A. J. H., van der Schoot, M., van Wesel, F., de Vries, M. H., & Jolles, J. (2013). *Contemporary Educational Psychology*, 38, 271-279.
- Bouffard, T., & Narciss, S. (2011). Benefits and risks of positive bias in self-evaluation of academic competence. *International Journal of Educational Research*, 50(4), 205-256.
- Bull, E., & Lee, K. (2014). Executive functioning and mathematics achievement. *Child Development Perspectives*, 8(1), 36-41.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nded.). New York: Academic Press.
- Dupeyrat, C., Escríbe, C., Huet, N., & Régner, I. (2011). Positive biases in self-assessment of mathematics competence, achievement goals, and mathematics performance. *International Journal of Educational Research*, 50, 241-250.
- Finn, B., & Metcalfe, J. (2014). Over-confidence in children's multi-trial judgments of learning. *Learning and Instruction*, 32, 1-9.
- García, T., Álvarez-García, D., Cueli, M., González-Castro, P., & Álvarez, L. (2013). Psychometric properties of the Executive Functions Scale for Families (EFS-F). *Manuscript submitted for publication*.
- García, T., González-Pienda, J. A., Rodríguez, C., Álvarez-García, D., & Álvarez, L. (2014). Psychometric characteristics of the BRIEF scale for the assessment of executive functions in Spanish clinical population. *Psicothema*, 26(1), 47-54.
- García, T., Rodríguez, C., González-Castro, P., Álvarez-García, D., & González-Pienda, J.A. (In Press). Metacognition and Executive Functioning in Primary School. *Anales de Psicología*.
- García, T., Rodríguez, C., González-Castro, P., González-Pienda, J.A., & Torrance, M. (2015). Elementary students' metacognitive processes and post-performance calibration on mathematical problem-solving tasks. *Metacognition and Learning*, 10(2).
- Garner, J. K. (2009). Conceptualizing the relations between executive functions and self-regulated learning. *Journal of Psychology: Interdisciplinary and Applied* 143, 405-426.
- González-Pienda, J. A., Cueli, M., García, T., Suárez, N., Fernández, E., Tuero-Herrero, E., et al. (2012). Gender differences in attitudes towards Mathematics in compulsory school. *Revista Iberoamericana de Psicología y Salud*, 3(1), 55-73.
- Hacker, D. J., Bol, L., & Keener, M.C. (2008). Metacognition in education: A focus on calibration. In J. Dunlosky, & R. Bjork (Eds.), *Handbook of Memory and Metacognition* (pp. 429-455). Mahwah, NJ: Erlbaum.
- Howe, P., & Roebers, C. (2007). Developmental progression in the confidence-accuracy relationship in event recall: New insights provided by a calibration perspective. *Applied Cognitive Psychology*, 21(7), 871-893
- Jacobse, A. E., & Harskamp, E. G. (2012). Towards efficient measurement of metacognition in mathematical problem-solving. *Metacognition and Learning*, 7(2), 133-149.
- Kolkman, M. E., Hoijtink, H. H., Kroesbergen, E. H., & Leseman, P. P. M. (2013). The role of executive functions in numerical skills. *Learning and Individual Differences*, 24, 145-151.
- Montague, M., Enders, G., & Dietz, S. (2011). Effects of cognitive strategy instruction on math problem solving of middle school students with learning disabilities. *Learning Disability Quarterly*, 34(4), 262-272.

- Narciss, S., Koerndle, H., & Dresel, M. (2011). Self-evaluation accuracy and satisfaction with performance: Are there affective costs or benefits of positive self-evaluation bias? *International Journal of Educational Research*, 50, 230-240.
- Özsoy, G. (2012). Investigation of fifth grade students' mathematical calibration skills. *Educational Sciences: Theory & Practice*, 12(2), 1190-1194.
- Presentación, M. J., Siegenthaler, R., Pinto, V., Mercader, J., & Miranda, A. (2015). Math skills and executive functioning in preschool: clinical and ecological evaluation. *Revista de Psicodidáctica*, 20(1), 65-82.
- Rinne, L. F., & Mazocco, M. M. (2014). Knowing right from wrong in mental arithmetic judgments: Calibration of confidence predicts the development of accuracy. *PLoS ONE* 9(7), e98663.
- Roebers, C. M., Cimeli, P., Röthlisberger, M., & Neuenschwander, R. (2012). Executive functioning, metacognition, and self-perceived competence in elementary school children: an explorative study on their interrelations and their role for school achievement. *Metacognition and Learning*, 7, 151-173.
- Sheldrake, R., Mujtaba, T., & Reiss, M. J. (2014). Calibration of self-evaluations of mathematical ability for students in England aged 13 and 15, and their intentions to study non-compulsory mathematics after age 16. *International Journal of Educational Research*, 64, 49-61.
- Stolp, S., & Zabrocky, K. M. (2009). Contributions of metacognitive and self-regulated learning theories to investigations of calibration of comprehension. *International Electronic Journal of Elementary Education*, 2(1), 3-31.
- Thorell, L. B., Veleiro, A., Siu, A. F., & Mohammadi, H. (2013). Examining the relation between ratings of executive functioning and academic achievement: findings from a cross-cultural study. *Child Neuropsychology*, 19(6), 630-638.
- Welsh, M. C., & Pennington, B. F. (1988). Assessing frontal lobe functioning in children: Views from developmental psychology. *Developmental Neuropsychology*, 4, 199-230.
- Van Loon, M. H., de Bruin, A. B. H., van Gog, T., & Van Merriënboer, J. J. G. (2013). Activation of inaccurate prior knowledge affects primary-school students' metacognitive judgments and calibration. *Learning & Instruction*, 24, 15-25.
- Whimbey, A., & Lochhead, J. (1999). *Problem-solving and comprehension*. Hillsdale, NJ: Erlbaum.
- Williams, J. R. (2008). Revising the Declaration of Helsinki. *World Medical Journal*, 54, 120-125.
- Zelazo, P. D., & Carlson, S. M. (2012). Hot and Cool Executive Function in Childhood and Adolescence: Development and Plasticity. *Child Development Perspectives*, 6(4), 354-360.