UNIVERSIDAD DE OVIEDO

DEPARTAMENTO DE PSICOLOGÍA

Programa de Doctorado en Psicología regulado
por el Real Decreto 1393/2007

TESIS DOCTORAL

HABILIDADES DE AUTORREGULACIÓN Y EVALUACIÓN ON-LINE DEL PROCESO DE RESOLUCIÓN DE PROBLEMAS MATEMÁTICOS

Self-regulation Skills and On-line Assessment of the Process Involved in Solving Mathematical Problems

Autora: Trinidad García Fernández

Oviedo, 2015

UNIVERSIDAD DE OVIEDO

DEPARTAMENTO DE PSICOLOGÍA

Programa de Doctorado en Psicología regulado
por el Real Decreto 1393/2007

TESIS DOCTORAL

HABILIDADES DE AUTORREGULACIÓN Y EVALUACIÓN ON-LINE DEL PROCESO DE RESOLUCIÓN DE PROBLEMAS MATEMÁTICOS

Self-regulation Skills and On-line Assessment of the Process Involved in Solving Mathematical Problems

Autora: Trinidad García Fernández

Directores: Julio Antonio González García

Paloma González Castro

Oviedo, 2015

RESUMEN DEL CONTENIDO DE TESIS DOCTORAL

1.- Título de la Tesis	
Español/Otro Idioma: HABILIDADES DE AUTORREGULACIÓN Y EVALUACIÓN ON-LINE DEL PROCESO DE RESOLUCIÓN DE PROBLEMAS MATEMÁTICOS	Inglés: SELF-REGULATION SKILLS AND ON-LINE ASSESSMENT OF THE PROCESS INVOLVED IN SOLVING MATHEMATICAL PROBLEMS

2.- Autor	
Nombre: TRINIDAD GARCÍA FERNÁNDEZ	DNI/Pasaporte/NIE:
Programa de Doctorado: PSICOLOGÍA, R.D. 1393/2007	
Órgano responsable: UNIVERSIDAD DE OVIEDO	

RESUMEN (en español)

La resolución de problemas matemáticos es muy frecuente en la edad escolar. Sin embargo, se trata de una actividad compleja, cuya correcta ejecución depende de múltiples factores. En este sentido, la evidencia científica sugiere que, para ser exitosos, los estudiantes necesitan, entre otras cosas, ser autorregulados en su actividad de resolución de problemas, ejerciendo un control eficaz sobre sus procesos cognitivos, afectivos y conductuales. No obstante, estudios previos señalan que estudiantes en diferentes etapas educativas muestran a menudo dificultades para gestionar estos aspectos. Una evidencia a favor de esta afirmación procede del análisis del proceso metacognitivo llevado a cabo por los estudiantes durante estas tareas.

Este análisis ha revelado la existencia de patrones de resolución ineficaces por parte de muchos estudiantes; con escasas habilidades de planificación y de evaluación de la actividad, así como una marcada preferencia por procedimientos familiares como la ejecución de cálculos como medio para resolver los problemas. Estos patrones de resolución –denominados procesos en esta Tesis Doctoral- se traducen a menudo en un bajo rendimiento en la tarea, así como en la formulación de juicios poco realistas acerca del resultado alcanzado.

Este componente de juicios está estrechamente relacionado con la "calibración", o grado en el que la percepción de una persona sobre su ejecución se corresponde con su ejecución real, establecida ésta en función de una medida objetiva como la puntuación en un test. La calibración supone un importante mecanismo metacognitivo, y ejerce un efecto positivo sobre el rendimiento en matemáticas y resolución de problemas. No obstante, los estudiantes son a menudo imprecisos en sus juicios, mostrando un exceso de confianza. Estos sesgos de percepción tienen además a ser estables, con efectos negativos sobre el rendimiento, pero también sobre el esfuerzo, la persistencia y el interés de los estudiantes hacia la asignatura o la tarea. Esto hace preciso examinar los factores relacionados con la formación de estos juicios.

Un aspecto importante en esta Tesis Doctoral es la distinción entre los juicios realizados antes (predicciones), y después (postdicciones) de la ejecución. Los últimos han mostrado ser más fiables y precisos, puesto que se basan en el feedback procedente de la realización de la tarea, y serían informativos de mecanismos de monitorización durante la misma. Este aspecto hace a estos juicios especialmente interesantes desde el punto de vista del análisis del proceso en resolución de problemas. No obstante, el potencial explicativo del proceso sobre estos juicios no ha sido explorado hasta la fecha. Adicionalmente, pocos de los estudios realizados con el fin de establecer los determinantes de estos juicios se han hecho en Educación Primaria, ni se han realizado desde un punto de vista comprensivo, centrándose en aspectos aislados en su lugar.

En este sentido, la presente Tesis Doctoral ha tenido estos objetivos:

- Diseñar y poner a prueba un método de evaluación del proceso en resolución de problemas matemáticos.
- Delimitar el potencial explicativo del proceso en la precisión de los juicios realizados tras la ejecución de estas tareas.
- Analizar el efecto que otras variables cognitivas, afectivas, motivacionales y conductuales tienen sobre estos juicios.

Se contó con 524 estudiantes de quinto y sexto curso (10-13 años), los cuales realizaron dos problemas matemáticos verbales basados en la vida real. Los resultados indicaron la presencia de patrones de resolución de problemas poco efectivos y un bajo rendimiento en las tareas. No obstante, las estrategias de representación y organización de la información mostraron ser importantes en el éxito en resolución de problemas. Los estudiantes fueron además poco precisos en sus juicios, mostrando un exceso de confianza. Los más calibrados mostraron estrategias de planificación más eficaces en resolución de problemas, mientras que los poco calibrados tendieron a emplear estrategias basadas en ensayo y error. El rendimiento en matemáticas mostró ejercer un efecto sobre la calibración, explicando ciertas diferencias previas en planificación. La precisión de los juicios se relacionó también con: mejores habilidades de funcionamiento ejecutivo en planificación, organización, memoria y atención, y menor impulsividad; creencias, motivaciones y actitudes más positivas hacia las matemáticas; y un mayor tiempo empleado en la resolución de problemas. Finalmente, el rendimiento en matemáticas, la utilidad percibida de la asignatura y el tiempo en resolver el primer problema matemático predijeron significativamente la precisión de estos juicios, clasificando correctamente al 71% de la muestra.

Estos resultados muestran la utilidad del método de evaluación diseñado como medida del proceso de resolución de problemas matemáticos y la necesidad de considerar un amplio rango de variables cuando las diferencias en calibración quieren ser explicadas.

RESUMEN (en Inglés)

Mathematical problem solving is very common in educational settings. However, it is a complex activity which depends upon multiple factors for correct performance. In this context, much scientific evidence suggests that, to be successful, students need to be self-regulated in their problem-solving activities, thereby exercising effective control over their cognitive, affective and behavioral processes. However, previous studies have indicated that students of various educational levels often show difficulties managing these aspects of self-regulation. Evidence for this assertion has been gained through analyses of the metacognitive processes undertaken by students when carrying out such tasks.

These analyses revealed the existence of ineffective patterns of performance among students, such as poor planning and evaluation skills, and a strong preference for using familiar procedures such as performing calculations as a means to solve problems. These patterns (or "processes" in the context of this PhD-thesis) often impair task performance and can lead a student to make unrealistic judgments about the result they have obtained.

This component of self-judging the accuracy of one's performance is closely related to "calibration", or the degree to which one´s perception of performance corresponds to one´s actual performance, the latter being established according to an objective measure, such as the score on a test. Calibration is an important metacognitive mechanism, which has a positive effect on performance in math and problem solving. However, students are often inaccurate in their judgments, in most cases being overconfident. These perceptual biases also tend to be consistent, with negative effects not only on performance, but also on a student´s effort,

persistence and interest towards a subject or task. Thus, it is essential to examine various factors which can be related to the formation of one's self-judgement of results.
In this context, it is crucial to scrutinize the distinction between student judgments made before (predictions) and after (postdictions) of their task performance in mathematical problem solving. The latter have proven to be more reliable and accurate, as they are largely based on a student's "self-feedback" whilst performing the task, and are suggestive of self-monitoring mechanisms during the task. This aspect of performance judgments is particularly interesting from the viewpoint of analysing the underlying processes. However, the explanatory potential of the processes behind these judgments has not been explored to date. In addition, the very few studies aimed at establishing the determinants of these judgments have not focused on primary schools years, nor were they conducted from a wide-ranging perspective, relying instead upon isolated aspects of the components involved.
To address this apparent shortfall in the literature, this PhD-thesis had the following objectives:

- Design and test a method for evaluating the processes involved in solving mathematical problems.
- Explore the explanatory potential of these processes in judgments-accuracy after task completion.
- Analyze the effects of other cognitive, affective, motivational, and behavioural variables on these judgments.

A sample of 524 fifth and sixth grade students (10-13 years) took part in the studies, which performed two math word problems based on real-life situations. The students showed ineffective solving patterns and poor performance on tasks. However, strategies relating to how the information was represented and/or organised were important for successful problem solving. Students were inaccurate and over-confident in their judgments. Highly calibrated students showed more effective planning strategies in problem solving, while lower calibrated students used strategies based on trial and error. Performance in mathematics was shown to exert an effect on calibration, thereby explaining some previously reported differences in planning. Judgments accuracy was also associated with: superior executive functioning skills in planning, organization, memory and attention, and less impulsivity; more positive beliefs, motivations and attitudes towards mathematics; and increased time performing the problems. Finally, mathematics achievement, perceived usefulness of the subject, and the time spent on the first problem, significantly predicted judgments accuracy which, in turn, classified 71% of the sample correctly.
These results show the usefulness of an evaluation method that has been designed as a measure of the processes involved in solving mathematical problems, and also the need to consider a wider range of variables when differences in calibration cannot otherwise be explained.

SR. DIRECTOR DE DEPARTAMENTO DE PSICOLOGÍA
SR. PRESIDENTE DE LA COMISIÓN ACADÉMICA DEL PROGRAMA DE DOCTORADO EN PSICOLOGÍA

Esta Tesis Doctoral se ha realizado con el apoyo de una beca predoctoral del Programa Severo Ochoa, de la Fundación para el Fomento en Asturias de la Investigación Científica Aplicada y la Tecnología-FICYT (Ref.: BP11.067), así como de un proyecto concedido por el Ministerio de Ciencia e Innovación a Don Julio Antonio González García (Ref.: EDU2010-19798)

FICYT
FUNDACION PARA EL FOMENTO EN ASTURIAS DE LA INVESTIGACION CIENTIFICA APLICADA Y LA TECNOLOGIA

GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA E INNOVACIÓN

Agradecimientos

Porque el proceso es tan o más importante que el producto final en muchas ocasiones, una vez culminada la elaboración de esta Tesis Doctoral, me gustaría dar las más que merecidas gracias a todas aquellas personas que me han acompañado durante estos años de trabajo. Todas y cada una de estas personas han hecho posible que esté hoy aquí, escribiendo mi Tesis Doctoral:

A mis directores de Tesis, los Doctores Julio Antonio González García y Paloma González Castro. Gracias por confiar y creer en mí para iniciar este proceso juntos y por haberme enseñado tanto. Gracias por darme la oportunidad de crecer y aprender dentro de un equipo de excelentes profesionales y aún mejores personas.

Al Doctor Celestino Rodríguez Pérez, ejemplo de esfuerzo y generosidad, y un gran referente para mí. Gracias por ser uno de mis mayores apoyos ahora y siempre.

Al resto de miembros del Grupo de Investigación ADIR. Todos y cada uno de vosotros habéis contribuido para que este trabajo saliera adelante. Con vuestros consejos, vuestra ayuda, vuestros proyectos y vuestro apoyo habéis hecho que creciera en mi la ilusión y el gusto por el trabajo bien hecho.

I would also like to thank Doctors Mark Torrance and Lucy Betts, from Nottingham Trent University (United Kingdom), and Evelyn H. Kroesbergen, from Utrecht University (The Netherlands). I greatly appreciate you accepted my visit, your kindness and generosity. It was a pleasure for me to work with you, and I hope we will keep on collaborating in future projects. I also thank Dr. Steve Loew for his kind assistance on the revision of some parts of the articles that are presented in this Dissertation. Gracias, amigo.

A mi familia y amigos. Os agradezco enormemente vuestro apoyo durante estos años. Gracias por vuestra paciencia, por entenderme, por aceptar mis ausencias y por poner un toque de locura y espontaneidad en mi vida. Quisiera agradecer especialmente a mi madre su esfuerzo y dedicación. Gracias, madre, por darlo todo y no pedir nada a cambio. A Miguel Ángel, mi pareja, mi mejor amigo y compañero de viaje. Gracias por tu ayuda, y por transmitirme tu optimismo y alegría. Espero que allá donde nos lleve el destino nunca perdamos la ilusión. A Sole, Miguel y Belén, muchas gracias por vuestra compañía, vuestros consejos y complicidad.

Finalmente, quisiera expresar mi más sincero agradecimiento a los estudiantes, familias y responsables de los centro escolares que participaron en los trabajos que forman parte de esta Tesis Doctoral. Sin vosotros nada de esto habría sido posible.

Listado de trabajos originales

Publicaciones

García, T., Betts, L., González-Castro, P., González-Pienda, J. A. & Rodríguez, C. (In press). On-line assessment of the process involved in Maths problem-solving in fifth and sixth grade students: self-regulation and achievement. *Revista Latinoamericana de Investigación en Matemática Educativa.*

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),* 47-52.

García, T., Rodríguez, C., González-Castro, P., Álvarez-García, D. & González-Pienda, J. A. (In press). Metacognición y funcionamiento ejecutivo en Educación Primaria [Metacognition and executive functioning in Elementary School]. *Anales de Psicología.*

Trabajos Complementarios

García, T., Rodríguez, C., González-Castro, P., González-Pienda, J. A. & Torrance, M. (2013). Elementary students´ metacognitive process and post-performance. *Submitted for publication.*

García, T., Cueli, M., Rodríguez, C., Krawec, J. & González-Castro, P. (2014). Metacognitive knowledge and skills in students with Deep approach to learning. Evidence from mathematical problem solving. *Submitted for publication.*

García, T., Kroesbergen, E. H., Rodríguez, C., González-Castro, P. & González-Pienda, J. A. (2014). Analysis of the factors involved in making post-performance judgments in mathematical problem solving. *Submitted for publication.*

Índice

Resumen

Resolver problemas matemáticos es considerado un aspecto clave en matemáticas, lo cual explica el gran volumen de investigación realizada en este ámbito. Dada su relevancia en el desarrollo de habilidades básicas de análisis, comprensión, razonamiento y aplicación, la resolución de problemas matemáticos es entendida como una competencia central que los individuos deben adquirir para adaptarse al mundo actual. Sin embargo, los estudios previos en este sentido sugieren que los estudiantes en diferentes etapas educativas presentan a menudo dificultades para realizar este tipo de tareas, a la vez que reconocen que los intentos realizados hasta ahora por enseñar a los estudiantes estrategias eficaces de resolución de problemas no han sido especialmente fructíferos (Blanco, Guerrero, y Caballero, 2013; Castro, 2008; Santos, 2007). Los resultados publicados referentes al último estudio PISA (OCDE, 2014), así como los procedentes de ejercicios anteriores, constatan esta realidad.

Obviamente, existen una infinidad de factores que pueden explicar estos resultados. Abarcar cada uno de ellos es simplemente inviable. Entre estos posibles factores, el uso de estrategias basadas en el Aprendizaje Autorregulado ha demostrado ejercer un importante papel sobre la resolución de problemas matemáticos (Lazakidou y Retalis, 2010; Montague, 2008; Pennequin, Sorel, Nanty, y Fontaine, 2010; Pereis, Dignath, y Schmitz, 2009; Schmitz y Perel, 2012). Estos estudios enfatizan la relevancia de los procesos cognitivos, afectivos y motivacionales presentes durante la realización de este tipo de tareas, y el grado de control que los estudiantes ejercen sobre ellos. Desde la perspectiva del control de los procesos cognitivos implicados en la resolución de problemas matemáticos, la evidencia sugiere sin embargo que los estudiantes muestran generalmente pobres habilidades metacognitivas durante la realización de estas tareas. Esto tiene su impacto no solamente en el resultado final en la tarea, sino también en las actitudes, motivación, y las percepciones o juicios de los estudiantes sobre su propia ejecución (Cleary y Chen, 2009; Kramarski y Gutman, 2006; Montague, 2008; Montague, Enders, y Dietz, 2011; Pennequin et al, 2010; Pereis et al., 2009; Veenman, 2005).

