SELF-CONFIDENCE JUDGMENTS IN MENTAL ROTATION ABILITY IN SIXTH GRADERS

JUICIOS DE CONFIANZA SOBRE LA CAPACIDAD DE ROTACIÓN MENTAL EN ESTUDIANTES DE SEXTO CURSO

Laura M. Fernández-Méndez[1], María Trinidad Maldonado-Recio[2], & María José Contreras[3]

Cómo referenciar este artículo/How to reference this article:

Fernández-Méndez, L. M., Maldonado-Recio, M. T., & Contreras, M. J. (2024). Self-confidence Judgments in Mental Rotation Ability in Sixth Graders [Juicios de confianza sobre la capacidad de rotación mental en estudiantes de sexto curso]. *Acción Psicológica, 21*(1-2), 5–18. https://doi.org/10.5944/ap.21.1-2.40238

Abstract

Monitoring self-performance is a significant metacognitive process in students' learning, helping students to adjust their performance in the tasks they are carrying out. However, to date, studies that evaluate these processes in primary school children, more specifically in spatial tasks, are scarce. The aim of this study is to analyze the self-confidence judgments and calibration index in two mental rotation tasks considering difficulty level. A total of 40 sixth graders, children aged between 11 and 12 years old applied a 5-point scale to evaluate, item by item, the confidence of their responses in two different mental rotation tasks (with high and low difficulty). It was calculated an index of calibration (Brier Score) as well for each task. The results indicated similar levels of confidence judgments in spatial tasks of varying difficulty where the calibration was different having the students more precision in easy test in comparison with difficult test. This evidence was discussed highlighting the importance of self-monitoring spatial performance strengthening the development of strategies that could regulate performance at this stage in this type of spatial tasks.

Keywords: metacognition; spatial ability; mental rotation; self-confidence judgments; Primary school children.

Resumen

La monitorización del propio rendimiento es un proceso metacognitivo significativo en el aprendizaje de los alumnos, que les ayuda a ajustar su rendimiento en las tareas

Correspondence address [Dirección para correspondencia]: Laura M. Fernández-Méndez, Facultad de Ciencias de la Salud, Universidad Rey Juan Carlos, Spain.
Email: laura.fernandez@urjc.es

ORCID: Laura M. Fernández-Méndez (https://orcid.org/0000-0002-4455-9928), María Trinidad Maldonado-Recio (https://orcid.org/0000-0002-9678-4852), & María José Contreras (https://orcid.org/0000-0001-6302-5238).
[1] Universidad Rey Juan Carlos (URJC), Spain.
[2] Universidad de Málaga (UMA), Spain.
[3] Universidad Nacional de Educación a Distancia (UNED), Spain.

Received: April 26, 2024.
Accepted: September 14, 2024.

que están realizando. Sin embargo, hasta la fecha son escasos los estudios que evalúan estos procesos en niños/as de primaria, más concretamente en tareas espaciales. El objetivo de este estudio es analizar los juicios de confianza y el índice de calibración en dos tareas de rotación mental considerando el nivel de dificultad. Un total de 40 alumnos/as de sexto curso con edades comprendidas entre 11 y 12 años evaluaron la confianza de sus respuestas ítem por ítem aplicando una escala de 5 puntos en dos tareas de rotación mental diferentes (con dificultad alta y baja). Se calculó también un índice de calibración (Brier Score) para cada tarea. Los resultados indicaron niveles similares de juicios de confianza en tareas espaciales de dificultad variable donde la calibración fue diferente, teniendo los/as estudiantes más precisión en la prueba fácil en comparación con la difícil. Estas evidencias se discuten resaltando la importancia de la monitorización del propio rendimiento espacial fortaleciendo el desarrollo de estrategias que puedan regular el rendimiento en esta etapa en este tipo de tareas espaciales.

Palabras clave: metacognición; capacidad espacial; rotación mental; juicios de confianza; niños/as de Educación Primaria.

Introduction

On a daily basis, people must make decisions that require reflecting on their own mental processes, which can occur in a more or less conscious way. When students decide whether they have studied enough for an exam or whether they should study more, their decision is based on their confidence to make an accurate judgment about their knowledge. Metacognitive judgments have implications in the regulation of our behavior when interacting with our environment and with other people (Koriat, 1995; Nelson & Narens, 1990). Metacognition refers to the processes used to monitor and regulate our mental activities (Fleming & Dolan, 2012). Schraw and Moshman (1995) proposed two different components of metacognition: the first, in reference to metacognitive knowledge, which contains all the knowledge and intuitions about our own cognition; and the second, in reference to the regulation of cognition that involves the planning, monitoring and evaluation of behavior. With respect to metacognitive monitoring, it represents a subjective evaluation of the possibility that one's own responses are correct or incorrect (Bjork et al., 2013; Nelson & Narens, 1980). Traditionally, to assess metacognitive monitoring, Likert-type scales have been used in which the participant was asked the degree of confidence regarding whether a given answer is correct or not. This assessment is known as a Confidence Judgment (CJ; Nelson & Narens 1980), being one of the most used measures of metacognition (Fleming & Lau, 2014). In these judgments, overconfidence can be observed when the CJ is greater than the actual performance obtained in the test. This bias reflects a lack of error detection, such as having confidence in an incorrect answer (Pallier et al., 2002; Rinne & Mazzocco, 2014). Conversely, lack of confidence occurs when confidence judgments are lower than actual performance, again reflecting imprecise detection of errors, such as a lack of confidence in correct answers (Pallier et al., 2002; Rinne & Mazzocco, 2014). Overconfidence can lead to a false sense of mastery that results in allocating less cognitive resources than necessary towards solving a problem; conversely, a lack of confidence can lead to the continued unnecessary allocation of resources to a well-performed problem (Dentakos et al., 2019). Studies have shown that individuals tend to be poor judges of their own state of knowledge, so children and adults are likely to display biased confidence judgments, leaning towards overconfidence rather than underconfidence (Bjork et al., 2013; Soderstrom et al., 2016). In the so-called hard-easy effect (Juslin et al., 2000; Lichtenstein & Fischhoff, 1977), people usually perform well on easier items although the feeling is related to low confidence in the response issued, while the more difficult items result in low performance but are evaluated with high confidence. Something similar occurs with the so-called Dunning-Kruger effect (Kruger & Dunning, 1999), whereby those who perform worse on tasks tend to overestimate their own results compared to those who perform better; In fact, evidence suggests that low performers have poorer perception or are less aware of their own thought processes than good performers (McIntosh et al., 2019). It has been shown that the difficulty is related to calibration, defined as the relationship between performance and item-by-item monitoring judgments (Dunlosky & Thiede, 2013; Hacker et al., 2009; Nietfeld et al., 2006; Schraw, 2009). A smaller discrepancy between confidence scores (CJ) and accuracy predicts better performance, where higher and more realistic self-confidence maximizes effective learning (Kleitman & Moscrop, 2010). Longitudinal studies have shown that students with better monitoring ability of their performance achieve better results (Rinne & Mazzocco, 2014).