Un importante mecanismo metacognitivo relacionado con estos juicios es la denominada *“calibración”* o el grado en que los juicios de una persona sobre su ejecución se corresponden con su ejecución real en una tarea, determinada ésta en base a una medida

objetiva como la puntuación en un test o la calificación en un examen (Hacker, Bol, & Keener, 2008a). Este tipo de juicios pueden ser realizados antes –*predicciones*– o después –*postdicciones*– de la ejecución de la tarea, y han demostrado ser importantes predictores tanto del rendimiento en matemáticas como en la resolución de problemas. Sin embargo, estudios previos han demostrado que los estudiantes tienden a ser poco calibrados, mostrando una tendencia hace el exceso de confianza, lo que tiene un efecto negativo en el esfuerzo, persistencia e interés de los estudiantes hacia la asignatura o la tarea (Desoete y Roeyers, 2006; Jacobse y Harskamp, 2012; Özsoy 2012; Rinne y Mazzocco, 2014).

En este sentido, aunque ambos procesos –predicciones y postdicciones- son entendidos como parte importante de la calibración, los juicios realizados tras la ejecución de la tarea serían más precisos y fiables, e informarían acerca de los procesos de monitorización durante la tarea (Hacker et al. 2008a; McCormick, 2003). Este segundo tipo de juicios serían por tanto susceptibles de ser examinados desde el punto de vista del proceso metacognitivo llevado a cabo por los estudiantes durante la resolución de problemas matemáticos. En este sentido, si bien el origen de estos juicios ha sido ampliamente estudiado, pocos trabajos han sido llevados a cabo en Educación Primaria, centrándose la mayoría de ellos en niveles educativos posteriores. Estos estudios no han analizado tampoco el potencial explicativo del proceso sobre la calibración. Finalmente, pocos trabajos han abordado el estudio de los determinantes de la calibración desde un punto de vista comprensivo, centrándose en su lugar en examinar aspectos aislados.

Dada esta realidad, la presente Tesis Doctoral ha tenido como objetivo responder a las siguientes cuestiones:

1. ¿Cómo se puede evaluar el proceso y qué puede decir éste acerca de las estrategias metacognitivas implicadas en la resolución de problemas matemáticos?
2. ¿Puede el proceso informar acerca de los mecanismos implicados en la realización de juicios tras la ejecución de este tipo de tareas?
3. ¿Hay otras variables afectivas, motivacionales, cognitivas o conductuales que puedan explicar las diferencias en estos juicios?

Con el fin de responder a estas cuestiones, esta Tesis Doctoral se ha concretado en la elaboración de tres publicaciones en revista de conocido prestigio en el área de Psicología y Educación, así como tres trabajos complementarios. Los resultados obtenidos en los diferentes trabajos se describen a continuación, en relación a cada una de las cuestiones planteadas.

Para la realización de estos estudios se contó con una muestra de 524 estudiantes de quinto y sexto curso, pertenecientes a 12 centros educativos en Asturias (Norte de España), los cuales resolvieron dos problemas matemáticos verbales basados en situaciones de la vida diaria. Estos problemas fueron extraídos del libro *"Comprender y Resolver Problemas"* traducido al castellano (Whimbey y Lochhead, 1993). Se eligió este tipo de problemas ya que no contienen expresiones algebraicas u otros medios que puedan guiar a los estudiantes en su resolución, ni tampoco requieren el conocimiento de contenidos o conceptos matemáticos específicos. Por lo tanto, se entiende que el éxito o fracaso de los estudiantes en estas tareas respondería a aspectos más estratégicos (Harskamp y Suhre, 2006). Una descripción más detallada de los problemas matemáticos empleados está disponible en los propios estudios.

Primera Cuestión

¿Cómo se puede evaluar el proceso y qué puede decir éste acerca de las estrategias metacognitivas empleadas por los estudiantes durante la resolución de problemas matemáticos?

Esta primera cuestión es tratada en el *Primer Artículo* que forma parte de esta Tesis Doctoral: **On-line assessment of the process involved in Maths problem-solving in fifth and sixth grade students: self-regulation and achievement.**

En este primer estudio se presentó y puso a prueba la técnica de evaluación del proceso diseñada, basada en de la adaptación de la técnica de la *Triple Tarea* (Kellog, 1987; Olive, Kellog, y Piolat, 2001; Olive y Piolat, 2002), inicialmente propuesta para la evaluación de los procesos implicados en tareas de composición escrita. Esta técnica es una de los denominadas *"on-line"* (Veenman, 2011), definidas como las medidas tomadas de forma concurrente a la ejecución de la tarea. A la versión de la Triple Tarea empleada, basada en el modelo de Zimmerman (2000, 2008) y el modelo IDEAL de Resolución de Problemas de Bransford y Stein (1993), se le ha dado el nombre de *Triple Task Procedure in Mathematics (TTPM).*

En este primer estudio se describió el perfil del proceso metacognitivo mostrado por una muestra de 510 estudiantes de quinto y sexto curso mientras resolvían dos problemas matemáticos, y se analizó la relación entre dicho proceso y el resultado – o producto- en estas tareas.

Se realizaron análisis separados para cada problema matemático. El análisis del proceso seguido por los estudiantes mostró en primer lugar la existencia de una gran variabilidad entre los procesos mostrados por los estudiantes, lo que se reflejó en unas elevadas desviaciones típicas. Los resultados indicaron así mismo la presencia de unas estrategias de planificación ineficaces y una ausencia de mecanismos de evaluación por parte del grupo general de estudiantes, así como una preferencia hacia el empleo de procedimientos familiares como la realización de cálculos como medio para resolver los problemas. El análisis de las diferencias entre los grupos con diferente rendimiento en las tareas (Éxito vs. Fracaso) reveló los sub-procesos implicados en la planificación, especialmente el empleo de estrategias de organización y representación de la información, como determinantes importantes en el éxito de los estudiantes, ejerciendo un efecto mayor en el segundo problema, presumiblemente debido a un incremente en la dificultad de la tarea.

Si bien la muestra con la que se contó en esta Tesis Doctoral fue de 524 estudiantes, la gran variabilidad encontrada en las variables del proceso hizo necesario eliminar a 14 participantes en este estudio, debido a la presencia de puntuaciones muy extremas en las mismas. Este aspecto fue subsanado en el segundo estudio realizado pruebas no paramétricas.

Segunda Cuestión

¿Puede el proceso informar sobre los mecanismos implicados en la realización de juicios tras la ejecución de este tipo de tareas?

El *Primer Trabajo Complementario* incluido en esta Tesis Doctoral trata de responder a esta cuestión: **Elementary students´ metacognitive process and post-performance calibration in mathematical problem solving.**

Este es el estudio más amplio de los realizados. En él se contó con la muestra completa de 524 estudiantes de quinto y sexto curso. En primer lugar, se analizó la precisión de los juicios post-ejecución de los estudiantes – alta o baja calibración; exceso o falta de confianza-, así como la estabilidad de estos juicios y de la ejecución real de los estudiantes en los dos problemas matemáticos. A continuación, se analizaron las diferencias en el proceso metacognitivo mostrado por los estudiantes con diferente precisión en sus juicios (Precisos vs. Imprecisos). Se examinó también la posible emergencia de diferentes patrones de calibración y de diferencias en el propio proceso metacognitivo en función de diferentes niveles rendimiento académico en matemáticas (Bajo-Medio-Alto) y el curso (Quinto vs. Sexto). Nuevamente, se realizaron análisis separados para cada problema matemático.

Los resultados indicaron la presencia de unas escasas habilidades de calibración en los estudiantes, los cuales mostraron una fuerte tendencia hacia el exceso de confianza en sus juicos. Adicionalmente, los estudiantes mostraron una importante estabilidad tanto en sus juicios como en la ejecución en los problemas matemáticos. En cuanto a las diferencias en el proceso metacognitivo, los estudiantes imprecisos en sus juicios mostraron emplear estrategias de organización o representación de la información en menor medida que sus compañeros en el otro grupo, mostrando en su lugar patrones de resolución relacionados con el uso de mecanismos de "ensayo y error". Finalmente, los diferentes niveles de rendimiento en matemáticas dieron lugar a diferentes patrones de calibración (la precisión de los juicios aumentó progresivamente junto con el nivel de rendimiento), así como en el propio proceso, fundamentalmente en lo referente al uso de estrategias de organización y representación de la información (mostrando también una relación positiva entre ambos). La variable curso no dio lugar a un patrón de resultados claro en este sentido.

Tercera Cuestión

¿Hay otras variables afectivas, motivacionales, cognitivas o conductuales que puedan explicar las diferencias en estos juicios?

Una vez mostrada la relación entre el proceso metacognitivo y la calibración, el objetivo del resto de estudios realizados fue realizar un análisis más profundo de los factores que pudieran determinar las habilidades de calibración de los estudiantes, tratando de identificar variables adicionales que pudieran predecir significativamente las diferencias en calibración encontradas. Con el fin de responder a esta cuestión se realizaron cuatro trabajos.

Dada la caracterización de la calibración como un importante proceso metacognitivo, los dos primeros estudios se centran en examinar la potencial utilidad de las habilidades de funcionamiento ejecutivo y de un enfoque profundo de aprendizaje como posibles predictores de la calibración post-ejecución. Para ello se partió del análisis de la relación entre estas variables y la metacognición (componentes de conocimiento y habilidades metacognitivas, las últimas definidas como la aplicación de este conocimiento en situaciones de aprendizaje). Finalmente, el último trabajo de esta Tesis Doctoral examina el grado en que las variables que mostraron ser significativas en los dos estudios anteriores, junto con varios componentes afectivo-motivacionales relacionados con las matemáticas, el rendimiento en la asignatura, y ciertas características de la tarea, predicen las diferencias en calibración post-ejecución.

Previamente al análisis de la relación entre metacognición y funcionamiento ejecutivo se realizó un primer estudio, con el fin de conocer la potencial utilidad de un instrumento de evaluación de las funciones ejecutivas, basado en las informaciones proporcionadas por parte de las familias (la Escala *Behavior Rating Inventory of Executive Functions-BRIEF*: Gioia, Isquith, Guy, y Kenworthy, 2000), en la evaluación de las funciones ejecutivas en niños y adolescentes. Este instrumento no se encontraba disponible en idioma español en el momento de la realización del trabajo, con lo cual el principal objetivo de ese primer estudio fue analizar sus características psicométricas en una muestra de población española. Este trabajo se realizó con una muestra clínica de estudiantes, dado el importante papel de las funciones ejecutivas en diferentes dificultades del aprendizaje y desórdenes en la infancia y adolescencia. Una breve descripción de estos cuatro estudios se muestra a continuación:

En primer lugar, en el *Segundo Artículo* de esta Tesis Doctoral, se presentan las propiedades psicométricas de la escala BRIEF (Gioia et al., 2000) en muestra clínica española: **Psychometric characteristics of the BRIEF scale for the assessment of executive functions in Spanish clinical population.**

En este estudio se empleó una muestra clínica de 125 participantes de 5 a 18 años. Se analizó la estructura interna y fiabilidad de las puntuaciones de la escala traducida al español, así como su relación con otras medidas comportamentales a través del análisis de sus correlaciones con la escala de *Evaluación del Déficit de Atención con Hiperactividad- EDAH* (Farré y Narbona, 1987). Los resultados se compararon con los del estudio de validación original. Los datos mostraron la presencia de una misma estructura interna de las puntuaciones, así como una aceptable consistencia interna y correlaciones estadísticamente significativas con los componentes de Déficit de Atención e Hiperactividad de la escala EDAH. Este estudio aportó evidencia preliminar sobre la utilidad de la escala BRIEF en contextos culturales diferentes al originario, concretamente en población clínica española. No obstante, ciertas limitaciones en el instrumento, principalmente su longitud, condujeron al diseño de un nuevo instrumento de evaluación de las funciones ejecutivas por parte del propio equipo de investigación de la Universidad de Oviedo. Concretamente, la escala de *Evaluación del Funcionamiento Ejecutivo- EFE* en sus versiones para familias (García, Álvarez-García, Cueli, González-Castro, y Álvarez, 2013) y profesorado (García, Álvarez-García, González-Castro, Álvarez, y Segurola, 2014a). Estas escalas se encuentran en estos momentos en fase de validación. En el Anexo 1 de presenta una muestra de los ítems evaluados por las mismas.

En segundo lugar, el análisis de la relación entre funciones ejecutivas y metacognición se trata en el *Tercer Artículo* de esta Tesis Doctoral: **Metacognición y funcionamiento ejecutivo en Educación Primaria.**

Si bien los estudios que relacionan ambos aspectos son prácticamente inexistentes hasta la fecha, la propia caracterización de las funciones ejecutivas como mecanismos de control de orden superior hace pensar que ambos componentes deben guardar algún tipo de asociación. Si esta asociación fuese encontrada, las funciones ejecutivas serían por tanto buenas candidatas a explicar parte de la variabilidad en los juicios realizados por los estudiantes, sobre todo teniendo en cuenta la importante relación entre funcionamiento ejecutivo y rendimiento en matemáticas evidenciada en la literatura.

Este estudio se analizó la relación entre metacognición y funciones ejecutivas, partiendo del conocimiento metacognitivo como base para el desarrollo de las habilidades metacognitivas. De este modo, este estudio analizó inicialmente si los estudiantes con diferentes niveles de conocimiento metacognitivo (Alto vs. Bajo) presentaban diferencias en sus habilidades metacognitivas, definidas como la aplicación de este conocimiento en situaciones de aprendizaje, específicamente en las fases de Planificación, Ejecución y Evaluación del Aprendizaje Autorregulado. El componente de conocimiento metacognitivo fue evaluado mediante un test de reconocimiento de estrategias, concretamente el test de *Conocimiento de Estrategias de Aprendizaje- CEA* (Núñez et al., 2011; Rosário, Mourão, Núñez, González-Pienda, y Solano, 2006), mientras que las habilidades metacognitivas se evaluaron mediante auto-informe, empleando el *Inventario de Procesos de Autorregulación del Aprendizaje- IPAA* (Rosário et al., 2010). Como medida de las funciones ejecutivas se empleó la escala de *Evaluación del Funcionamiento Ejecutivo- EFE* en sus versiones para familias (García et al., 2013) y profesorado (García et al., 2014a), previamente comentada.

Los resultados mostraron que los estudiantes con alto conocimiento metacognitivo indicaron emplear sus habilidades metacognitivas, tales como pensar en los pasos a seguir o los materiales necesarios para realizar una tarea, comparar el resultado final en la tarea con el esperado a su comienzo, o evaluar el propio progreso en la tarea, en mayor medida que los estudiantes con bajo conocimiento metacognitivo. Estas diferencias fueron estadísticamente significativas en las fases de Planificación y Ejecución del aprendizaje. También se encontraron diferencias estadísticamente significativas entre los grupos en las habilidades de atención sostenida, focalización de la atención, memoria funcional y fundamentalmente en planificación. Estos resultados fueron consistentes empleando ambas escalas de funciones ejecutivas. De este modo, tanto familias como profesorado informaron acerca de unos mejores niveles de funcionamiento ejecutivo en los estudiantes con alto conocimiento metacognitivo. La voluntariedad en el estudio y el requisito de contar con las informaciones de familias y profesorado hizo que la muestra se redujera a 114 participantes.

En tercer lugar, la relación entre un enfoque profundo de aprendizaje y la metacognición se analiza en el *Segundo Trabajo Complementario* que forma parte de esta Tesis Doctoral: **Metacognitive knowledge and skills in students with Deep approach to learning. Evidence from mathematical problem solving.**

Partiendo de los resultados mostrados en la literatura previa sobre el tema y que sugieren la existencia de una asociación entre el empleo de un enfoque profundo de aprendizaje y el uso de estrategias metacognitivas en resolución de problemas, este estudio tuvo por objetivo analizar la relación entre este enfoque profundo y los componentes metacognitivos de conocimiento y habilidades. Para ello se dividió la muestra inicial de 524 estudiantes en tres grupos en función del grado en que presentaban un enfoque profundo de aprendizaje (Bajo-Medio-Alto) y se analizaron las diferencias en ambos componentes de la metacognición entre estos grupos. Para la evaluación del enfoque profundo de aprendizaje se empleó el *Inventario de Procesos de Estudio- IPE* (Núñez et al., 2011; Rosário et al., 2013). Nuevamente, el conocimiento metacognitivo fue evaluado mediante la anterior prueba de reconocimiento de estrategias, mientras que las habilidades metacognitivas fueron evaluadas en esta ocasión a través de la media del proceso diseñada, en situaciones de resolución de problemas matemáticos. Se llevaron a cabo análisis separados para cada problema.

Los resultados indicaron la existencia de diferencias estadísticamente significativas entre los grupos en la variable de conocimiento metacognitivo. En este sentido, un mayor empleo de un enfoque profundo se relacionó con un mejor conocimiento metacognitivo general. No obstante, no se encontró un patrón claro de diferencias en cuanto al proceso metacognitivo, si bien se encontraron diferencias aisladas en el segundo de los problemas, fundamentalmente en el sub-proceso de revisión. En concreto, un alto empleo de un enfoque profundo se relacionó con un mayor empleo de estrategias de revisión. Estos resultados son discutidos desde el punto de vista del papel del conocimiento metacognitivo como base para el desarrollo de las habilidades metacognitivas, así como teniendo en cuenta la distinción entre pruebas de evaluación "off-line" y "on-line" y el grado de correspondencia entre las mismas. Las primeras son medidas tomadas antes o después de la tarea, generalmente cuestionarios, mientras que las últimas son registradas simultáneamente a la ejecución –Triple Task Procedure in Mathematics en este caso-.

Este patrón de resultados, junto con la especificidad de la variable a predecir (es decir, juicios post-ejecución en resolución de problemas matemáticos) y el argumento sostenido por numerosos autores sobre el hecho de que no un único enfoque de aprendizaje sino varios pueden estar presentes en uno individuo, variando en función de factores personales, de la tarea y del contexto, llevaron a desestimar el enfoque profundo de aprendizaje como posible variable predictora de las diferencias en calibración.

Finalmente, en el *Tercer Trabajo Complementario* incluido en esta Tesis Doctoral, se analiza el valor predictivo que las funciones ejecutivas, junto con variables adicionales que la literatura ha mostrado como relevantes en relación a la calibración, podrían tener sobre la precisión de los juicios post-ejecución: **Analysis of the factors involved in making post-performance judgments in mathematical problem solving.**

De este modo, este estudio analizó el grado en que las siguientes variables podrían predecir la precisión de estos juicios: funciones ejecutivas, varios componentes afectivo-motivacionales relacionados con las matemáticas, el rendimiento académico en la asignatura, y ciertas características de la tarea (dificultad percibida y tiempo total empleado en resolver los problemas). Previo a establecer su valor predictivo, se analizaron las diferencias en estas variables entre los grupos con diferente precisión en sus juicios (Precisos vs. Imprecisos). Los estudiantes fueron asignados a estos dos grupos en base a la precisión de sus juicios en los dos problemas matemáticos para evitar respuestas azarosas. Aquellos que proporcionaron juicios certeros solamente en uno de los problemas fueron por tanto excluidos de los análisis, razón por la cual se empleó una muestra de 188 estudiantes en este estudio. Las funciones ejecutivas fueron evaluadas mediante la escala EFE (García et al., 2013) anteriormente comentada, solamente en su forma para familias, con el fin de no perder excesivo tamaño muestral.

En cuanto a las diferencias entre los grupos, los resultados indicaron que los estudiantes precisos en sus juicios mostraron unas mejores actitudes, motivaciones y creencias relacionadas con las matemáticas, así como un mejor rendimiento en la asignatura. Este grupo fue evaluado por las familias como significativamente menos impulsivos, así como con unas mejores habilidades atencionales, memorísticas y de planificación y organización. Los estudiantes mejor calibrados mostraron así mismo emplear más tiempo en la resolución de los problemas, fundamentalmente en el primero de ellos. Finalmente, en cuanto al valor predictivo de estas variables, el rendimiento en matemáticas, la utilidad percibida o valor de la asignatura y el tiempo total empleado en resolver el primer problema predijeron significativamente la precisión de estos juicios, clasificando correctamente al 71.3% de los estudiantes en sus respectivos grupos.

Los resultados obtenidos a lo largo de los diferentes estudios evidencian la utilidad de la medida diseñada -*Tripe Task Procedure in Mathematics (TTPM)*- en la evaluación del proceso metacognitivo relacionado con la resolución de problemas matemáticos, así como el potencial valor explicativo del proceso sobre la calibración post-ejecución. Ésta se trata, no obstante, de un proceso complejo en el que intervienen múltiples variables, las cuales es necesario abordar a diferentes niveles y de una forma comprensiva.

Summary

Problem solving has always been regarded as a key point of mathematics, which explains the amount of research conducted on this issue. Because of its relevance enhancing the development of basic skills such as analysis, comprehension, reasoning and application, problem solving is considered a core competence individuals need to acquire for todays world. Nevertheless, previous studies have shown that students at different educational stages usually struggle with mathematics problems while recognizing that the attempts to teach students general problem solving strategies have been mostly unsuccessful (Blanco, Guerrero, & Caballero, 2013; Castro & Santos, 2008). The results of the Programme for International Students Assessment (PISA) of 2014 and previous years have highlighted this reality.

Obviously, there is a wide range of variables that may explain these results. Thus, addressing all of them is simply not feasible. Among the possible variables related to successful mathematical problem solving, strategy use based on Self-regulated Learning (SRL) has demonstrated to be an important determining factor (Lazakidou & Retalis, 2010; Montague, 2008; Pennequin, Sorel, Nanty, & Fontaine, 2010; Pereis, Dignath, & Schmitz, 2009; Schmitz & Perel, 2012). These studies emphasize the relevance of the cognitive, affective and motivational processes involved in solving mathematical problems and the control that students exercise over them. However, focusing on the cognitive sphere, previous research suggests that students tend to show poor metacognitive skills while involved in mathematical problem-solving situations. This has a negative impact not only on task performance but also on student attitudes, motivation, and perceptions or judgments of performance (Cleary & Chen, 2009; Kramarski & Gutman, 2006; Montague, 2008; Montague, Enders, & Dietz, 2011; Pennequin et al, 2010; Pereis et al., 2009; Veenman, 2005).

An important metacognitive mechanism related to these judgments is the so-called *"calibration",* referred to as the degree to which one´s judgments of performance corresponds to one´s actual performance, the latter determined on the basis of an objective measure such as a score on a test (Hacker, Bol, & Keener, 2008a). These judgments may be made before – "predictions"– or after –"postdictions"– task completion, and their accuracy has demonstrated to be an important predictor of mathematics achievement and problem-solving performance.

However, student have demonstrated to be poorly calibrated in previous studies, showing a tendency towards over-confidence, which has adverse effects on students´ effort, persistence and interests towards a subject or task (Desoete & Roeyers, 2006; Jacobse & Harskamp, 2012; Özsoy, 2012; Rinne & Mazzocco, 2014).

Although both components –predictions and postdictions– are understood as important parts of calibration, the latter –also called post-performance judgments– would be more accurate and reliable as they inform about monitoring mechanisms during task performance (Hacker et al., 2008a, McComick, 2003). Therefore, this sort of judgments can be examined from the perspective of the metacognitive process shown by students while solving mathematical problems. Although the origin of these judgments has been substantially studied, the research conducted on this issue presents some limitations: first, the number of references regarding calibration in Elementary School stages is rather scarce, as most of the studies have been carried out with college or undergraduate samples; second, the potential explanatory power of the process over this kind of judgments has not been analyzed to date; finally, just a few studies have addressed the analysis of the determining factors of calibration from a comprehensive perspective, focusing on analyzing a small range of variables instead.

Within this context, the present PhD-Thesis asked the following questions:

1. How can the process be assessed, and what does this process reveal about the metacognitive mechanisms involved in solving mathematical problems?
2. Can the process provide information about the mechanisms involved in making post-performance judgments in these tasks?
3. Are there other affective, motivational, cognitive or behavioral variables that may be explaining differences in these judgments?

In order to answer these questions, this PhD-Thesis resulted in the elaboration of three articles accepted and published in recognized journals in the field of Psychology and Education, as well as three complementary papers. The results obtained are described below, in relation to each posed question.

Five hundred and twenty-four fifth- and sixth-grade students from 12 schools in Asturias (Northern Spain) took part in the studies. They completed two mathematical problems based on daily-life situations. These problems were taken from the book "*Problem-solving and comprehension"* (Whimbey and Lochhead, 1993) and translated into Spanish. This specific type of problems was chosen as they are stated in everyday language and deal with fairly realistic situations. These tasks do not contain algebraic expressions or other mathematical means to guide students. As they do not involve specific mathematical contents, students´ failure in their resolution is assumed to be due to more strategic processes (Haskamp & Suhre, 2006). A more detailed description of the problems used in the studies is provided in the studies.

First question

How can the process be assessed, and what does this process reveal about the metacognitive mechanisms involved in solving mathematical problems?

This question was addressed in the *First Article* of this Dissertation: **On-line assessment of the process involved in Maths problem-solving in fifth and sixth grade students: self-regulation and achievement.**

The *Triple Task Procedure in Mathematics –TTPM–* was used as a measure of the process. The first study of this Dissertation presents this tool and examines its usefulness in the assessment of the metacognitive processes involved in mathematical problem solving. This measure is an adaptation of the Triple Task technique (Kellog, 19787; Olive, Kellog, & Piolat, 2001; Olive & Piolat, 2002), initially proposed to study the processes involved in composition writing. This is one of the so-called "on-line" assessment tools (Veenman, 2011), referred to as those measures taken concurrently with performance. This new version of the Triple Task is based on the Self-Regulation model of Zimmerman (2000, 2008) and the IDEAL problem-solving model of Bransford and Stein (1993).

In this first study, the metacognitive process shown by 510 fifth- and sixth- grade students while solving two mathematical problems was analyzed, as well as the relationship between this process and the result of their performance *–or product–* in these tasks. Separate analyses were conducted for each mathematics problem. Results evidenced the existence of a great variability in the process showed by students, as the high standard deviations found suggested. This process was characterized by the application of ineffective planning strategies, a preference to use familiar procedures, such as doing calculations, and a lack of evaluation mechanisms in the general simple. The analysis of the differences between groups with different performance in the problems (Success vs. Failure) revealed however the sub-process involved in planning, mainly the use of information representation and organization strategies, as important determining factor of students´ success. These differences were more evident in the second problem, possibly due to an increase in problem difficulty.

Although the initial sample was made up by 524 students, the high variability observed in the process shown by students made it necessary to exclude 14 participants from the statistical analyses. This aspect was addressed in the second study by conducting non-parametric analyses.

Second question

Can the process provide information about the mechanisms involved in making post-performance judgments in these tasks?

This second question was addressed in the *First Complementary Paper* included in the present Dissertation: **Elementary students´ metacognitive process and post-performance calibration in mathematical problem solving.**

This is the most extensive study of this Dissertation. The total sample of 524 fifth- and sixth- grade students took part in this study. First of all, an estimation of post-performance calibration accuracy was calculated –*how accurate are them in their judgments? Did they show a tendency towards over- or under- confidence?*– and the stability of both post-performance judgments and actual performance was analyzed. Then, differences in the metacognitive process between students with different calibration accuracy (Accurate vs. Inaccurate) were examined. The possible emergence of different patterns of calibration and differences in the process as function of mathematics achievement (Low-Medium-High) and grade (Fifth vs. Sixth) was also examined. Separate analyses were conducted for each mathematical problem.

The results indicated the presence of low levels of calibration accuracy among students, showing a strong tendency towards over-confidence. Additionally, students showed a high stability in their judgment and actual performance. As for the differences in the metacognitive process, inaccurate students used strategies of information representation and organization to a lesser extent than their peers in the accurate group, showing resolution patterns coherent with a use of "trial and error" mechanisms instead. Finally, different levels in mathematics achievement led to different calibration patterns. Specifically, judgments accuracy progressively improved as mathematics achievement level increased. Differences in the metacognitive process itself were also found, mainly regarding the use of information representation and organization strategies, being confirmed a positive relationship between them. Grade level did not generate a specific pattern of differences in this sense.

Third question

Are there other affective, motivational, cognitive or behavioural variables that may be explaining differences in post-performance judgments?

Once the existence of the relationship between process and calibration was examined, the main aim of the remaining studies was to conduct a deep analysis of the variables that may explain differences in calibration accuracy, trying to identify additional components that could significantly predict differences in calibration. Four studies were conducted in this sense.

Given the characterization of calibration as an important metacognitive mechanism, two of these studies focused on examining the potential usefulness of executive functions and a deep approach to learning predicting these judgments, starting from the analysis of the relationship between these variables and metacognition (components of knowledge and skills, the latter defined as application of this knowledge in learning situations). Finally, the last study included in this Dissertation analyzed the extent to which the variables that showed to be significant in these previous studies, along with affective-motivational variables related to mathematics, achievement in the subject, and task characteristics, may predict differences in post-performance calibration.

Prior to analyzing the relationship between executive functions and metacognition, a preliminary study was conducted. This study was aimed at analyzing the potential utility of a questionnaire, based on family reports, for the assessment of executive functions in children and adolescents (*Behavior Rating Inventory of Executive Functions-BRIEF*: Gioia, Isquith, Guy, & Kenworthy, 2000). This instrument was not available in Spanish language when the study was conducted. Thus, the main aim of this study was to analyze its psychometric characteristics in a sample of Spanish students. This study was carried out with a clinical sample, given the relevance of executive functions in learning disabilities and multiple clinical disorders in childhood and adolescence.

A brief description of these studies and the results obtained is provided below:

First, the *Second Article* of this Dissertation focused on analyzing the psychometric properties of the BRIEF scale (Gioia et al., 2000) in a clinical Spanish sample of students: **Psychometric characteristics of the BRIEF scale for the assessment of executive functions in Spanish clinical population.**

A sample of 125 participants aged 5-18 years took part in this study. Internal structure and reliability of the translated scale were analyzed, as well as its relationship with other behavioral measures through the analysis of their correlations with the *Assessment of Attention Deficit Hyperactivity Disorder Scale* (Evaluación del Déficit de Atención con Hiperactividad-EDAH, in Spanish: Farré & Narbona, 1997). The results were compared with those from the original validation study. The data revealed the presence of the same internal structure, as well as acceptable internal consistency and significant correlations with the Attention Deficit and Hyperactivity components of the EDAH scale. This study provides preliminary evidence of the utility of the BRIEF scale in cultural contexts different from the original, particularly in Spanish clinical population. However, some limitations of this assessment tool, mainly its excessive length (i.e., it is made up of 86 items), led the research group at the University of Oviedo to desing a new instrument for the assessment of executive functions: the *Executive Functions Scale* (Escala de Funcionamiento Ejecutivo- EFE in Spanish) in its versions for families (García, Álvarez-García, Cueli, González-Castro, & Álvarez, 2013) and teachers (García, Álvarez-García González-Castro, Álvarez, & Segurola, 2014a).

Second, the analysis of the relationship between executive functions and metacognition is addressed in the *Third Article* of this Dissertation: **Metacognition and executive functioning in Elementary School.**

While studies linking these components are practically non-existent to date, the characterization of executive functions as high-order control mechanisms suggests that both executive functions and metacognition should keep some kind of association. If this association were found, executive functions would therefore be a good candidate to explain some of the variability in the judgments made by students, especially considering the important relationship between executive functioning and performance in mathematics, evidenced in the literature.

This study examined the relationship between executive functions and metacognition focusing on metacognitive knowledge as it is suggested to be the basis for metacognitive skills development. In this sense, this study starts by analyzing if students with different metacognitive knowledge levels (Low vs. High) differed in their metacognitive skills, conceptualized as the application of this knowledge in general learning situations, specifically in the Self-Regulated Learning (SRL) phases of planning, execution and evaluation.

These two components were analyzed by means of a strategy recognition test (*Learning Strategies Knowledge Questionnaire-LSKQ*: Núñez et al., 2011; Rosário et al., 2006), and self-report (*SRL Processes Inventory-SRLPI*: Rosário et al., 2010), respectively. Differences in executive functions between the students with high and low metacognitive knowledge (High vs. Low) were also analyzed. Executive functions were assessed through the administration of the Executive Functions Scale in its versions for families (García et al., 2013) and teachers (García et al., 2014a), previously presented. These scales are now in process of validation. A sample of the items evaluated by these scales is provided in Appendix 1.

The results showed that students with high metacognitive knowledge reported using metacognitive strategies, such as thinking about the steps or materials needed to perform a task, comparing their final result in a task with the result they expected to obtain at the beginning, or keeping track of their own progress during the task, more frequently than their peers with low metacognitive knowledge. These differences were statistically significant in the Planning and Execution phases of SRL. Regarding executive functions, statistically significant differences between groups in the components of sustained attention, focus, working memory and planning were found. These results were consistent using both scales of executive functions. Families and teachers reported enhanced levels of executive functioning in the group of students with high metacognitive knowledge. The voluntary nature of the study and the requirement to obtain information from both families and teachers meant the sample was reduced to 114 participants in this study.