Some studies have shown that having better metacognitive regulation skills is beneficial for academic achievement, as mathematical performance (Desender & Sasanguie, 2022). In this sense, the study by Jacobse and Harskamp (2012) showed that metacognitive skills are positively related to geometric performance, verbal problem-solving skills, and arithmetic performance. Likewise, Erickson and Heit (2015) demonstrated in two experiments that students tend to overestimate their own mathematical performance, which would be in line with the aforementioned Dunning-Kruger effect. The relationship between metacognition and task performance varies according to the type of cognitive task being performed (Van der Stel & Veenman, 2008). Over the years, research in the field of metacognition from an educational perspective has examined the processes involved in learning, remembering, and understanding. However, there has been a recent increase in research on metacognitive processes in other domains such as reasoning and problem solving (Ackerman & Thompson, 2017) which had received less attention until recent studies. A growing focus of interest in the field of metacognition is that related to spatial cognition, where inaccurate monitoring of one's spatial abilities could have several negative implications. In this regard, spatial reasoning can be defined as the ability to generate, retrieve, maintain, and manipulate visuospatial information (Lohman, 1996). Within of visuospatial abilities, Mental Rotation (MR) is defined by Linn and Petersen (1985) as a process in which people can mentally rotate a figure to align it on one plane with another reference figure, reflecting on whether both are equal or not. It is valuable that people can accurately monitor and evaluate their own spatial cognitive performance because perceptions of their spatial performance can influence their use of spatial strategies and decisions to engage in tasks that require spatial thinking (Ariel & Moffat, 2018). The monitoring of spatial thinking has been the focus of interest in few studies related to developmental factors (Ariel & Moffat, 2018).

Regarding metacognitive processes during development, preschoolers show a rather imprecise monitoring, where the relationship between judgments and objective performance is weak (van Loon & Roebers, 2017). Subsequently, and especially between 7 and 10 years of age, there is an improvement in the ability to monitor one's own performance, although it is far from perfect (Krebs & Roebers, 2012; Roebers et al., 2019; Schneider & Löffler, 2016; Shin et al., 2007). In primary school, children gain experience and apply their metacognitive skills to identify the current state of their learning progress (Desoete & Roeyers, 2003; Dignath et al., 2008). In this sense, Zelazo's (2004) study related the increase in conscious reflection on one's own knowledge with a better allocation of available cognitive resources to face the demands of a task. Another study (Chevalier & Blaye, 2016) showed that children aged between 6 and 10 years could effectively monitor their cognitive resources and proactively prepare for the different types of stimuli that could appear in a task that changed unpredictably. However, there are relatively few studies that have evaluated metacognition processes with spatial tasks in primary school. These experiments showed that on-task behavior was more affected by children's confidence in their responses than by the accuracy of those responses (Contreras et al., 2020). These results have educational implications, demonstrating that primary school children are sensitive to the difficulty of more or less demanding items, and teachers can take the opportunity to recommend that they become aware of the difficulty and thus spend more or less time thinking about whether their answers to a problem are correct or not, and to ask questions when they are in doubt. With this awareness of the difficulty, the person can learn to seek feedback when more support is needed (Wall et al., 2016). Therefore, this raises the question about what supervision processes would be like at early ages in other types of tasks, specifically in spatial cognition tasks, where studies that evaluate the importance of metacognitive reasoning are scarce (Contreras et al., 2020).

Montoya et al. (2021) emphasize the importance of establishing research lines regarding these judgments, as they serve as signals or inducers of knowledge and regulation (Dunlosky & Metcalfe, 2009). However, the literature has demonstrated a lack of studies explaining part of the performance in a MR ability due to non-cognitive factors, making it necessary to investigate in depth CJ of spatial monitoring during the development.

The objective of the present study was to analyze the relationship between the difficulty of a MR task, confidence judgments and calibration in primary school children. Performance (percentage of correct answers) will be analyzed across two MR tasks of different difficulty (high and low). High performance would be expected in an easy task compared with another similar task more difficult. With respect to confidence judgment (measured using a 5-point scale), high level would be expected in easy task in relation to difficult one. If an effect of task difficulty appears, a higher level of precision (calibration measure) is expected in the task considered easier compared to the task considered to be more difficult.

Methods

Participants

A total of 40 students (23 girls and 17 boys) between 11-12 years of age (M_{age} = 11.34, SD_{age} = 0.54) who were in the 6th year of Primary Education took part in the study. All participants and their families were informed of the description of the study, children consented to participate, and their legal representatives signed the information and informed consent sheet. The study was included in a competitive research project whose procedures were approved by the UNED Ethics Committee.

Materials

Mental Rotation Task and Confidence Judgments

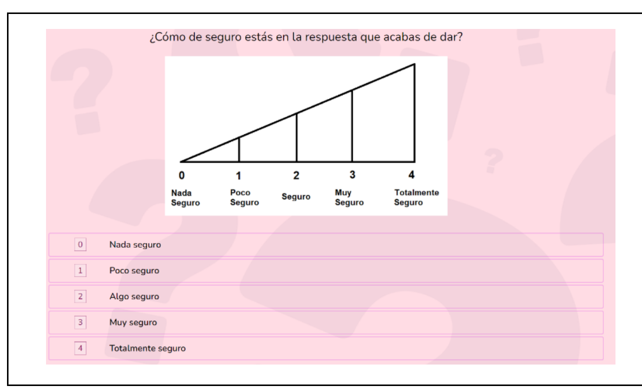
The MR test used in this study was an adaptation of the MR training designed by Rodán et al. (Rodán et al., 2016, Rodán et al., 2019), the PERM-2D for Primary Education and Secondary Education (which in our study correspond to low and high difficulty, respectively). The difficulty was operationalized considering items from Primary Education training (Rodán et al., 2019) as a low difficulty (PERM 1 hereinafter) and items from Secondary Education training (Rodán et al., 2016) as a high difficulty (PERM 2 hereinafter).

The items from the third and last Primary Education training session (the last session contained the most complex items for second grade, which correspond to 7-8 years of age) and the items from the first Secondary Education training session (the first session contained the easiest items for first year of secondary school, which correspond to 12-13 years of age) were used. These sessions were chosen since the ages of the sample of the present study were between the primary and secondary grades to which the original sessions were applied, so the first test had to be accessible for the students while the second test contemplated a more difficult level for these ages. Each training task at each stage is described in more detail below.

In addition to the rotation task, participants had to make a CJ after solving each MR trial. Therefore, a CJ was made for each rotation item resolved. Following each MR item, the participant was shown a question regarding their confidence in the answer they had just given to the MR trial to which the participant decided, on a scale from 0 to 4, how confident they were with the answer they had just given. The scale ranged from 0 "not at all sure" to 4 "totally sure" in their response (see Figure 1). The scale was adapted from the scale used by Neys and Fereman (2013). The index of confidence judgments was the average of responses on the scale from 0 to 4.

Figure 1.

Scale to measure the confident of the response



Note: Translation of the confidence judgement item. "¿Cómo de seguro estás en la respuesta que acabas de dar?" = "How sure are you about the answer you just gave?"; "nada seguro" = "not at all sure"; "poco seguro" = "not very sure"; "seguro" = "sure"; "muy seguro" = "very sure"; "totalmente seguro" = "totally sure".

PERM 1 with Confidence Judgment.

The original PERM-2D task for Primary Education consisted of 30 practice cards (10 for each of the three sessions) and 450 training cards divided into two blocks (150 for each of the sessions carried out in two blocks of 75, with a break in between). It was applied to boys and girls aged between 7 and 8 years (Gimeno, 2014; Rodán et al., 2019).

PERM 1 in the present study was composed of 15 experimental items. The items were concrete figures as a hairdryer, teacup, action toys, musical notes and some abstract figures, among others (see supplemental online material). The participant had to imagine movements corresponding to transformations and MR, deciding in each case which of the two drawings on the right "1" and "2" (see Figure 2) fitted rotated into the reference mold on the left, taking into account that only one of them could fit in. The participant had to press the Z or M key (Drawing 1 or Drawing 2, respectively) on their keyboard, as shown in Figure 2. Four practice items were also included, where, after his/her answer, the participant was shown a series of images in movement that simulated the rotation that the piece had to do so that it fitted into the mold image, thus making it easier for the participant to understand the test. The last 3 trials incorporated the confidence scale. Additionally, the test consisted of 15 confidence items that were displayed after each MR item. In the MR task, each correct item corresponds to one point, with 15 being the maximum score. To make performance in this task comparable with the next task, of greater difficulty with a greater number of items, the probability of correct answers (number of correct items/total number of items) was used as an index. The estimated time to perform this test was 15 minutes and there was no time limit to solve it.