Third, the relationship between a deep approach to learning and metacognition is analyzed in the *Second Complementary Paper* of the present Dissertation: **Metacognitive knowledge and skills in students with Deep approach to learning. Evidence from mathematical problem solving.**

Based on previous studies on the subject, which suggest the existence of an association between the use of a deep learning approach and better metacognitive strategies in problem solving, this study aimed at analyzing the relationship between this approach and the components of metacognitive knowledge and skills. In order to accomplish the objective, the sample of 524 students was divided into three groups based on the degree to which they show a deep approach to learning (Low-Medium-High), and differences among groups in metacognitive knowledge and skills were analyzed. Deep approach to learning was assessed by means of the *Processes Study Inventory-PSI* (Núñez et al., 2011; Rosário et al., 2013). Metacognitive knowledge was evaluated by means of a strategy recognition test (as in the previously described study) whereas metacognitive skills were assessed during problem solving tasks, through the administration of the process measure designed and described in Studies 1 and 2. Different analyses were conducted for each mathematical problem.

The results indicated the existence of statistically significant differences between groups in metacognitive knowledge. In this regard, the increased use of a deep approach to learning was associated with better overall metacognitive knowledge. A clear pattern of differences in metacognitive skills during problem-solving tasks was not found, although minimal differences were found in the second mathematical problem, mainly in the sub-process of revision. Specifically, a high use of a deep approach was associated with increased use of revision strategies. These results are discussed from the point of view of the role of metacognitive knowledge as a basis for the development of metacognitive skills, and the distinction between "off-line" vs. "on-line" assessment measures; defined the former as those measures taken after of before task performance –such as questionnaires–, and the latter as measures taken concurrent with task performance –Triple Task Procedure in Mathematics in this case–.

The pattern of results obtained in this study, alongside the specificity of the variable to be predicted in the last phase of this Dissertation (i.e., post-performance judgments in mathematical problem solving), and the argument supported by numerous authors that no a single but different approaches to learning should be considered within an individual –their choice depends on personal, task and context features–, led to reject a deep approach to learning as a possible predictor of differences in calibration.

Finally, the predictive value that executive functions, and other variables that literature revealed as having influence on calibration, may have over post-performance judgments accuracy was examined in the *Third Complementary Paper* included in this Dissertation:

Analysis of the factors involved in making post-performance judgments in mathematical problem solving.

This study analyzed the extent to which the following variables predicted post-performance judgments accuracy: executive functions, affective-motivational components related to mathematics, achievement in this subject, and task characteristics (perceived difficulty, and total time spent on the problem-solving tasks). Prior to analysing their predictive value, differences in these variables between groups with different calibration (Accurate vs. Inaccurate) were examined. Students were assigned to these two groups based on the accuracy of their post-performance judgment in both mathematical problems. Students who judge accurately their performance only in one of the problems were excluded from the analyses to avoid hazard responses. Thus, a sample of 188 students was used in this study. Executive functions were assessed by means of the EFE scale (García et al., 2013) described above, only in its form to families, in order not to lose excessive sample size.

Regarding differences between groups with different calibration accuracy, the results indicated that accurate students showed improved motivations, beliefs and attitudes towards mathematics, as well as higher levels of achievement in the subject than their peers in the inaccurate group. Accurate students were evaluated by families as significantly less impulsive, also reporting better attention, memory, planning and organization skills in this group. They also spend more time solving problems than the inaccurate group, mainly in the first problem. Finally, with regard to the predictive value of these variables, mathematics performance, perceived usefulness or value of the subject, and the total time spent solving the first problem, significantly predicted differences in calibration, classifying 71.3% of the sample correctly.

The results obtained in these five studies showed the usefulness of the designed measure –*Tripe Task Procedure in Mathematics (TTPM)*– in the evaluation of the metacognitive process involved in solving mathematical problems, and the potential value of this process explaining differences in post-performance calibration. Calibration is, however, a complex mechanism which depends upon multiple variables, being necessary to address this analysis from a comprehensive perspective.

Introducción

Las matemáticas se han convertido en una herramienta básica en nuestro día a día. En pleno siglo XXI una buena competencia matemática es, junto con la competencia digital, esencial. Este aspecto hace necesario desarrollar estrategias de Enseñanza/Aprendizaje que ayuden a los estudiantes a construir un buen conocimiento en esta área. En este sentido, hoy se entiende que las matemáticas implican algo más que el conocimiento de hechos numéricos o el dominio de una serie de habilidades básicas y procedimientos. Desarrollar un buen conocimiento matemático requiere entender y ser capaces de usar las matemáticas en respuesta a las exigencias de un mundo cambiante. Para autores como (Das y Das, 2013), esto implicaría cuatro aspectos básicos: primero, disponer de un conocimiento conceptual, entendido como la capacidad para entender conceptos matemáticos, operaciones y relaciones; segundo, tener un conocimiento procedimental, o la habilidad para llevar a cabo diferentes procedimientos de forma flexible, precisa y eficiente; tercero, presentar una competencia estratégica, o la capacidad para formular, representar y resolver problemas; y finalmente, disponer de unas buenas capacidades de razonamiento adaptativo, definido como pensamiento lógico, capacidad de juicio y reflexión. Este trabajo de Tesis Doctoral se centra en la resolución de problemas, entendida ésta como un medio que permite construir ideas propias, favorece el pensamiento lógico y la generalización de habilidades a situaciones poco familiares.

La resolución de problemas matemáticos en el contexto educativo actual

La resolución de problemas matemáticos es una de las actividades más frecuentes a las que se enfrentan estudiantes de muy diversas edades. Sin embargo, se trata también de una actividad compleja cuya correcta ejecución implica múltiples procesos. De este modo, resolver un problema matemático no solamente implicaría entender o representar la situación del problema, crear algoritmos, procesar diferentes tipos de información o realizar cálculos matemáticos, sino también la capacidad de identificar, manejar y evaluar diferentes estrategias en respuesta al tipo de problema (Zhu, 2007). Dada su complejidad, no es extraño que numerosos estudiantes tengan dificultades en este ámbito ya desde edades tempranas. Esto es especialmente relevante teniendo en cuenta que, sin una adecuada intervención, estas dificultades en la resolución de problemas tienden a agravarse, dando lugar a resultados como los obtenidos en el último informe PISA (Organización para la Cooperación y el Desarrollo Económico- OCDE, 2014) entre otros. Los últimos datos publicados se refieren a las evaluaciones realizadas a 85.000 estudiantes de 15 años pertenecientes a 44 países diferentes, 28 de ellos pertenecientes a la OCDE, durante mayo de 2012. Los resultados obtenidos indicaron que el rendimiento medio del alumnado español en resolución de problemas matemáticos fue significativamente menor a la media del resto de países de la OCDE (477 frente a 500 puntos). La puntuación media obtenida por España se correspondería con el nivel 2 de competencias de los 6 establecidos en PISA, y situaría al 28.5% de nuestros estudiantes en el grupo de baja competencia en esta área, mientras que solamente el 7.8% lograría estar entre los alumnos considerados excelentes. Estos resultados siguen la línea de los encontrados en ejercicios anteriores en el área de las matemáticas. En concreto, en PISA 2012 España obtuvo 484 puntos en matemáticas, 10 puntos menos que el promedio de países de la OCDE, siendo estas cifras similares a las obtenidas en 2009 y 2006 (OCDE, 2014). Estos ciclos no incluyeron, no obstante, una medida específica de resolución de problemas. En el ciclo 2003 (OCDE, 2004) si se incorporó la resolución de problemas como una materia de carácter transversal. En esta ocasión, un total de 250.000 estudiantes de 41 países se enfrentaron a problemas basados en contextos de la vida real, siendo el rendimiento de los estudiantes españoles en estas pruebas nuevamente inferior al promedio (482 puntos frente a 500 puntos de media de los países de la OCDE).

Cabe plantearse en este sentido la siguiente cuestión: ¿A qué pueden deberse estas cifras? Más allá de la discusión en torno al tipo de problemas usados o el procedimiento empleado para la obtención de los datos, parece evidente que los estudiantes en estas etapas educativas (tercer curso de Educación Secundaria Obligatoria) están empleando estrategias de resolución de problemas ineficaces. Entre otros muchos factores, cabe esperar que el origen de estas dificultades se encuentre en ciertos patrones de resolución de problemas presentes en etapas educativas anteriores, siendo los últimos cursos de Educación Primaria especialmente relevantes en este sentido. Partiendo de esta premisa, la Tesis Doctoral aquí presentada ha contado para su desarrollo con una muestra de 524 estudiantes de quinto y sexto curso de Educación Primaria, pertenecientes a 12 centros educativos del Principado de Asturias.

La relevancia dada a estos niveles educativos se debe en parte a que se trata de una etapa de transición hacia niveles educativos posteriores, lo cual conlleva un incremento significativo del volumen y la complejidad de la información a procesar. Esta etapa coincide además para numerosos autores con el momento en que los estudiantes empiezan a mostrar actitudes, creencias y motivaciones más claras en relación al aprendizaje de las matemáticas (Adelson y McCoach, 2011; Sakiz, Pape, y Woolfolk, 2012; Waters, Cross, y Runions, 2009). En este sentido, los estudiantes en estas etapas educativas serían capaces de expresar si les gusta o no la materia, el grado en que valoran tener un buen conocimiento de la misma, hasta qué punto se sienten capaces de aprender matemáticas o de realizar una determinada tarea, así como de realizar evaluaciones o juicios más o menos precisos sobre su ejecución. Todas estas variables tienen su efecto en el esfuerzo y la persistencia en la tarea, así como en el rendimiento final en la misma.

Relevancia del proceso en resolución de problemas matemáticos y su relación con la calibración

Ahora bien, existen infinidad de factores que pueden influir en la resolución de problemas matemáticos en estas edades, desde variables del individuo o de la tarea hasta variables del contexto, sin olvidar sus posibles interacciones (Blanco, Guerrero, y Caballero, 2013; Ramírez, Gunderson, Levine, y Beilock, 2013; Stacey, 2005; Stein, Remillard y Smith, 2007). De este modo, mientras que numerosos estudios han señalado la importancia de disponer de un conocimiento flexible y bien organizado sobre hechos matemáticos, símbolos, fórmulas, conceptos y reglas como aspecto central para una buena resolución de problemas (Gálvez et al., 2011; Geary, 2004; Jarero, Aparicio, y Sosa, 2013; Verschaffel, Greer, y De Corte, 2000), son también muchos los trabajos que ha mostrado como el éxito en estas tareas se relaciona con el uso de estrategias basadas en el *Aprendizaje Autorregulado* (Lazakidou y Retalis, 2010; Montague, 2008).

El término *Aprendizaje Autorregulado* (Self-Regulated Learning – SRL– en Inglés) se define como "*un proceso activo en el cual los estudiantes establecen los objetivos que guían su aprendizaje intentando monitorizar, regular y controlar su cognición, motivación y comportamiento con la intención de alcanzarlos*" (Rosário, 2004, p. 37). Como señalan Núñez, Solano, González-Pienda y Rosário (2006), con el fin de lograr estos objetivos, el aprendizaje autorregulado se presenta como un proceso abierto que se produce de forma cíclica. Esta actividad cíclica se basaría en tres fases: una fase previa de planificación, una fase de realización o ejecución, y una fase de auto-reflexión o evaluación (Zimmerman, 2000, 2008). Estas fases estarían formadas cada una de ellas por diferentes subprocesos, y su comportamiento obedecería a las fluctuaciones en los componentes personal, conductual y contextual.

En este sentido, la evidencia empírica muestra que los estudiantes autorregulados presentan un mejor rendimiento en resolución de problemas matemáticos, presumiblemente porque ejercen un control más eficaz sobre sus propios procesos cognitivos, motivacionales y comportamentales durante la realización de este tipo de tareas (Pennequin, Sorel, Nanty, y Fontaine, 2010; Pereis, Dignath, y Schmitz, 2009; Schmitz y Perel, 2012). No obstante, son numerosos los trabajos que señalan que los estudiantes en estas etapas educativas (últimos años de Educación Primaria) e incluso posteriores tienden a ser poco autorregulados cuando se enfrentan a tareas de resolución de problemas, ya sea porque carecen del conocimiento y/o de las habilidades necesarias para hacerlo, y esto se refleja a menudo en el proceso de resolución de problemas que éstos llevan a cabo (Azevedo y Cromley, 2004; Butler y Cartier, 2005; Ifenthaler, 2012; Kramaski y Gutman, 2006; Nietfeld, Cao, y Osborne, 2005; Zimmerman, Moylan, Hudesman, White, y Flugman, 2011).

Desde la perspectiva del análisis del proceso metacognitivo implicado en la resolución de problemas matemáticos, es común observar ciertos patrones de resolución ineficaces. Una de las características más destacadas en este sentido es el hecho de que numerosos estudiantes tienden a mostrar escasas habilidades de planificación, así como un uso infrecuente de mecanismos de evaluación tanto del progreso como del resultado de su actividad de resolución de problemas (Cleary y Chen, 2009; Kramarski y Gutman, 2006; Montague, Enders, y Dietz, 2011). Esto se traduce a menudo en una tendencia a dar respuestas azarosas o impulsivas, en el uso de mecanismos de ensayo y error como método de resolución de problemas, o en una excesiva confianza en procedimientos familiares como la realización de cálculos como medio para resolver cualquier problema, relegando a un segundo plano el uso de estrategias potencialmente más efectivas como puedan ser el uso de diferentes formas de representación u organización de la información (Alexander 2013; Cleary 2009; Efklides y Misailidi, 2010; Schunk y Pajares, 2009).

Este tipo de patrones de resolución de problemas –procesos en el contexto de esta Tesis Doctoral– tienen su impacto en el rendimiento final en la tarea –es decir, en el producto de la ejecución–, llevando a menudo a los estudiantes a fracasar en sus intentos por realizar exitosamente la tarea, pero también sobre la motivación, las expectativas y la percepción o juicios de los estudiantes sobre su ejecución. De este modo, y ligadas a estos juicios, las habilidades de *"calibración"* suponen un mecanismo de control metacognitivo muy importante.

La calibración se define como el grado en que los juicios de una persona sobre la adecuación o corrección de su ejecución en una tarea se corresponde con su ejecución real, establecida ésta en función de una medida objetiva, como la puntuación en un test o la calificación en un examen (Hacker, Bol, y Keener, 2008a). Estos juicios pueden ser realizados antes –*predicciones*– o después –*postdicciones*– de la realización de la tarea y han sido extensivamente estudiados en matemáticas y resolución de problemas en particular, mostrando la existencia de una importante asociación entre la precisión de estos juicios y el rendimiento en estos ámbitos (Desoete y Roeyers, 2006; Jacobse y Harskamp, 2012; Özsoy, 2012; Rinne y Mazzocco, 2014).

No obstante, los estudios previos coinciden en señalar que los estudiantes tienden a ser poco calibrados, mostrando a menudo un exceso de confianza. Lo cierto sin embargo es que tanto el exceso como la falta de confianza sobre la propia ejecución han mostrado tener un efecto negativo sobre el rendimiento en la tarea, así como sobre las creencias y expectativas de los estudiantes (Dupeyrat, Escribe, Huet, y Régner, 2011; Sheldrake, Mujtaba, y Reiss, 2014). Esto hace necesario determinar qué factores influyen en la formación y precisión de estos juicios, sobre todo teniendo en cuenta que tanto los juicios como los posibles sesgos en este sentido tienden a ser estables en el tiempo, lo cual les hace más resistentes a los efectos de la intervención (Bouffard, Vezeau, Roy, y Lengelé, 2011; Hacker, Bol, Horgan, y Rakow, 2000; Hacker et al., 2008a; Nietfeld, Cao, y Osborne, 2006).

Dada esta realidad, desde el punto de vista del análisis de los factores implicados en la calibración, un aspecto relevante en el contexto de esta Tesis Doctoral es la distinción previamente planteada entre *predicciones* y *postdicciones*. Si bien ambos aspectos son entendidos como importantes componentes de control metacognitivo, la evidencia empírica sugiere que éstos responderían a mecanismos diferentes (Bol, Hacker, O'Shea, y Allen, 2015; Bol, Hacker, Walck, y Nunnery, 2012; McCormick 2003; Hacker et al. 2008a; Sheldrake et al., 2014). De este modo, mientras las predicciones supondrían un tipo de juicio de auto-eficacia, las postdicciones serían informativas de los mecanismos de monitorización durante la tarea. Estos estudios sugieren además que éstas últimas serían medidas más precisas y fiables de la calibración, debido al feed-back procedente de la propia exposición y experimentación con la tarea, lo que las hace susceptibles de ser analizadas desde el punto de vista del proceso.