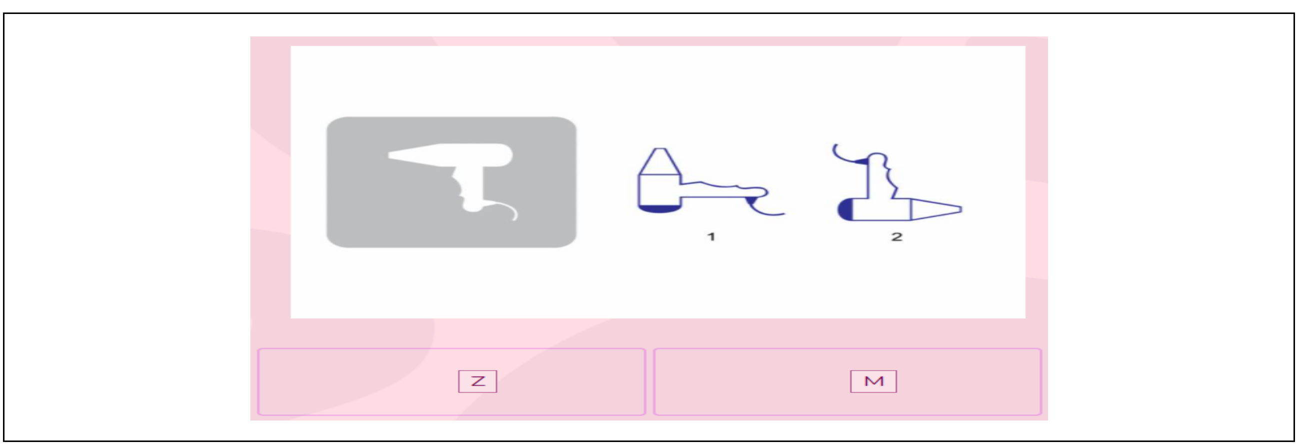
PERM 2 with Confidence Judgment

The original PERM-2D task for Secondary Education consisted of a total of 30 practice slides (10 per session) and 300 training slides, in which two MR decisions had to be made per slide (100 slides per session divided into two blocks of 50) applied to adolescents between 14 and 15 years of age (Rodán et al., 2016).

PERM 2 in the present study consisted of 30 experimental MR items (15 pictures and two decisions for each) and eight practice items. The items of PERM 2 were figures made mainly by objects, squares, rectangles, circles, or lines with different dimensions and colors (see supplemental online material). In the experimental items, the participant was shown a drawing or mold on the left and two drawings "1" and "2" on the right. In each trial, first they were asked whether drawing 1 fit the mold, then

Figure 2.

Example of a practice item sheet for the PERM 1 test.



the same image was shown again, but this time they were asked whether drawing 2 fit the mold. In this sense, both drawing could fit to mold. To answer, each participant had to imagine the rotation of the object and decide whether it fitted or not by pressing the Z (yes it fits) or M (it does not fit) keys, as can be seen in Figure 3. Prior to the experimental items, participants were presented eight practice items where, after the participant's response, it was shown whether it fitted into the mold or not through the rotation of the object using a series of moving images that simulated the rotation. The confidence scale was incorporated in 7 of the 8 practice trials. One point was awarded for each correct answer issued, with 30 being the maximum score. In the comparison analyses of the two tasks by difficulty, to make performance comparable with the PERM1 task that had fewer items, the probability of correct answers (number of correct items/total items) was used as an index. The estimated time to perform this test was 25 minutes and there was no time limit to solve it. The procedure for evaluating the confidence judgments for each response issued was the same as that described for the PERM1 task. The reliability indexes were of 0.90 and 0.92 for CJ in PERM 1 and PERM 2, respectively.

Procedure

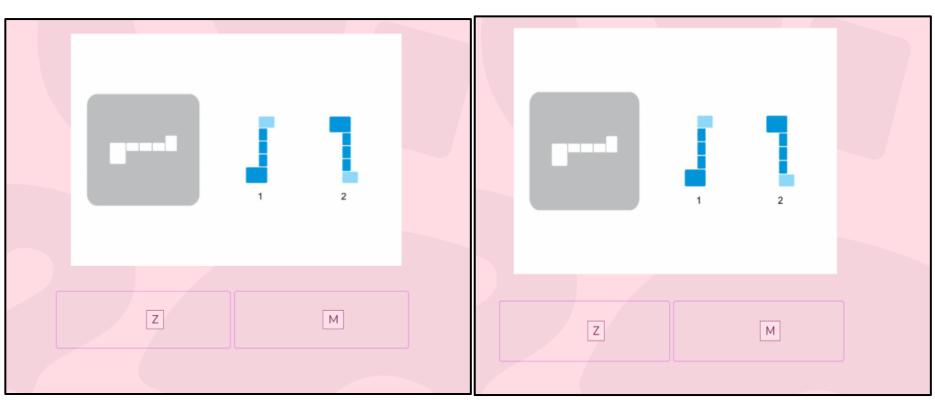
All tests were carried out collectively online with laptops. The collect data was done in a quiet room in two different schools and the participants completed the tasks under supervision. Participants completed the tasks in approximately 55 minutes in a single session. The children were tested after their mid-morning recess break session. In the first school, the tests involved 30 students supervised by three examiners. In the second school, the tests involved a total of 10 students divided into groups of three and four students with two examiners. First, the participant completed PERM 1 and after they completed the second task, PERM 2. Between the two tests, a 5-minute break was provided.

Data Analyses

Data were analyses using the SPSS 21.0 statistical support (IBM, 2011), and all analyzes were performed with a significance level of 0.05 (95% confidence interval, CI). To assess whether normal distributions by group, and since our sample was less than 50 cases, the Shapiro-Wilk test was performed. This indicated that the distribution was normal for calibration and confidence judgment, and

Figure 3.

Examples of practice sheets for the adapted PERM 2 test. Practice item number 1 (A) and practice item number 2 (B).



therefore it was decided to use parametric tests. However, the Shapiro-Wilk test no showed that the percentage of correct answers (corresponding of PERM1 and PERM2) had a normal distribution, conducting no parametric tests.