Si bien el origen de estos juicios ha sido extensivamente estudiado, atendiendo a factores tan variados como las habilidades cognitivas, el rendimiento académico, el grado de experiencia formulando estos juicios, las propias características de la tarea, variables más de tipo afectivo-motivacional como creencias de auto-eficacia o establecimiento de objetivos, o rasgos más estables como estilos atribucionales o de personalidad (Alexander 2013; Bol et al. 2012; Dinsmore y Parkinson, 2013; Hacker, Bol, y Bahbahadi, 2008b; Hadwin y Webster, 2013; Stolp y Zabrucky, 2009), lo cierto es que el trabajo realizado hasta ahora en este ámbito presentaría no obstante algunas limitaciones: en primer lugar, los estudios en Educación Primaria escasos siendo la gran mayoría de trabajos llevados a cabo en etapas educativas posteriores; en segundo lugar, el potencial explicativo del proceso llevado a cabo durante la tarea sobre la calibración no ha sido propiamente analizado; y finalmente, pocos trabajos han abordado el estudio de los determinantes de la calibración desde un punto de vista comprensivo, centrándose en su lugar en examinar aspectos aislados.

La presente Tesis Doctoral

Dicho lo anterior, la Tesis Doctoral que aquí se presenta va dirigida a responder las siguientes cuestiones:

1. ¿Cómo se puede evaluar el proceso y qué puede decir éste acerca de las estrategias metacognitivas empleadas por los estudiantes durante la resolución de problemas matemáticos?
2. ¿Puede el proceso informar sobre los mecanismos implicados en la realización de juicios tras la ejecución de este tipo de tareas?
3. ¿Hay otras variables afectivas, motivacionales, cognitivas o conductuales que puedan explicar las diferencias en estos juicios?

En relación a la última de las cuestiones, se ha hecho especial énfasis en las funciones ejecutivas como mecanismos de control de orden superior (Anderson, Jacobs, y Anderson, 2008; García, González-Castro, Areces, Cueli, y Rodríguez, 2014; Meltzer, 2013), así como un enfoque profundo de aprendizaje, definido como una combinación de intenciones, motivos y estrategias que daría lugar a un aprendizaje más efectivo y duradero (Baeten, Kyndt, Struyven, y Dochy, 2010; McNamara, 2011; Phan, 2011; Ranellucci et al., 2013), como potenciales variables explicativas de las diferencias en calibración. Previo a establecer esta relación, y dada la caracterización de la calibración como un importante mecanismo metacognitivo, se examina la asociación de las estas dos variables con los componentes de conocimiento y habilidades metacognitivas.

Para dar respuesta a la totalidad de las cuestiones planteadas en la presente Tesis Doctoral, se han realizado tres publicaciones en revistas de conocido impacto en el área de Psicología y Educación, así como tres trabajos complementarios. Estos trabajos se muestran más adelante en el presente documento.

Objetivos

El objetivo principal de esta Tesis Doctoral has sido responder a las tres cuestiones planteadas previamente. Se han establecido para ello una serie de objetivos específicos en torno a estas cuestiones:

1. ¿Cómo se puede evaluar el proceso y qué puede decir éste acerca de las estrategias metacognitivas empleadas por los estudiantes durante la resolución de problemas matemáticos? Artículo 1.

   Objetivos específicos:

   - Diseñar y aplicar un procedimiento de evaluación del proceso de resolución de problemas matemáticos.
   - Analizar el proceso metacognitivo mostrado por los estudiantes durante la realización de estas tareas, así como su relación con el "producto" o resultado final en las mismas.

2. ¿Puede el proceso informar sobre los mecanismos implicados en la realización de juicios tras la ejecución de este tipo de tareas? Trabajo Complementario 1.

   Objetivos específicos:

   - Analizar el grado de precisión de estos juicios mediante el cálculo de un índice de calibración.
   - Examinar las diferencias en el proceso mostradas por los estudiantes con diferente precisión en sus juicios post-ejecución.

3. ¿Hay otras variables afectivas, motivacionales, cognitivas o conductuales que puedan explicar las diferencias en estos juicios?

   Objetivos específicos:

   - Delimitar el potencial valor explicativo de las habilidades de funcionamiento ejecutivo y el uso de un enfoque profundo de aprendizaje sobre estos juicios, partiendo de la relación de estos componentes con el conocimiento y las habilidades metacognitivas. Artículos 2 y 3, Trabajo Complementario 2.

- Analizar el valor predictivo sobre la calibración post-ejecución de los componentes que han mostrado ser significativos en la etapa anterior, incluyendo en este análisis variables afectivo-motivacionales, el rendimiento en matemáticas y las características de la tarea. Trabajo Complementario 3.

Publicaciones

Artículo 1:

García, T., Betts, L., González-Castro, P., González-Pienda, J. A. & Rodríguez, C. (In press). On-line assessment of the process involved in Maths problem-solving in fifth and sixth grade students: self-regulation and achievement. *Revista Latinoamericana de Investigación en Matemática Educativa.*

Artículo 2:

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),* 47-52.

Artículo 3:

García, T., Rodríguez, C., González-Castro, P., Álvarez-García, D. & González-Pienda, J. A. (In press). Metacognición y funcionamiento ejecutivo en Educación Primaria [Metacognition and executive functioning in Elementary School]. *Anales de Psicología.*

Nota. Dos de estos artículos actualmente se encuentran en proceso de publicación, por lo que se proporcionan a continuación los justificantes de aceptación de cada uno de ellos.

Revista Latinoamericana de Investigación en Matemática Educativa
RELIME

Dirección Editorial

Asunto: Constancia de Publicación

Ref. Relime: CONST-PUBL/2015/093

Trinidad García
Universidad de Oviedo - España
Lucy Betts
Nottingham Trent University - UK
Paloma González-Castro
Julio Antonio González-Pienda
Celestino Rodríguez
Universidad de Oviedo - España

Por medio de la presente constancia, me complace informarles que el Comité de Redacción dictaminó que el manuscrito que propusieron para su publicación, titulado:

"On-line assessment of the process involved in Maths problem-solving in fifth and sixth grade students: self-regulation and achievement"

ha sido aceptado para su publicación. El artículo se publicará en uno de los próximos volúmenes de nuestra revista.

Agradecemos su interés por publicar en *Relime*.

Se extiende la presente constancia a los cuatro días del mes de marzo de dos mil quince.

Dr. Ricardo Cantoral Uriza
Director Editorial

Oficina # 101. Departamento de Matemática Educativa. Cinvestav del IPN
Avenida Instituto Politécnico Nacional # 2508 ❑ 07360, San Pedro Zacatenco, México DF ❑ MÉXICO
Teléfono: (52) 55 – 57 47 60 07 ❑ Telefacsímil: (52) 55 – 57 47 38 23 ❑ relime@clame.org.mx

anales de psicología
UNIVERSIDAD DE MURCIA
SERVICIO DE PUBLICACIONES
ISSN edición impresa: 0212-9728
ISSN edición web: 1695-2294
http://revistas.um.es/analesps/

Revista científica de Psicología
Periodicidad cuatrimestral (un volumen anual con tres números en enero, mayo y octubre).
Sistema estandarizado de revisión científica externa de originales

anales de psicología se encuentra en los siguientes **repertorios de factor de impacto**:
- JCR (*Journal Citation Reports*):
  Factor de impacto (jun 2011): **1.338**
  Ranking: 47 de 120 (en *Psychology-Multidisciplinary*)
- SJR (*Scimago Journal Rank*, de Elsevier).
  Factor de impacto 2011: 0.028
- In-Recs (Granada, España):
  (http://ec3.ugr.es/in-recs/ii/Psicologia-2010.htm)
  Factor de impacto 2010 (publ. 2011): 1.068
  Ranking: nº 2 en Psicología

Principales repertorios de índices/sumarios y bases de datos en que se encuentra indexada:
- SSCI (*Social Sciences Citation Index*), SCIexp (*Science Citation Index Expanded*), Web of Knowledge, Web of Science (todas de ISI Thomson Reuters)
- SCOPUS (de Elsevier)
- PsycInfo database, Psychological Abstracts y PsycLit (de A.P.A., American Psychological Ass., Washington, U.S.A.)
- Indice Español de Ciencias Sociales A: Psicología y Educación (CINDOC, CSIC, Madrid),
- PSICODOC (Col. Ofic. Psicólogos, España).
- Anuario de Psicología Clínica producida en lengua española (Madrid),
- Ulrich's International Periodicals Directory
- PIO (Periodical Index Online)
- En la web (open Access): Open Science Directroy (EBSCO), DOAJ, SCIRUS, Psycline, PsychSpider, Google Scholar, Dialnet, e-revistas, Recolecta, RedAlyc

Principales índices de calidad editorial y sistemas de evaluación de revistas:
- *ERIH* (European Reference Index for the Humanities, de la European Science Foundation)
- Evaluación FECYT-RECYT (Madrid): Sello de excelencia, edición 2011
- Criterios Latindex
- Criterios CNEAI
- DICE (Difusión y Calidad Editorial de las Revistas Españolas de Humanidades y Ciencias Sociales y Jurídicas; CSIC, CINDOC, ANECA)
- RESH (Revistas Españolas de Ciencias Sociales y Humanas: Valoración integrada; IEDCYT-CSIC)
- MIAR (Matriu d'informació per a l'avaluació de Revistes. Base de datos con la que se obtiene el ICDS)

DIRECCIONES:

Envío y recepción de originales:
Director de "Anales de Psicología"
Facultad de Psicología,
Universidad de Murcia
Campus de Espinardo
Aptdo. 4021, 30080 Murcia (España).
Correo electrónico: analesps@um.es

Solicitud de canje con otras revistas:
Servicio de Intercambio Científico.
Universidad de Murcia
Biblioteca General (Campus de Espinardo)
Aptdo. 4021, 30080 Murcia (España).
Correo electrónico: mdem@um.es
Internet, dirección World Wide Web:http://revistas.um.es/analesps/

DATOS DEL ARTÍCULO:

Título: Metacognición y funcionamiento ejecutivo en Educación Primaria (Metacognition and executive functioning in Primary School)

Autores: Trinidad García, Celestino Rodríguez, Paloma González-Castro, David Álvarez-García, Julio Antonio González-Pienda

CÓDIGO: 202891

Recibido: 18/07/2014

Asunto: ACEPTACIÓN DEL ARTÍCULO

Agustín Romero Medina
Director de **anales de psicología**
Facultad de Psicología, Universidad de Murcia
Aptdo. 4021, 30080 Murcia
Tlf.: 868 883483, Fax: 868 884115
E mail: analesps@um.es

Murcia, a 19 de febrero de 2015

Estimado compañero:

Por la presente, te comunico que, con fecha de hoy, el artículo enviado a **anales de psicología** cuyos datos son los mencionados más arriba, **ha sido aceptado** para su publicación en esta revista. La separata del artículo la obtendrás en formato pdf en la página web de la revista (http://revistas.um.es/analesps/) en cuanto sea publicado en número correspondiente.

Gracias por tu colaboración y espero que sigas contando con esta revista para publicar tus trabajos.

Recibe un cordial saludo,

anales DE psicología ISSN: 0212-9728
UNIVERSIDAD DE MURCIA SERVICIO DE PUBLICACIONES

Fdo.: Agustín Romero Medina

ON-LINE ASSESSMENT OF THE PROCESS INVOLVED IN MATHS PROBLEM SOLVING IN FIFTH AND SIXTH GRADE STUDENTS: SELF-REGULATION AND ACHIEVEMENT

TRINIDAD GARCÍA, LUCY BETTS*, PALOMA GONZÁLEZ-CASTRO, JULIO ANTONIO GONZÁLEZ-PIENDA & CELESTINO RODRÍGUEZ

UNIVERSITY OF OVIEDO (SPAIN), * NOTTINGHAM TRENT UNIVERSITY (UNITED KINGDOM)

IN PRESS

ABSTRACT. The aim of this study was to test a method to assess the processes involved in mathematical problem solving, based on the Triple Task methodology and Self-regulated Learning principles. This protocol was administered to 510 fifth and sixth grade students from Northern Spain, who carried out two mathematical problems. The results derived from the total sample indicated the presence of ineffective planning strategies and a lack of evaluation mechanisms. However, comparisons between groups with different achievement in the problems revealed the sub-processes involved in planning (especially the use of representation strategies) as important determining factors in students´ success rates, exerting a greater effect on the second problem, maybe related to an increase in task difficulty.

Keywords: Self-Regulated Learning, Mathematics, problem solving, process.

RESUMEN. El objetivo de este estudio ha sido poner a prueba un método de evaluación del proceso implicado en la resolución de problemas matemáticos, basado en la metodología de la Triple Tarea y en los principios del Aprendizaje Autorregulado. Este protocolo se administró a 510 estudiantes de quinto y sexto curso procedentes del Norte de España, los cuales realizaron dos tareas matemáticas de diferente dificultad. Los resultados indicaron la presencia de unas estrategias de planificación ineficaces, así como la ausencia de mecanismos de revisión. Sin embargo, el análisis de las diferencias entre los grupos con diferente rendimiento en las tareas reveló los sub-procesos implicados en la planificación, y especialmente el empleo de estrategias de representación de la información, como determinantes importantes en el éxito de los estudiantes, ejerciendo un efecto en el segundo problema, quizás debido a diferencias en la dificultad de los problemas.

Palabras clave: Aprendizaje Auto-Regulado, matemáticas, resolución de problemas, proceso.

1. BACKGROUND

It is widely known that the development of problem-solving skills is an important issue in formal education. In fact, authors such as Lazakidou and Retalis (2010) linked it to Life-long Learning in the sense that *"engagement in problem-solving activities helps students to acquire useful attitudes such as thinking, flexibility, creativity, and productivity, which are very important to real life"* (p.3). In the current educational context, mathematical problem solving takes a great relevance. This activity, which starts in the first years of elementary education, may be the basis to develop basic problem-solving skills.

Mathematical problem solving is a complex cognitive activity that involves multiple processes. A student's achievement in these tasks relies on the integrated application of cognitive, metacognitive, and motivational components (Cleary & Chen, 2009; Kajamies, Vauras, & Kinnunen, 2010; Montague, Enders, & Dietz, 2011; Voyer, 2011). In fact, while it has been supported for many years that a well-organized and flexibly accessible knowledge about mathematical facts, symbols, algorithms, concepts and rules may be the core of problem-solving skills, current literature has established that strategy use based on Self-regulated Learning (SRL) is linked to mathematical capacities which are an important determining factor in solving mathematical problems (Gálvez, Cosmelli, Cubillos et al., 2011; Geary, 2004; Jarero, Aparicio, & Sosa, 2013; Lazakidou & Retalis, 2010; Montague, 2008; Pennequin, Sorel, Nanty, & Fontaine, 2010; Pereis, Dignath, & Schmitz, 2009; Schmitz & Perel, 2012; Verschaffel, Greer &, De Corte, 2000).

SRL refers to those proactively initiated thoughts, feelings and behaviors, which are planned and cyclically adapted, based on self-generated or performance feedback in order to attain personal goals (Zimmerman, 2000). SRL involves those aspects which facilitate the control and regulation of students' cognitive systems and learning processes, and comprises three sequential phases: forethought (i.e., processes that precede efforts to learn or perform), performance control (i.e., processes occurring during learning efforts), and self-reflection (i.e., processes occurring after learning or performance such as evaluate the effectiveness of one´s learning methods or results; Zimmerman, 2000; in Cleary & Chen, 2009). These processes, which are based on an interplay of personal and task characteristics and the strategies available in a learning situation, help students to monitor and check their thoughts and are particularly useful when solving novel or challenging problems (Lazakidou & Retalis, 2010; Pennequin et al., 2010; Rosenzweig, Krawec, & Montague, 2011; Throndsen, 2011; Zimmerman & Schunk, 2008). However, previous research has shown that students tend to demonstrate poor metacognitive skills while engaged in mathematical problem-solving situations. Many students forge ahead without considering alternative decisions, jumping immediately into calculations, giving impulsive responses, and using trial and error as strategies. Students often get stuck in irrelevant details of the task or fail to verify solution paths and evaluate answers, focusing on superficial measures of progress (Cleary & Chen, 2009; Kramarski & Gutman, 2006; Montague, 2008; Montague et al., 2011; Pennequin et al, 2010; Pereis et al., 2009; Veenman, 2005). In this sense, not only do students with learning disabilities show problems in these issues. In fact, typical students may also not realize the importance of regulating their thoughts and behavior, or they may simply not know how to self-regulate properly. Thus, many students often need support to regulate their learning processes (Azevedo & Cromley 2004; Butler & Cartier 2005; Ifenthaler, 2012; Kramaski & Gutman, 2006).

In this context, one of the biggest challenges for researchers is to design measures that provide access to internal cognitive structures and functions involved in both, self-regulation and problem solving. Most studies based on applying standardized measures such as questionnaires or structured interviews; this kind of measures involves asking students about how they solve a problem or the extent to which they use different strategies. Additionally, students are also asked about different situations in which they would use (or not use) a specific strategy or what approximation would be the best in each case. Students' responses to the metacognitive questions are scored depending on the quality of the response and a total score is calculated. These kind of assessments, although widely applied (Greene & Azevedo, 2010; Shea & Bidjerano, 2010; Zimmerman & Schunk, 2008), may yield inconsistent or inaccurate information due to inaccurate memory or perhaps even response biases such as social desirability, and do not provide information about how learners transfer their knowledge to regulate their problem-solving activities. As many authors have pointed out, these tools are useful to assess a student's declarative and situational metacognitive knowledge, but do not provide any information about the mental processes underlying task performance and the student's achievement levels which together represent important issues in learning and instruction (Clearly & Chan, 2006; Veenman, 2011). In response to this, Veenman (2011) established the distinction between off-line and on-line methods in the assessment of strategy use and learning. Off-line methods refer to measures administered either before or after task performance, such as questionnaires and interviews, while on-line methods concern measurements taken concurrent to task performance, such as Think-aloud or

Triple Task procedures. These measures may be especially useful to provide data about student reasoning abilities during problem-solving activities.

In *Think-aloud* protocols (Rosenzweig et al., 2011) learners verbalize their thoughts and cognitive activities while engaged in task execution. The usefulness of these procedures lie in the fact that they provide access to students' short-term memory abilities, which reflects cognitive processing during task completion. This approach allows the researcher to obtain accurate information about metacognitive skills and strategies while students are asked to simply verbalize what they are thinking at each moment. After transcription and coding, undirected verbalizations are recognized as valid expressions of the students' cognitive and metacognitive processes. Studies in this area are limited, although some interesting data support the usefulness of the protocols in assessing metacognitive processes during mathematical problem-solving tasks (Lazakidou & Retalis, 2010; Ostad & Sorenson, 2007; Throndsen, 2011).

On the other hand, *Triple Task* procedures have been widely applied in the study of the processes involved in written composition, and more recently in note-taking activities (Fidalgo, Torrance, Robledo, & García, 2009; García, Rodríguez, Pacheco, & Diez, 2009; Piolat, Kellog, & Farioli, 2001; Olive, Kellog, & Piolat, 2002; Piolat, Barbier, & Roussey, 2008; Piolat & Olive, 2000; Piolat, Olive, & Kellogg, 2005; Torrance & Galbraith, 2006).

The Triple Task method is rooted in the so-called *"Double Task"* protocols. Although the present study focuses on the use of Triple Task procedures, a description of both methods is provided below:

The *Double Task* procedure provides information about the cognitive effort engaged in higher-order cognitive tasks, such as comprehension or text production. Participants are asked to perform concurrently a primary task and a secondary probe task. For example, while composing a text (the primary task), participants must react as fast as possible to tones (the secondary task by pressing a mouse button or by saying 'stop' to a microphone linked to a vocal key) that are periodically distributed in a random interval (generally between 15 and 45s). Reaction time (RT) in this dual-task situation is compared with a control condition when the probe is responded to as a single task. The degree of interference in RT (IRT) caused by the primary task provides a measure of the amount of cognitive effort designated to the task. However, although widely used, this measure does not provide information about the process that underlies the cognitive task under consideration.

It is the *Triple Task* procedure which provides additional information about the temporal organization and the cognitive processes underlying a cognitive activity by adding a third task, consisting of asking students for an immediate directed introspection after each tone's detection. Students have to categorize their thoughts at the moment the tone is presented. Thus, in Triple Task studies students perform: a) the primary task under investigation, b) the secondary probe task, and c) a third task in which participants are asked to label the process that was interrupted by the probe. For this last task, participants are trained in direct retrospection in order to better identify and report the cognitive processes present at each moment. As a measure of the process, the Triple Task method can be differentiated from Think-aloud protocols in the following aspects: a) the Triple Task uses directed introspection as opposed to the undirected introspection involved in Think-aloud, b) a system of categories can be given in order to facilitate categorization and facilitate exploration in to sub-processes, c) response transcription and coding are unnecessary, and d) inter-rater agreement does not have to be calculated. To date however, this method has neither been applied to the analysis of the processes underlying mathematical problem solving, nor from the point of view of investigating the metacognitive skills involved in these tasks. Additionally, as happens with Think-aloud protocols, most Triple Task studies have been conducted with small sample sizes, mainly due to their own methodology features, which make them less economic to implement.

In this context, we introduce our own process assessment tool, the *Triple Task Procedure in Mathematics–TTPM* (García & González-Pienda, 2012). It is based on the Triple Task technique and suitable for application in larger samples through a MOODLE platform. One of the novelties of this method is the introduction of a new system of categories to analyze student processes during problem solving. It was designed according to the PLEJE model of SRL ("**Pl**anificación-**Ej**ecución-**E**valuación" in Spanish: Rosário, Mourão, Núñez, González-Pienda, & Solano, 2008; Zimmerman, 2000), which establishes three main phases: Planning, Execution and Evaluation; and Bransford and Stein´s (1993) IDEAL problem-solving model, which stablishes five stages to successfully solve a problem: **I**dentifying potential problems, **D**efining and representing the problem, **E**xploring possible strategies, **A**cting on those strategies, and **L**ooking back and evaluating the effects of those activities. A more detailed description of the assessment process and this tool, which was administered to 510 students from fifth and sixth grade belonging to 12 private and state schools in Asturias (Northern Spain), is provided in the method section.

The main aim of this study was to analyze the usefulness of TTPM as a measure of the process during the performance of two mathematical problems.

Specifically, this study sought to answer the following questions:

1) Do students follow Self-regulated Learning stages while they try to solve mathematical problems?
2) Are there differences in the process followed by students with different achievement in the tasks?

2. METHOD

2.1. Participants

This study involved 510 students from fifth and sixth grade students from 12 private and state primary schools in northern Spain (Mean age: 11 years, SD:0.71). Students belonged to 32 different classrooms. Gender distribution was balanced. Specifically, 255(50%) were female. Of the total sample, 210 students (41.2%) attended fifth and 300 (58.8%) sixth grade. Gender and level (fifth or sixth grade) were taken into account as potential mediating variables in subsequent analyses.

Sample selectionwas made through convenience or accessibility procedures. Students volunteered for the study and presented informed consent from their parents.

2.2. Measures

Mathematical problems: students completed two relatively complex mathematical word problems taken from the book *"Problem-solving and comprehension"* (Whimbey & Lochhead, 1993) translated into Spanish. This book provides a systematic review of the characteristics that define a good problem solver, while proposing strategies and activities to develop these skills, mainly based on mathematical reasoning. These problems differ in the amount of information to be processed, the number of associations presented and the number of parameters to calculate.

Process measures: evidence of student metacognitive process was obtained by means of the TTPM (García & González-Pienda, 2012). It consists in the following: after a RT probe task, and while completing the mathematical problems, student are presented with an electronic tone at a randomized interval of 35-45 seconds. At this point students are asked to categorize their activities or thoughts. In order to help students categorize these processes, students are provided with a category system (see Table 1). They have to indicate by mouse-click which of nine the categories (reading, drawing or summarizing, recalling similar problems, thinking about a solution, mental calculation, writing, reviewing, correcting mistakes, or “other”) best describes the processes in which they are involved when the tone is presented. The category system is based on Bransford and Stein’s (1993) IDEAL problem-solving model and the PLEJE model of SRL (Rosário et al., 2008; Zimmerman, 2000), previously described in this study. The category called “other” was incorporated to gather all those thoughts or actions unrelated to the mathematical task (e.g., day-dreaming). Nevertheless, previous research indicates that such unrelated processes are generally very scarce (Olive et al., 2002). The eight proposed processes are also organized into three, higher-level categories corresponding to SRL phases: Planning, Execution and Evaluation. Table 1 shows the category system used for students to categorize their processes during the task, based on the correspondence between the stages and phases involved in both SRL and IDEAL models.

TABLE 1

Category system. Based on the Self-Regulation Model (Rosário et al., 2008; Zimmerman, 2000) and the IDEAL Model (Bransford and Stein, 1993)

SRL Model	IDEAL Model	Process categories (I am ...)
Planning	Identification of the problem	Reading
	Definition and representation	Drawing or summarizing
		Recalling similar problems
	Exploration of possible strategies	Thinking about a solution
Execution	Action based on the strategy	Calculating
		Writing a response
Evaluation	Look at effects of solutions	Reviewing
		Correcting mistakes
“Other”		Doing something unrelated

2.3. Procedure

The study was conducted in accordance with The Helsinki Declaration of the World Medical Association (Williams, 2008), which reflects the ethical principles for research involving humans. Participants were tested colectivelly within a 45-minute time frame. A maximum of 20 participants were tested at a time. The first phase consisted of training students to familiarize them with the system of categories and the assessment procedure. The hypothetical case of a boy who was the same age as them (Álex) and who tried to solve a mathematical problem was used for this purpose. After training, students performed a category recognition test consisting of 12 multiple-choice items with four alternatives. Students were asked to indicate the category that best expressed each proposed activity (e.g., for the statement *"Alex realised he made a mistake, so he is erasing", the* alternatives were the following: Álex is thinking about a solution, writing, reviewing or correcting mistakes).

Once the system of categories was understood, students were informed that they would occasionally hear a tone coming from the computer at varying intervals, and they were instructed to quickly react to every tone by clicking on the computer mouse with their dominant hand. This tone was presented in a randomized interval of 10-15 seconds. Then they were informed that this same sound signal would appear while they tried to solve the tasks. They were asked to choose the category that best represented what they were doing in each moment. The tone was presented in a randomized interval of 35-45 seconds. Concurrent to the tone, a pop-up appeared on the computer screen. It showed a box with the category system. Students were able to select a category with a mouse-click. Mathematical problems were provided on paper. Students could use that paper to write whatever they needed, with the condition that they had to write their answer on the paper when they finished the task. As data were collected from individual students simultaneously, headphones were provided in order that other students were not disturbed. Data collection was implemented through Moodle platform (https://moodle.org). A special module was created on the platform, which hosted the evaluation procedure and stored data. In order to accomplish that, a multidisciplinary team, including psychologists, teachers, and a computer engineer (responsible for all technical issues) collaborated during the study. Students accessed this platform through an individual username and password in order to guarantee their anonymity. Once data were obtained and stored, they were automatically transferred to an Excel file for subsequent processing.

Finally, students who showed difficulties understanding the category system (less than 90% correct responses in categorization test) were excluded from the analyses. Process variables were based on frequency counts. Frequency of each sub-process or category was established by dividing the frequency of election of that category by the total number of elections done across categories. These frequency counts were then transformed into percentages by multiplying the quotient by 100. Finally, students´ achievement in the mathematical tasks was established in terms of success (1) or failure (0) according to the answers given on paper. Given the main aims of this study, focused on the process, RT´s as a measure of effort were not taken into account for data analyses.

2.4. Data analysis

Ex-post-facto descriptive and comparative design was used in this study. After describing the profile of process followed by the total group in each task, Multivariate Analysis of Covariance (MANCOVA) analyses with process variables (8 categories and 3 higher-level categories or phases) as dependent variables, and gender and level as covariates, were conducted in order to investigate differences between groups with different achievement in the problems. These problems were analysed separately in order to determine the process related to each one. Effects were significant at $p < .05$ unless otherwise stated. Data were analyzed with SPSS 18. The additional category named "Other" was not included for further analyses.

3. RESULTS

3.1. Student problem-solving process

Figure 1 shows the profile of the process followed by students in Tasks 1 and 2. This profile indicates similar processes for both tasks, characterized by a higher frequency of selected categories in Planning and Execution phases. The means (Table 2) indicated a percentage of 49.46% in Problem 1 and 45.53% in Problem 2 of the students´ reports referring to planning categories, while those related to execution were 40.43% and 45.95%, respectively. With respect to the Evaluation phase, mean percentages of categories reported within this phase were 10.15% in Problem 1 and 7.57% in Problem 2 across categories. It must be noted in this case the high means found in the Planning phase (even above those found in the Execution phase). This is because the Planning phase consists of four sub-processes or categories, while the other phases comprise two.

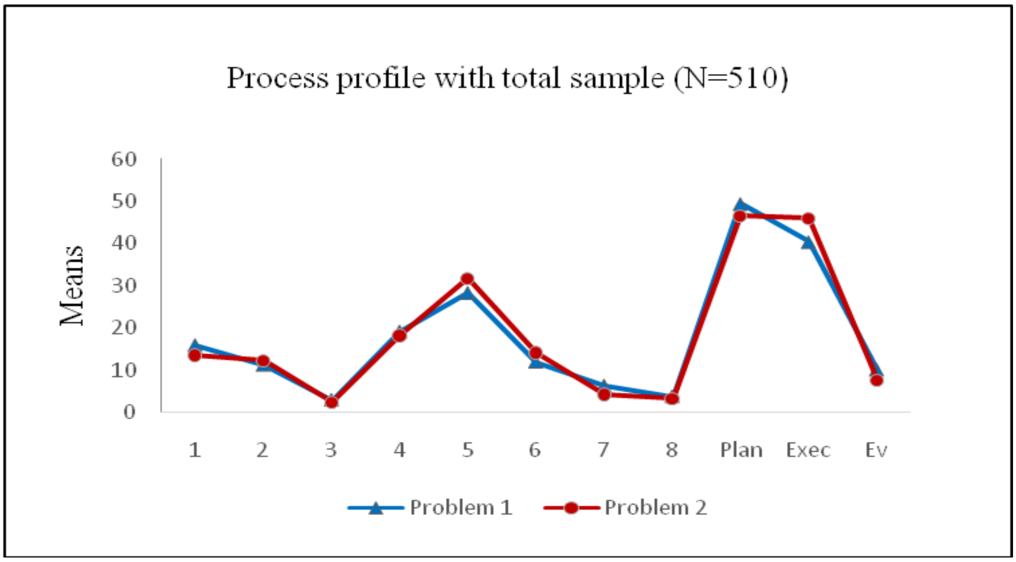


Figure 1. Process profile of the total sample. Problems 1 and 2.

1 = Reading; 2 = Drawing or summarizing; 3 = Recalling similar problems;

4 = Thinking about solutions; 5 = Calculations; 6 = Writing; 7 = Reviewing;

8 = Correcting mistakes; Plan = Planning phase; Exec = Execution phase;

Ev = Evaluation phase

Regarding the categories or sub-processes within each phase, in the Planning phase, "thinking about solutions" and "reading" were the most frequently reported categories, with mean percentages of 19.20% in Problem 1 and 17.02% in Problem 2 for the first category, and 15.98% and 13.57% respectively for the second category. "Recalling similar problems" was the least reported category or sub-process, although higher in the second problem (2.97% in Problem 1 and 6.67% in Problem 2). "Drawing or summarizing" represented 11.29% of the total categories reported in Problem 1 and 12.36% in Problem 2, being an important sub-process within planning. With respect to Execution phase, "mental calculation" was the category most frequently reported in this phase, with mean percentages of 28.37% in Problem 1 and 31.83% in Problem 2. "Writing" also represented an important percentage of the categories used, with a 12.09% frequency reported in Problem 1, and 14.14% in Problem 2. In relation to the Evaluation phase, "reviewing" and "correcting mistakes" were two of the least frequently reported categories by students in both tasks (6.36% in Problem 1 and 4.29% in Problem 2 for "reviewing"; 3.79% and 3.29% respectively for "correcting mistakes").

Finally, standard deviations in Table 2 showed a high variability among the processes shown by students in both tasks, mainly in the categories less reported ("recalling similar problems", "reviewing" and "correcting mistakes").

TABLE 2

Means *(M)* and Standard Deviations *(SD)* in process variables.
Total sample (N=510). Problems 1 and 2

	Problem 1		Problem 2	
Process variables	*M*	*SD*	*M*	*SD*
Reading	15.98	14.60	13.57	14.66
Drawing or summarizing	11.29	16.56	12.36	19.53
Recalling similar problems	2.97	6.67	2.46	6.53
Thinking about a solution	19.20	17.02	18.19	18.80
Calculations	28.37	22.67	31.82	25.10
Writing	12.06	12.07	14.14	14.27
Reviewing	6.36	9.25	4.29	7.73
Correcting mistakes	3.79	7.99	3.29	7.93
Planning phase	49.46	23.10	46.53	26.33
Execution phase	40.43	22.47	45.95	25.73
Evaluation phase	10.15	13.13	7.57	11.96

3.2. Process differences between groups with different achievement (success/failure)

Tables 3 and 4 present the mean scores and standard deviations in process variables for groups with different achievement (Group 1 = success; Group 2 = failure) in Tasks 1 and 2, respectively. Group compositions revealed the presence of a high proportion of students solving the mathematical problems unsuccessfully. Specifically, 356 students (69.80%) failed to solve the first problem while 327 students (60.11%) gave an incorrect answer on the second problem.