To assess the absolute precision of the confidence judgments, Murphy's (1973) calibration component called the Brier Score was calculated for each participant in each item in the PERM 1 and PERM 2 tests. This index relates the objective performance with the CJ (with a range of 0 to 1), where calibration values close to zero reflect an absolute perfect match between the objective response and the metamnemonic judgment, while those increasing values reflect deviations from the response-judgment pairing up to 1, implying a low match. Various studies have shown that the Brier Score is an adequate index to estimate the accuracy of confidence judgments in MR tasks (Ariet et al., 2018; Ariel & Moffat, 2018). However, in these studies, the CJ scores had a range of 0 to 100 and since the CJ that the boys and girls had to make in the present study had a range of 0 to 4, the score had to be adapted from ordinal to percentages. To establish equivalence, a confidence value of 1 corresponds to 25 % confidence, a confidence value of 2 corresponds to 50 % confidence, a confidence value of 3 corresponds to a 75 % confidence level, and a confidence level of 4 corresponds to being 100 % sure that the answer issued is correct. With this equivalence, the calibration index was calculated.

The differences between the probability of correct answers between PERM 1 and PERM 2 were analyzed through Wilcoxon test. With respect to the confidence judgments and the calibration (intra-subject factor) between PERM 1 and PERM 2 were analyzed through two repeated measures ANOVA’s.

As there was a problem with the computers of some of the children, the size of the total sample for the different analyses carried out varies depending on the test administered, with a maximum of 40 and a minimum of 16. In addition, there was a computer error in the last study question of the PERM 2 task, causing two items to be invalidated. Thus, they were not computed into the statistical analyses.

Results

Descriptive statistics are showed in Table 1.

Table 1.

Descriptive statistics of the performance, confidence index, and calibration in PERM1 and PERM2.

	N	Mean	*SD*
1. Percentage of correct answers (hits) PERM 1	40	.80	.15
2. Percentage of correct answers (hits) PERM 2	40	.63	.15
3.Confidence Index PERM 1	16	3.15	.53
4. Confidence Index PERM 2	16	3.20	.54
5. Calibration PERM 1	16	.14	.06
6. Calibration PERM 2	16	.25	.13

Note: *SD*: Standard Deviation.

The results of the Wilcoxon test showed a main effect of the difficulty of the task (Z = -5.014, p = .000, η_p^2 = .607), with the PERM 1 task being easier with a higher percentage of correct answers (M = .80, SD = .15) than the PERM 2 task (M =.63, SD = .15).

For confidence judgments, the results of the ANOVA were not significant.

The ANOVA showed the calibration effect [$F_{(1,15)}$ = 18.213; p = .001; η_p^2 = .548, β = .978], where the participants were more adjusted in calibrating their performance on the PERM 1 task (M = .14, SD = .06) compared to the PERM 2 task (M = .25, SD = .13). The mean square error (MSE) was .006.

Discussion

The main objective of this research was to analyze the relationship between the difficulty of a MR task, confidence judgments and calibration in 6th Primary Education students.

Confidence judgments constitute metacognitive monitoring processes that in turn influence control processes, where individuals assign more or fewer resources or change strategies to cope with the task they are performing

(Nelson & Narens, 1990). A good match between the objective performance and the CJ made, known as "calibration" (Ariel et al., 2018), is decisive for the success of the task. This is especially important in students, where a good calibration will entail adequate management of cognitive resources, assigning such resources to areas or knowledge where a lack of security is detected and releasing resources in those tasks that are considered to have been answered assuredly. A poor calibration can lead to either overconfidence in tasks with low objective performance, which will imply a lower allocation of resources than necessary, or it can lead to a lack of confidence in tasks where the objective performance is high, implying unnecessary and continuous allocation of resources to solve the task. In the exploratory analysis involved in this study, the data showed that a greater difficulty did not imply either an increase or a decrease in judgment confidence. Although a decrease in the level of confidence would be expected in the more difficult test (PERM 2), the children did not show a lower level of confidence with respect to the judgments made in the simpler MR test (PERM 1) although their performance was significantly lower in the more difficult rotations. Note that the average confidence in both tests is quite high. The high scores in the confidence judgments of both tests are in line with other studies where it is shown that individuals tend to bias their confidence judgments, leaning towards overconfidence (Bjork et al., 2013; Soderstrom et al., 2016). One possible explanation is that their previous experience with the simple test made them overconfident for the subsequent, more complex task. Likewise, asking participants themselves to provide confidence judgments about the accuracy of their own responses can affect the task they are going to perform (Double & Birney, 2019; Song et al., 2021). It is possible that this poor evaluation and overconfidence may also have influenced their spatial performance (Ariel & Moffat, 2018).

Regarding calibration, the results show that students were more accurate in the easier task compared to the more complex task. In line with the hard-easy effect (Juslin et al., 2000; Lichtenstein & Fischhoff, 1977), it seems that there is worse calibration when the complexity of the task increases, indicating that the feeling of security remains high, even if performance decreases.

The developmental stage is a factor to take into account regarding the development of high-level cognitive processes, such as metacognitive skills, where it is expected that as children grow, both the frequency and quality of metacognitive evaluations will increase (Ohtani & Hisasaka, 2018). In fact, it appears that the developmental stage modulates the relationship between metacognition and academic performance. For example, Murayama et al. (2013) showed that metacognitive strategies predict mathematical achievement in 7th grade children but not in 5th grade children. Taking into consideration these results as well as those of the present study, further research is necessary regarding monitoring processes at early ages in other types of tasks, specifically in spatial cognition tasks, where studies that evaluate the importance of metacognitive reasoning are scarce (Contreras et al., 2020).