TABLE 3

Means *(M)* and Standard Deviations *(SD)* in process variables for each group. Problem 1

	Group 1 (success) N = 154		Group 2 (failure) N = 356	
Process variables	*M*	*SD*	*M*	*SD*
Reading	16.79	15.30	15.62	14.30
Drawing or summarizing	13.66	17.78	10.26	15.91
Recalling similar problems	3.37	7.44	2.80	6.31
Thinking about a solution	17.01	16.17	20.15	17.30
Calculations	26.62	22.18	29.12	22.86
Writing	12.76	11.26	11.76	12.40
Reviewing	5.88	8.31	6.56	9.63
Correcting mistakes	3.80	6.90	3.78	8.42
Planning phase	50.95	21.07	48.82	23.92
Execution phase	39.37	21.69	40.88	22.81
Evaluation phase	9.68	12.08	10.36	13.58

For the first problem, means in Table 3 indicated a trend to plan more, and to execute and review less, in the Group 1 (success). MANCOVA analyses showed the presence of statistically significant differences between the groups (*Wilks´λ* = .957; $F(11,496) = 2.012$; $p = .026$; $\eta p^2 = .043$). However, these differences were only found in “drawing or summarizing” ($F(1,506_{)} = 4.548$; $p = .033$; $\eta p^2 = .009$). As happened with planning, this sub-process was more frequently reported by students who successfully solved the task. Gender and grade level, which were included as covariates, did not generate any differences in the process ($p = .461$ and $p = .123$, respectively).

With respect to the second problem, the means in Table 4 showed a different pattern of results, with students who successfully solved the problem reporting using less sub-processes related to Planning, and more related to the Execution and Evaluation phases.

TABLE 4

Means *(M)* and Standard Deviations *(SD)* in process variables for each group. Problem 2

	Group 1 (success) N = 187		Group 2 (failure) N = 327	
Process variables	*M*	*SD*	*M*	*SD*
Reading	11.58	11.10	14.68	16.22
Drawing or summarizing	15.03	19.83	10.86	19.23
Recalling similar problems	1.55	4.37	2.97	7.43
Thinking about a solution	15.46	15.54	19.72	20.26
Calculations	34.58	23.92	30.27	25.65
Writing	14.14	11.28	14.14	15.72
Reviewing	4.68	7.17	4.07	8.03
Correcting mistakes	3.18	7.37	3.35	8.23
Planning phase	43.54	23.61	48.21	27.63
Execution phase	48.67	23.42	44.43	26.86
Evaluation phase	7.86	11.36	7.41	12.30

MANCOVA indicated the presence of statistically significant differences between groups in the process (*Wilks´λ* = .951; $F(11,496) = 2.303$; $p = .009$; $\eta p^2 = .049$). These differences were found in those categories or sub-processes within the Planning phase: "reading" ($F(11,506) = 5.499$; $p = .019$; $\eta p^2 = .011$), "drawing or summarizing" ($F(11,506) = 4.604$; $p = .032$; $\eta p^2 = .009$), "recalling similar problems" ($F(11,506) = 5.209$; $p = .023$; $\eta p^2 = .010$) and "thinking about solutions" ($F(11,506) = 5.387$; $p = .021$; $\eta p^2 = .011$). Differences in the Planning phase were close to statistical significance ($F(11,506) = 3.825$; $p = .051$; $\eta p^2 = .006$). Means revealed that students who successfully solved the task (Group 1) reported reading, recalling similar problems, and thinking about solutions less frequently than the other group. As with the first problem, successful students reported using representation strategies (i.e., drawing or summarizing) more frequently than their peers. Gender and level did not generate differences in the process ($p = .590$ and $p = .230$, respectively).

4. DISCUSSION AND CONCLUSIONS

It has been argued that it is essential to identify economic, fast, reliable, and valid techniques to elicit and analyse cognitive processes involved in different cognitive tasks (Ifenthaler, 2008, Ifenthaler, Masduki, & Seel, 2011). In this context, the main aim of this study was to analyze the usefulness of a new assessment tool based on the Triple Task procedure and designed to assess the metacognitive and self-regulated processes shown by the 510 fifth and sixth grade students during the performance of two mathematical problems. Specifically, this study tried to answer these questions:

Do students follow Self-regulated Learning stages while they try to solve mathematical problems? Global results indicated the presence of important differences among processes followed by students, as indicated by the high standard deviations. This variability was even higher in the categories or sub-processes less frequently reported by students in general: recalling similar problems, reviewing and correcting mistakes. The pattern of results was similar in both tasks, and was characterized by a lack of evaluation strategies. However, sub-processes related to Planning and Execution phases were the most frequently reported, respectively. These result may seem to contradict some previous research pointing out that student commonly tend to be impulsive in problem-solving situations, spending little time making a plan to develop (Cleary & Chen, 2009; Kramarski & Gutman, 2006; Montague, 2008; Montague et al., 2011; Pennequin et al, 2010; Pereis et al., 2009).

However, despite the fact that these results could indicate that once students properly planned an activity, making mistakes became less likely and reviewing activities less necessary, students' achievements in both tasks revealed a high proportion of failure. This probably means that planning strategies were not effective enough. In fact, an analysis of the sub-processes involved in the planning phase revealed that thinking about solutions and reading were the most frequently reported sub-processes within planning, while the use of potentially more useful strategies such as information representation and organization were less frequent among students. This result could indicate a trend to use passive solving strategies or even comprehension difficulties.

Are there differences in the processes showed by students with different achievements in the tasks? Analyses revealed the existence of statistically significant differences between groups. These differences were located in the sub-process of drawing or summarizing in Problem 1, while significant differences were found in all the sub-process within the planning phase in Problem 2. The existence of more significant differences in this second problem could be related to its characteristics (i.e., more complex relationships to stablish, parameters to calculate, etc.). Thus, this task may result more difficult, suggesting a more important role of planning strategies when problem difficulty and demands increase. Regarding the pattern of differences, students who successfully solved the tasks tended to draw and summarize more, but they recalled similar problems, read and thought about solutions proportionally less than those who failed to solve the tasks. These results are consistent with the statement above about the effectiveness of planning strategies, and suggest the use of more active processes by students with high achievement. In this sense, the strategy that showed to be effective in both tasks was drawing or summarizing (i.e. using information representation and organization strategies). Making this kind of representations would correspond to the "translation" process, characterized as an essential part of solving mathematical problems by many authors (Abdullah, Zakaria, & Halim, 2012; Csíkos, Szitányi, & Keleme, 2012; Díez-Palomar, Menéndez, & Civil, 2011; Pantziara, Gagatsis, & Elia, 2009; Stylianou, 2011). An example of graphical representation, made by a participant in this study is shown in Figure 2.

The use of graphical representations supposes an important part of planning strategies. As Montague et al. (2011) argued, *"problem representation involves translating and transforming linguistic and numerical information into verbal, graphic, symbolic, and quantitative representations that show the relationships among the problem parts prior to generating appropriate mathematical equations or algorithms for problem solution"* (p. 263). In fact, Abdullah et al. (2012) showed in their study that when the teaching approaches encourage students to apply thinking strategies through using visual representation, students are able to gain a better conceptual understanding and eventually improve their mathematics achievement. They attributed this effect to a more active involvement of students in their learning process. However, representation is not only a matter of copying what one sees. As the example above shows, it involves a process of personal re-organization, inventing or adapting conventions of a representational system for the purpose at hand (Stylianou, 2011).

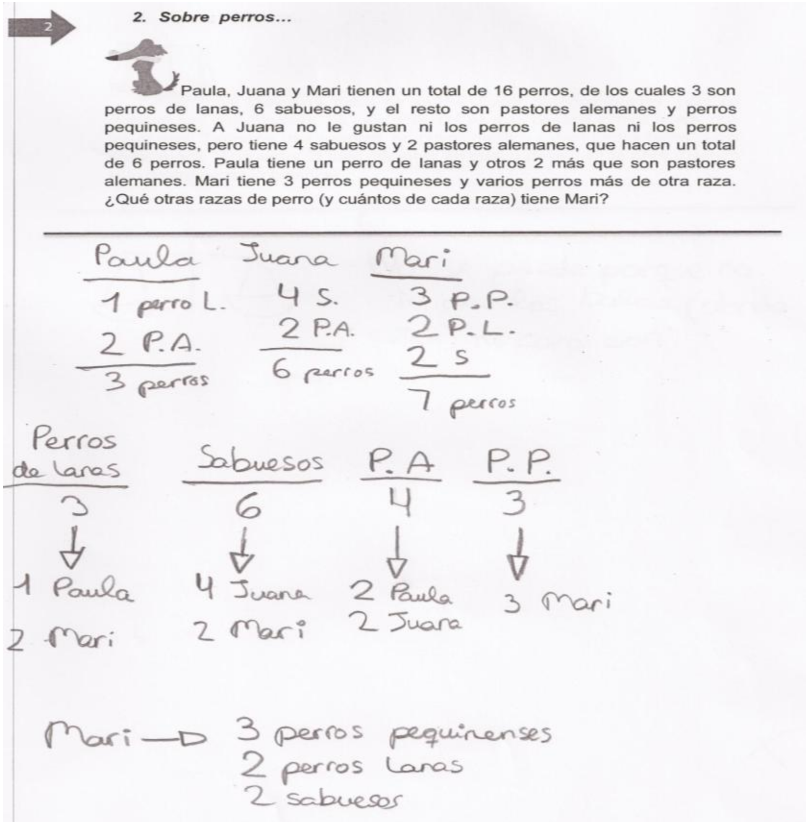


Figure 2. Sample of student answer in the second problem.

It presents an example of information representation and organization strategies

In summary, the results found in this study would seem to contradict those studies that indicated that students do not plan, jumping immediately to calculations without considering alternatives to solve problems or giving impulsive responses (Cleary & Chen, 2009; Kajamies et al., 2010; Kramaski & Gutman, 2006). In fact, students in this study reported having planned. However, an in-depth analysis of the sub-process involved in planning revealed the presence of ineffective planning strategies implemented by the group in general, which was translated into a greater percentage of failure in both problems. In this sense, another important finding of this study is that students have serious difficulties to evaluate their progress. They also showed a tendency to perform calculations as a method to solve the problems. Thus, students in this study did not self-regulate or implement effective metacognitions during problem solving. However, differences between groups with different achievement-levels suggested that, despite the fact that students did not show a properly self-regulated process, a proportion of them were able to use more effective and active strategies while solving the problems, such as organizing or representing the information. This is in fact an important determinant in successful mathematical problem solving.

Implications

These findings add evidence to the utility of process measures, specifically those based on directed and concurrent self-reporting, in the assessment of cognitive activities (in this case mathematical problem-solving tasks). The usefulness of the assessment tools presented in this study lies, however, in the systems of categories designed. This system makes it possible to divide the mathematical problem solving process into sub-processes or isolated activities, and provides in-depth insights about a student's strengths or weaknesses, thereby helping to understand their success or failure from the point of view of SRL. This is not only important for instruction, but specially for student learning. In this sense, as Ifenthaler (2012) pointed out, the key link between knowledge about (and the regulation of) one's own problem-solving activities may lie in reflective thinking. Making students aware of their own learning processes may help them to generate information about the efficiency of their problem-solving strategies and successfully implement that knowledge in the ongoing problem-solving process, thereby being able to control and regulate their cognition, effort and behavior.

Limitations

It is necessary to acknowledge the following limitations in the present study: First, as self-regulation is a cyclical process (Zimmerman, 2000), analyzing the recursion of the process followed by students may bring useful information about its temporal course and the metacognitive mechanisms involved. In this sense, identifying different patterns of performance may also help to reduce the high variability whithing-subjets found in this study. Second, motivational and affective components, such as self-efficacy beliefs, task interest and perceived instrumentality, have been demonstrated to play an important role in mathematical performance and problem solving situations (Ahmed, Minnaert, Kuyper, & Van der Werf, 2012; Cueli, García, & González-Castro, 2013; Dettmers, Trautwein, Lüdtke, Goetz, Frenzel, & Pekrun, 2011; Hoffman, 2010). The study conducted by Ahmed et al. (2012) revealed that a good students´ performance in addition and subtraction tasks was related not only to the student's use of advanced mathematics strategies and SRL competence, but also to the ability attribution for success, effort attribution for failure, and high perceived self-efficacy when using specific strategies. Third, it would be interesting to extend the number of mathematical problems, also adding different levels of difficulty, which would provide more accurate information about the influence of task difficulty in problem-solving processes. Additionally, in order to provide support to the usefulness of the tool applied in this study as a measure of self-regulatory processes involved in mathematical problem solving, analyzing its relationship with other standardized instruments such as questionnaires and interviews would be useful. Complementary use of both measures would provide a better understanding of self-regulatory mechanisms exhibited by the students, thus helping to design more adapted instructional strategies. Finally, and given the characteristics of the sample (students belonged to different schools in a specific area in Northern Spain), extending the research to other areas would be advisable in order to better stablish the scope of our findings. This is, however, the first approach to the study of the usefulness of the Triple Task technique in the assessment of the process involved in solving mathematical problems. Therefore, the results from the present study actually open the way for future research. In this sense, it would be interesting, for example, to conduct multi level modeling that would take in to account school level effects.

References

Abdullah, N., Zakaria, E. & Halim, L. (2012). The effect of a thinking strategy approach through visual representation on achievement and conceptual understanding in solving mathematical word problems. *Asian Social Science, 8*(16), 16-340. doi: 10.5539/ass.v8n16p30

Ahmed, W., Minnaert, A., Kuyper, H. & Van der Werf, G. (2012). Reciprocal relationships between math self-concept and math anxiety. *Learning and individual differences, 22,* 385-389.

Azevedo R. & Cromley, J.G. (2004) Does training of self- regulated learning facilitate student's learning with hypermedia? *Journal of Educational Psychology, 96*, 523-535.

Bransford, J. D. & Stein, B. S. (1993). *The ideal problem solver: A guide for improving thinking, learning and creativity (2nd Ed.).* New York, NY: W.H. Freeman.

Butler D. L. & Cartier S.C. (2005). *Multiple complementary methods for understanding self-regulated learning as situated in context.* Accepted for presentation at the April 2005 Annual Meetings of the American Educational Research Association, Montreal, QC.

Cleary, T. J. & Chen, P. P. (2009). Self-regulation, motivation, and math achievement in middle school: Variations across grade level and math context. *Journal of School Psychology, 47,* 291-314.

Csíkos, C., Szitányi, J. & Keleme, R. (2012). The effects of using drawings in developing young children's mathematical word problem solving: A design experiment with third-grade Hungarian students. *Educational Studies in Mathematics, 81,* 47-65. doi: 10.1007/s10649-011-9360-z

Cueli, M., García, T. & González-Castro, P. (2013). Autorregulación y rendimiento académico en Matemáticas. [Self-regulation and academic achievement in Mathematics]. *Aula Abierta, 41*(1), 39-48.

Dettmers, S., Trautwein, U., Lüdtke, O., Goetz, T., Frenzel, A. & Pekrun, R. (2011). Students' emotions during homework in mathematics: Testing a theoretical model of antecedents and achievement outcomes. *Contemporary Educational Psychology, 36*, 25-35. doi: 10.1016/j.lindif.2011.04.006

Díez-Palomar, J., Menéndez, J. M. & Civil, M. (2011). Learning mathematics with adult learners: drawing from parents´ perspective. *Revista Latinoamericana de Investigación en Matemática Educativa, 14*(1), 71-94.

Fidalgo, R., Torrance, M., Robledo, P. & García, J. N. (2009). Dos enfoques metacognitivos de intervención: auto-conocimiento del producto textual frente a auto-regulación del proceso de escritura. [Two approaches to metacognitive intervention: metalinguistic knowledge of the written product compared to self-regulation of the writing process]. *International Journal of Developmental and Educational Psychology, 1*(1), 303-312.

Gálvez, G., Cosmelli, D., Cubillos, L., Leger, A. M., Mena, A., Tanter, E., … & Soto-Andrade, J. (2011). Estrategias cognitivas para el cálculo mental. [Cognitive strategies for mental calculation]. *Revista Latinoamericana de Investigación en Matemática Educativa, 14*(1), 9-40.

García, T. & González-Pienda, J. A. (2012). Evaluación del proceso de Aprendizaje Autorregulado en el área de las Matemáticas mediante pizarras digitales. [Assessment of the self-regulated process in mathematics through digital whiteboards]. In J. Dulac & C. Alconada (Coord.), *III Congreso Pizarra Digital. Publicación de comunicaciones* (pp.105-117). Madrid: Pluma y Arroba.

García, J. N., Rodríguez, C., Pacheco, D. & Diez, C. (2009). Influence of cognitive effort, sustained attention, working memory and ADHD symptoms in the process and product of written composition. An experimental study. *Estudios de Psicología, 30*(1), 31-50. doi: 10.1174/021093909787536326

Geary, D. C. (2004). Mathematics and learning disabilities. *Journal of Learning Disabilities, 37,* 4-15.