Contreras et al. (2020) was the first study in which confidence judgments in spatial tests in second grade of Primary Education were analyzed, showing that children are overconfident in their answers. The present results extend those found in the previous study, indicating that, in sixth grade of Primary Education (11-12 years of age), children continue to show overconfidence in their responses in spatial tests, a notable result at this age where a better adjustment between actual performance and feeling confident would be expected. This work presents an advance regarding the calibration measure, the Brier Score index, an index that has been calculated as a metacognitive measure for the adult population (Ariel et al., 2018; Ariel & Moffat, 2018; Cooke-Simpson & Voyer, 2007) but, as far as we know, not in Primary Education children.

Through this research, we highlight the following limitations and future directions that this study raises. One of the main limitations of this study is the small sample size. For this reason, the results derived have an exploratory nature, although they are a good starting point to delve into the metacognitive factors related to spatial tasks.

A line of future research would be to propose tools or strategies that help students to improve these cognitive processes, including seeking more opportunities to explore the different meta-level processes, since the previous step before adjusting our strategies to face a task is precisely to accurately evaluate that we are not doing it right.

There is evidence that the type of metacognitive measure used in the study can moderate the metacognition-intelligence relationship (Ohtani & Hisasaka, 2018). In fact, simply having feedback about monitoring processes can improve both calibration and performance on the task (Nietfeld et al., 2006). Even obtaining participants' responses to metacognition measures will alter the processes for which the measures were constructed (Double & Birney, 2019), so designing an effective, accurate, and valid metacognitive measure of metacognition is a very important line of study for this field. Another field that could be explored in the future would be to evaluate sex differences in both calibration and confidence judgments at these ages, given that in the adult population it has been shown that women are less confident (Ariel et al., 2018).

In conclusion, in a MR task, children between 11 and 12 years of age did not decrease their judgment confidence when the difficulty of the task increases, showing a poor calibration when the cognitive demands of task grow. The confidence feeling about their responses indicates the children are unaware about their mistakes. In an academic context, this result could imply a serious problem because a poor calibration would not lead a good cognitive resources management. Future studies could propose intervention strategies to improve calibration in primary school children.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

This work was partially supported by 1) Spanish Ministry of Science, Innovation and Universities (grant number RTI2018-098523-B-I00), 2) Spanish Ministry of Science and Innovation (Grant number PID2021-125677OB-I00, and 3) Teaching Innovation Project for Teaching Innovation Groups Call 2022/2023: "Learning by researching through the Turtle Lab platform in the TFM subject: a pilot test".

Supplemental online material

The research materials of PERM1 and PERM 2 have been uploaded to a public repository and can be accessed via the following links: https://doi.org/10.5944/ap.21.1-2.40238

References

Ackerman, R. & Thompson, V. A. (2017). Meta-reasoning: Monitoring and Control of Thinking and Reasoning. *Trends in Cognitive Sciences. 21*, 607–617. https://doi.org/10.1016/j.tics.2017.05.004

Ariel, R., Lembeck, N. A., Moffat, S., & Hertzog, C. (2018). Are there Sex Differences in Confidence and Metacognitive Monitoring Accuracy for Everyday, Academic, and Psychometrically Measured Spatial Ability? *Intelligence, 70*, 42–51. https://doi.org/10.1016/j.intell.2018.08.001

Ariel, R. & Moffat, S. D. (2018). Age-related Similarities and Differences in Monitoring Spatial Cognition. Neuropsychology, development, and cognition. *Section B, Aging, Neuropsychology and Cognition, 25*(3), 351–377. https://doi.org/10.1080/13825585.2017.1305086

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, Techniques, and Illusions. *Annual Review of Psychology, 64*, 417–444. https://doi.org/10.1146/annurev-psych-113011-143823

Chevalier, N. & Blaye, A. (2016). Metacognitive Monitoring of Executive Control Engagement during Childhood. *Child Development, 87*(4), 1264–1276. https://doi.org/10.1111/cdev.12537

Contreras, M. J., Meneghetti, C., Uttal, D. H., Fernández-Méndez, L. M., Rodán, A., & Montoro, P. R. (2020). Monitoring the own Spatial Thinking in Second Grade of Primary Education in a Spanish

School: Preliminary Study Analyzing Gender Differences. *Education Sciences, 10*(9), Article 237. https://doi.org/10.3390/educsci10090237

Cooke-Simpson, A. & Voyer, D. (2007). Confidence and Gender Differences on the Mental Rotations Test. *Learning and Individual Differences, 17*(2), 181–186. https://doi.org/10.1016/j.lindif.2007.03.009

Dentakos, S., Saoud, W., Ackerman, R., & Toplak, M. E. (2019). Does Domain Matter? Monitoring Accuracy across Domains. *Metacognition and Learning, 14*(3), 413–436. https://doi.org/10.1007/s11409-019-09198-4

Desender, K. & Sasanguie, D. (2022). Math Anxiety Relates Positively to Metacognitive Insight into Mathematical Decision Making. *Psychological Research, 86*, 1001–1013. https://doi.org/10.1007/s00426-021-01511-8

Desoete, A. & Roeyers, H. (2003). Can off-line Metacognition Enhance Mathematical Problem Solving? *Journal of Educational Psychology, 95*(1), 188–200. https://doi.org/10.1037/0022-0663.95.1.188

Dignath, C., Büttner, G., & Langfeldt, H. (2008). How Can Primary School Students Learn Self-Regulated Learning Strategies most Effectively? A Meta-analysis on self-Regulation Training Programmes. *Educational Research Review, 3*(2), 101–129. https://doi.org/10.1016/j.edurev.2008.02.003

Double, K. S. & Birney, D. P. (2019). Reactivity to Measures of Metacognition. *Frontiers in Psychology, 10,* Article 2755. https://doi.org/10.3389/fpsyg.2019.02755

Dunlosky, J. & Thiede, K. W. (2013). Four Corner-Stones of Calibration Research: Why Understanding Students' Judgments Can Improve their Achievement. *Learning and Instruction, 24*, 58–61. https://doi.org/0.1016/j.learninstruc.2012.05.002

Erickson, S. & Heit, E. (2015). Metacognition and Confidence: Comparing Math to other Academic Subjects. *Frontiers in Psychology, 6*, Article 742. https://doi.org/10.3389/fpsyg.2015.00742

Fleming, S. M. & Dolan, R. J. (2012). The Neural Basis of Metacognitive Ability. *Philosophical Transactions of the Royal Society of London, 367*(1594), 1338–1349. https://doi.org/10.1098/rstb.2011.0417

Fleming, S. M. & Lau, H. C. (2014). How to Measure Metacognition. *Frontiers in Human Neuroscience, 8,* Article 443. https://doi.org/10.3389/fnhum.2014.00443

Gimeno, P. (2014). Intervención en las aptitudes visoespaciales y su relación con las matemáticas y la memoria operativa en estudiantes de 2º curso de Educación Primaria Obligatoria (EPO) [*Intervention in Visuospatial Skills and their Relationship with Mathematics and Working Memory in 2nd-grade Compulsory Primary Education Students (EPO)*]. [Master's thesis, National University of Distance Education]. UNED Repository. https://hdl.handle.net/20.500.14468/24628

Hacker, D. J., Dunlosky, J., & Graesser, A. C. (Eds.). (2009). *Handbook of Metacognition in Education.* Routledge. https://doi.org/10.4324/9780203876428

IBM Corporation. (2011). *SPSS Statistics for Windows*, Version 20.0. Author.

Jacobse, A. E. & Harskamp, E. G. (2012). Towards Efficient Measurement of Metacognition in Mathematical Problem Solving. *Metacognition Learning, 7,* 133–149. https://doi.org/10.1007/s11409-012-9088-x

Juslin, P., Winman, A., & Olsson, H. (2000). Naive Empiricism and Dogmatism in Confidence Research: A Critical Examination of the Hard–Easy

Effect. *Psychological Review, 107*(2), 384–396. https://doi.org/10.1037/0033-295X.107.2.384

Kleitman, S. & Moscrop, T. (2010). Self-confidence and Academic Achievements in Primary-School Children: Their Relationships and Links to Parental Bonds, Intelligence, Age, and Gender. In A. Efklides & P. Misailidi (Eds.), *Trends and Prospects in Metacognition Research* (pp. 293–326). Springer Science + Business Media. https://doi.org/10.1007/978-1-4419-6546-2_14

Krebs, S. S. & Roebers, C. M. (2012). The Impact of Retrieval Processes, Age, General Achievement Level, and Test Scoring Scheme for Children's Metacognitive Monitoring and Controlling. *Metacognition and Learning, 7*(2), 75–90. https://doi.org/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 and Social Psychology, 77*(6), 1121–1134. https://doi.org/10.1037/0022-3514.77.6.1121

Koriat, A. (1995). Dissociating Knowing and the Feeling of Knowing: Further Evidence for the Accessibility Model. *Journal of Experimental Psychology: General, 124*(3), 311–333. https://doi.org/10.1037/0096-3445.124.3.311

Lichtenstein, S. & Fischhoff, B. (1977). Do those who Know more also Know more about how much they Know? *Organizational Behavior and Human Performance, 20*(2), 159–183. https://doi.org/10.1016/0030-5073(77)90001-0

Linn, M. C. & Petersen, A. C. (1985). Emergence and Characterization of Sex Differences in Spatial Ability: A meta-analysis. *Child Development, 56* (6), 1479–1498. https://doi.org/10.1111/j.1467-8624.1985.tb00213.x

Lohman, D. F. (1996). Spatial Ability and G. In I. Dennis & P. Tapsfield (Eds.), *Human Abilities: Their Nature and Measurement* (pp. 97–116). Lawrence Erlbaum. https://doi.org/10.4324/9780203774007

McIntosh R. D., Fowler E. A., Lyu T., & Della Sala S. (2019). Wise up: Clarifying the Role of Metacognition in the Dunning-Kruger Effect. *Journal of Experimental Psychology: General, 148*(11), 1882–1897. https://doi.org/10.1037/xge0000579

Montoya, D., Orrego, M., Puente, A., & Tamayo, O. (2021). Metacognitive Judgments as an Emerging Research Field. A Systematic Review (2016-2020). *Revista Latinoamericana de Estudios Educativos, 17*(1), 188–223. https://doi.org/17151/rlee.2021.17.1.10

Murayama, K., Pekrun, R., Lichtenfeld, S., & vom Hofe, R. (2013). Predicting Long-Term Growth in Students' Mathematics Achievement: The Unique Contributions of Motivation and Cognitive Strategies. *Child Development, 84,* 1475–1490. https://doi.org/10.1111/cdev.12036

Murphy, A. H. (1973). A New Vector Partition of the Probability Score. *Journal of Applied Meteorology, 12,* 595–600. http://doi.org/10.1175/1520-0450(1973)012<0595:ANVPOT>2.0.CO;2

Nelson, T. O. & Narens, L. (1980). Norms of 300 General-Information Questions: Accuracy of Recall, Latency of Recall, and Feeling-of-Knowing Ratings. *Journal of Verbal Learning and Verbal Behavior, 19*(3), 338–368. https://doi.org/10.1016/S0022-5371(80)90266-2

Nelson, T. O. & Narens, L. (1990). Metamemory: A Theoretical Framework and New Findings. In G. Bower (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory* (pp. 125–173). Academic Press. https://doi.org/10.1016/S0079-7421(08)60053-5

Neys, W. D. & Fereman, V. (2013). Development of Heuristic Bias Detection in Elementary School.

Developmental Psychology, 49(2), 258–269. https://doi.org/10.1037/a0028320

Nietfeld, J. L., Cao, L., & Osborne, J. W. (2006). The Effect of Distributed Monitoring Exercises and Feed-back on Performance, Monitoring Accuracy, and Self-Efficacy. *Metacognition and Learning, 1*(2), 159–179. https://doi.org/10.1007/s10409-006-9595-6

Ohtani, K. & Hisasaka, T. (2018). Beyond Intelligence: A Meta-Analytic Review of the Relationship among Metacognition, Intelligence, and Academic Performance. *Metacognition and Learning, 13*, 179–212. https://doi.org/10.1007/s11409-018-9183-8

Pallier, G., Wilkinson, R., Danthiir, V., Kleitman, S., Knezevic, G., Stankov, L., & Roberts, R. D. (2002). The Role of Individual Differences in the Accuracy of Confidence Judgments. *The Journal of General Psychology, 129*(3), 257–299. https://doi.org/10.1080/00221300209602099

Rinne, L. F. & Mazzocco, M. M. (2014). Knowing Right from Wrong in Mental Arithmetic Judgments: Calibration of Confidence Predicts the Development of Accuracy. *PLoS One, 9*(7), Article e98663. https://doi.org/10.1371/journal.pone.0098663

Rodán, A., Contreras, M. J., Elosúa, M. R., & Gimeno, P. (2016). Experimental but not Sex Differences of a Mental Rotation Training Program on Adolescents. *Frontiers in Psychology, 7*, Article 1050. https://doi.org/10.3389/fpsyg.2016.01050

Rodán, A., Gimeno, P., Elosúa, M. R., Montoro, P. R., & Contreras, M. J. (2019). Boys and Girls Gain in Spatial, but not in Mathematical Ability after Mental Rotation Training in Primary Education. *Learning and Individual Differences, 70*, 1–11. https://doi.org/10.1016/j.lindif.2019.01.001

Roebers, C. M., Mayer, B., Steiner, M., Bayard, N. S., & van Loon, M. H. (2019). The Role of Children's Metacognitive Experiences for Cue Utilization and Monitoring Accuracy: A Longitudinal Study. *Developmental Psychology, 55*(10), 2077–2089. https://doi.org/10.1037/dev0000776

Schneider, W. & Löffler, E. (2016). The Development of Metacognitive Knowledge in Children and Adolescents. In J. Dunlosky & S.K. Tauber (Eds), *The Oxford Handbook of Metamemory* (pp. 491–518). Oxford University Press.

Schraw, G. (2009). A Conceptual Analysis of Five Measures of Metacognitive Monitoring. *Metacognition and Learning, 4*(1), 33–45. https://doi.org/10.1007/s11409-008-9031-3

Schraw, G. & Moshman, D. (1995). Metacognitive Theories. *Educational Psychology Review, 7,* 351–371. https://doi.org/10.1007/BF02212307

Shin, H., Bjorklund, D. F., & Beck, E. F. (2007). The Adaptive Nature of Children's Overestimation in a Strategic Memory Task. *Cognitive Development, 22*(2), 197–212. https://doi.org/10.1016/j.cogdev.2006.10.001

Soderstrom, N. C., Yue, C. L., & Bjork, E. L. (2016). Metamemory and Education. In J. Dunlosky & S. K. Tauber (Eds.), *The Oxford handbook of Metamemory* (pp. 197–215). Oxford University Press.

Song, J. H., Loyal, S., & Lond, B. (2021). Metacognitive Awareness Scale, Domain Specific (MCAS-DS): Assessing Metacognitive Awareness During Raven's Progressive Matrices. *Frontiers in Psychology, 11,* Article 3683. https://doi.org/10.3389/fpsyg.2020.607577

Van der Stel, M. & Veenman, M. V. J. (2008). Relation between Intellectual Ability and Metacognitive Skill Fulness as Predictors of Learning Performance of Young Students Performing Tasks in Different Domains. *Learning Individual Different 18,* 128–134. https://doi.org/10.1016/j.lindif.2007.08.003

Van Loon, M. H. & Roebers, C. M. (2017). Effects of Feedback on Self-Evaluations and Self-Regulation in Elementary School. *Applied Cognitive Psychology, 31*(5), 508–519. https://doi.org/10.1002/acp.3347

Wall, J. L., Thompson, C. A., Dunlosky, J., & Merriman, W. E. (2016). Children Can Accurately Monitor and Control their Number-Line Estimation Performance. *Developmental Psychology, 52*(10), Article 1493. https://doi.org/10.1037/dev0000180

Zelazo, P. D. (2004). The Development of Conscious Control in Childhood. *Trends in Cognitive Sciences, 8*(1), 12–17. https://doi.org/10.1016/j.tics.2003.11.001