Greene, J. A. & Azevedo, R. (2010). The measurement of learners' Self-Regulated cognitive and metacognitive processes while using Computer-Based Learning Environments. *Educational Psychologist, 45*(4), 203-209. doi: 10.1080/00461520.2010.515935

Hoffman, B. (2010). "I think I can, but I'm afraid to try": The role of self-efficacy beliefs and mathematics anxiety in mathematics problem solving efficiency. *Learning and Individual Differences, 20*(3), 276-283. doi: 10.1016/j.lindif.2010.02.001

Ifenthaler, D. (2008). Practical solutions for the diagnosis of progressing mental models. In D. Ifenthaler, P. Pirnay-Dummer & J. M. Spector (Eds.), *Understanding models for learning and instruction*. *Essays in honor of Norbert M. Seel* (pp. 43-61). New York, NY: Springer.

Ifenthaler, D., Masduki, I. & Seel, N. M. (2011). The mystery of cognitive structure and how we can detect it. Tracking the development of cognitive structures over time. *Instructional Science, 39*(1), 41-61. doi: 10.1007/s11251-009-9097-6

Ifenthaler, D. (2012). Determining the effectiveness of prompts for self-regulated learning in problem-solving scenarios. *Educational Technology & Society, 15*(1), 38-52.

Jarero, M., Aparicio, E. & Sosa, L. (2013). Pruebas escritas como estrategia de evaluación de aprendizajes matemáticos. Un estudio de caso a nivel superior. [Written test as a strategy assessment learning mathematics. A case study of a higher level]. *Revista Latinoamericana de Investigación en Matemática Educativa, 16* (2), 213-243. doi: 10.12802/relime.13.1623

Kajamies, A., Vauras, M. & Kinnunen, R. (2010). Instructing low-achievers in mathematical word problem solving. *Scandinavian Journal of Educational Research, 54*(4), 335-355. doi: 10.1080/00313831.2010.493341

Kramarski, B. & Gutman, M. (2006). How can self-regulated learning be supported in mathematical E-learning environments? *Journal of Computer Assisted Learning, 22*, 24-33. doi: 10.1111/j.1365-2729.2006.00157.x

Lazakidou, G. & Retalis, S. (2010). Using computer supported collaborative learning strategies for helping students acquire self-regulated problem-solving skills in mathematics. *Computers & Education, 54,* 3-13. doi: 10.1016/j.compedu.2009.02.020

Montague, M. (2008). Self-Regulation and Mathematics Instruction. *Learning Disabilities Research & Practice*, *22(*1), 75-83. doi: 10.1111/j.1540-5826.2007.00232.x

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. doi: 10.1177/0731948711421762

Olive, T., Kellogg, R. T. & Piolat, A. (2002). The Triple Task technique for studying the processes of writing: why and How? In G. Rijlaarsdam (Series Ed.), Studies in Writing & T. Olive, & C. M. Levy (Vol. Eds.), *Contemporary tools and techniques for studying writing* (pp. 31-59). Dordrecht: Kluwer Academic Publishers.

Ostad, A. & Sorenson, P. M. (2007). Private speech and strategy- use patterns: Bidirectional comparisons of students with and without mathematical disabilities in a developmental perspective. *Journal of Learning Disabilities, 40*, 2-14.

Pantziara, M., Gagatsis, A. & Elia, I. (2009). Using diagrams as tools for the solution of non-routine mathematical problems. *Educational Studies in Mathematics, 72*, 39-60. doi: 10.1007/s10649-009-9181-5

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 and Reasoning, 16*(3), 198-220.

Pereis, F., Dignath, C. & Schmitz, B. (2009). Is it possible to improve mathematical achievement by means of self-regulation strategies? Evaluation of an intervention in regular math classes. *European Journal of Psychology of Education, 24*(1), I7-29.

Piolat, A., Kellogg, R. T. & Farioli, F. (2001). The Triple Task technique for studying writing processes: on which task is attention focused? *Current Psychology Letters. Brain, Behavior and Cognition, 4,* 67-83.

Piolat, A. & Olive, T. (2000). Comment étudier le coût et le déroulement de la rédaction de textes? La méthode de la triple-tâche: Un bilan méthodologique [How to study the cost and the unfolding of writing? The method of Triple-Task: A methodological review]. *L'Anneé Psychologique, 100,* 465-502.

Piolat, A., Olive, T. & Kellogg, R. T. (2005). Cognitive effort during note taking. *Applied Cognititive Psychology, 19*, 291-312.

Piolat, A., Barbier, M. L. & Roussey, J. Y. (2008). Fluency and cognitive effort during first and second-language note taking and writing by undergraduate students. *European Psychologist, 13*(2), 114-125. doi: 10.1027/1016-9040.13.2.114

Rosário, P., Mourão, R., Núñez, J. C., González-Pienda, J. A. & Solano, P. (2008). Storytelling as a promoter of Self-Regulated Learning (SRL) throughoutschooling. In A. Valle, J.C. Núñez, R.G. Cabanach, J.A. González-Pienda & S. Rodríguez (Eds.), *Handbook of instructional resources and their applications in the classroom* (pp. 107-122). New York, NY: Nova Science.

Rosenzweig. C., Krawec, J. & Montague, M. (2011). Metacognitive strategy use of eighth-grade students with and without learning disabilities during mathematical problem solving: A Think-Aloud analysis. *Journal of Learning Disabilities, 44*(6), 508-520. doi: 10.1177/0022219410378445

Schmitz, B. & Perel, P. (2011). Self-monitoring of self-regulation during math homework behaviour using standardized diaries. *Metacognition and Learning, 6,* 255-273. doi: 10.1007/s11409-011-9076-6

Shea, P. & Bidjerano, T. (2010). Learning presence: Towards a theory of self-efficacy, self-regulation, and the development of a communities of inquiry in online and blended learning environments. *Computer & Education, 55*(4), 1721-1731. doi: 10.1016/j.bbr.2011.03.031

Stylianou, D. A. (2011). An examination of middle school students' representation practices in mathematical problem solving through the lens of expert work: towards an organizing scheme. *Educational Studies in Mathematics, 76,* 265-280.

Throndsen, I. (2011). Self-regulated learning of basic arithmetic skills: A longitudinal study. *British Journal of Educational Psychology, 81,* 558-578**.** doi: 10.1348/2044-8279.002008

Torrance, M. & Galbraith, D. (2006). The processing demands of writing. In MacArthur, C., Graham, S. & Fitzgerald, J. (Ed.), *Handbook of Writing Research* (pp. 67-80). New York, NY: The Guilford Press.

Veenman M. V .J. (2005). The assessment of metacognitive skills: what can be learned from multi-method designs? In B. Moschner & C. Artelt (Eds.), *Lernstrategien und Metakognition: Implikationen fur Forschung und Praxis* (pp. 75-97). Berlin: Waxmann.

Veenman, M. V. J. (2011). Learning to self-monitor and self-regulate. In R. Mayer & P. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 197-218). New York, NY: Routledge.

Verschaffel, L., Greer, B. & De Corte, E. (2000). *Making sense of word problems*. Lisse: Swets & Zeitlinger.

Voyer, D. (2011). Performance in mathematical problem solving as a function of comprehension and arithmetic skills. *Journal of Science and Mathematics Education, 9,* 1073-1092.doi: 10.1007/s10763-010-9239-y

Whimbey, A. & Lochhead, J. (1993). *Comprender y resolver problemas [Problem-solving and comprehension]*. Madrid: Visor

Williams, J. R. (2008). Revising the Declaration of Helsinki. *World Medical Journal*, *54*, 120-125.

Zimmerman, B. J. (2000). Attaining self-regulation. A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13-39). San Diego, CA: Academic Press.

Zimmerman, B. J. & Schunk, D. H. (2008). Motivation: An essential dimension of self-regulated learning. In D. H. Schunk & B. J. Zimmerman (Eds.), *Motivation and self-regulated learning. Theory, research, and applications* (pp. 1-30). New York, NY: Routledge.

Psychometric Characteristics of the BRIEF Scale for the Assessment of Executive Functions in Spanish Clinical Population

Trinidad García, Julio Antonio González-Pienda, Celestino Rodríguez, David Álvarez-García & Luis Álvarez

University of Oviedo (Spain)

Psicothema
Psicothema 2014, Vol. 26, No. 1, 47-52
doi: 10.7334/psicothema2013.149
ISSN 0214 - 9915 CODEN PSOTEG

www.psicothema.com

Psychometric characteristics of the BRIEF scale for the assessment of executive functions in Spanish clinical population

Trinidad García Fernández, Julio Antonio González-Pienda, Celestino Rodríguez Pérez, David Álvarez García and Luis Álvarez Pérez
Universidad de Oviedo

Abstract

Background: The Behavior Rating Inventory of Executive Functions (BRIEF) scale, completed by families, is widely known in the assessment of executive functions in children and adolescents. However, its application is limited to English-speaking population. **Method:** This study analyzes the preliminary results from its application in a Spanish clinical sample, comprising 125 participants aged 5-18 years. Internal structure and reliability of the translated scale were analyzed, as well as its relationship with other behavioral measures through the analysis of their correlations with the Assessment of Attention Deficit Hyperactivity Disorder Scale (EDAH). The results were compared with those from the original validation study. **Results:** The data revealed the presence of the same internal structure, as well as acceptable internal consistency and significant correlations with the Attention Deficit and Hyperactivity components of the EDAH scale. **Conclusions:** This study provides preliminary evidence of the utility of the BRIEF scale in cultural contexts different from the original, particularly in Spanish clinical population.

Keywords: BRIEF, Spanish, executive functions, family.

Resumen

***Características psicométricas de la escala BRIEF para la evaluación de funciones ejecutivas en población clínica española*. Antecedentes:** la escala Behavior Rating Inventory of Executive Functions (BRIEF), cumplimentada por familias, es ampliamente conocida en la evaluación de las funciones ejecutivas en niños y adolescentes. Sin embargo, su aplicación está limitada a población de habla inglesa. **Método:** en este estudio se analizan los primeros resultados procedentes de su aplicación a una muestra clínica española, formada por 125 participantes de 5 a 18 años. Se analizó la estructura interna y fiabilidad de las puntuaciones de la escala traducida al español, así como su relación con otras medidas comportamentales a través del análisis de sus correlaciones con la escala de Evaluación del Déficit de Atención con Hiperactividad (EDAH). Los resultados se compararon con los del estudio de validación original. **Resultados:** los datos mostraron la presencia de una misma estructura interna de las puntuaciones, así como una aceptable consistencia interna y correlaciones estadísticamente significativas con los componentes de Déficit de Atención e Hiperactividad de la escala EDAH. **Conclusiones:** este estudio aporta evidencia preliminar sobre la utilidad de la escala BRIEF en contextos culturales diferentes al originario, concretamente en población clínica española.

Palabras clave: BRIEF, español, funciones ejecutivas, familia.

The term *executive functions* refers to a set of skills involved in the generation, supervision, regulation, execution, and readjustment of behavior to achieve complex goals, especially those that require a novel and creative approach (Verdejo-García & Bechara, 2010).

Executive functions comprise different processes, such as goal setting, hypothesis formulation, planning, focal attention, concentration, strategy generation, monitoring, response to feedback, the capacity to solve problems, abstract thinking, cognitive flexibility, working memory, and emotional control (Korzeniowski, 2011). Alterations in these processes during childhood and adolescence cause difficulties in an extensive range of domains, including cognitive, behavioral, and social problems.

In this regard, many studies have shown the close relationship between the presence of executive deficits and specific problems in concrete areas such as reasoning, mathematics, reading, and writing (García, Rodríguez et al., 2013; Latzman, Elkovitch, Young, & Clark, 2010; Raghubar, Barnes, & Hecht, 2010; Stelzer & Cervigni, 2011; Van der Ven, Kroesbergen, Boom, & Leseman, 2013). Numerous studies have also shown how the inadequate development of these processes is related to difficulties in the individual's adaptation to the environment (Brock, Rimm-Kaufman, Nathanson, & Grimm, 2009; Garon, Bryson, & Smith, 2008; McClelland et al., 2007).

The study of the executive functions has also been particularly useful in the differential diagnosis of numerous childhood and adolescent disorders, mainly Attention Deficit with Hyperactivity Disorder (ADHD; Di Trani et al., 2011; Hale et al., 2011; Makris, Biederman, Monuteaux, & Seidman, 2009; Shimoni, Engel-Yeger,

Received: May 22, 2013 • Accepted: September 5, 2013
Corresponding author: Celestino Rodríguez Pérez
Facultad de Psicología
Universidad de Oviedo
33003 Oviedo (Spain)
e-mail: rodriguezcelestino@uniovi.es

& Tirosh, 2012), Autistic Spectrum Disorders (Han et al., 2011; Kenworthy, Black, Harrison, Della Rosa, & Wallace, 2009), and other clinical conditions (Donders, Den Braber, & Vos, 2010; Neri et al., 2012; Vințan, Palade, Cristea, Benga, & Muresanu, 2012; Wilson, Donders, & Nguyen, 2011).

However, although nowadays the relevance of the executive functions is clear in research, their assessment is often complex. This complexity is partially due to the type of measurement instruments employed, as most of them are performance-based. For some authors, this type of measures, such as the Stroop Test, the Tower of Hanoi or the Wisconsin Card Sorting Test, present low ecological validity and are not representative of the individual's functioning in real life settings (Gioia, Kenworthy, & Isquith, 2010; Henry & Bettenay, 2010; Lezak, Howieson, Bigler, & Tranel, 2012; Toplak, Bucciarelli, Jain, & Tannock, 2009; Verdejo-García & Bechara, 2010). These tasks, which are excessively structured and based on quantitative criteria, disregard much relevant information about children and adolescents' daily functioning, such as the type of strategies they employ to solve problems, their capacity to plan and to recall certain rules or guidelines, to inhibit behaviors or impulses, or to adapt to new settings or situations.

An alternative to this type of measures is the use of questionnaires based on the observation of the behavior, such as the BRIEF (*Behavior Rating Inventory of Executive Function;* Gioia, Isquith, Guy, & Kenworthy, 2000). There are several hetero-report versions of this questionnaire. The most frequently employed is based on information provided by the family. It is applicable from ages 5 to 18 years and it assesses the frequency with which children and adolescents display certain problematic behaviors related to deficits in executive functions at home and/or in school. For this purpose, it uses a Likert-type response format ranging from 1 to 3, where 1 is never, 2 is sometimes, and 3 is often. The scale is made up of 86 items (72 computable items and 14 additional ones). The latter do not contribute to the score but are useful to orient possible interventions. The remaining 72 items form 8 scales: Inhibit, Shift, Emotional Control, Initiate, Working Memory, Plan/Organize, Organization of Materials, and Monitor, which in turn are grouped into two main indexes: the *Behavioral Regulation Index* (BRI), made up of the first three scales, and the *Metacognition Index* (MI), made up of the remaining five. Both indexes made up the *Global Executive Composite* (GEC) score. High scores in these scales and indexes indicate executive deficit. It also includes two validity scales to identify problematic response styles. These scales are based on the analysis of the excessive frequency assigned to certain items (Negativity) or the correspondence between pairs of items (Inconsistency) and are not submitted to statistical analysis.

This instrument has shown its utility for the assessment of executive functions in a broad range of clinical conditions (Anderson & Reidy, 2012; Lee et al., 2011; Toplak et al., 2009; Wilson et al., 2011), especially in ADHD (Jarratt, Riccio, & Siekierski, 2005; McCandless & O'Laughlin, 2007; Toplak et al., 2009). In this regard, numerous studies have analyzed the executive profile in ADHD and its subtypes, as well as in comorbidity with other disorders such as Reading Difficulties (RD). These studies have pointed to response inhibition and working memory as key issues in the differentiation of ADHD, finding greater impairment of response inhibition in groups with hyperactivity/impulsivity and more impairment of working memory in groups with inattention (García et al., in press; Gioia et al., 2000; McCandless & O'Laughlin, 2007; Riccio, Homack, Jarratt, & Wolfe, 2006), as well as differential executive profiles in groups with ADHD and ADHD with associated RD. Specifically, García et al. (2013) found greater executive deficit in the group with ADHD and associated RD than in the group with isolated ADHD, mainly in working memory and plan, coherent with previous studies carried out by Pratt (2000) with the BRIEF scale, as well as by Bental and Tirosh (2007), Van De Voorde, Roeyers, Verté and Wiersema (2010), and Willcutt et al. (2010) with performance-based tests.

However, despite being one of the most extensively employed scales, its use has been limited to English-speaking countries, except for a new version available in Dutch (Huizinga & Smidts, 2011; Smidts & Huizinga, 2009). Therefore, to determine its potential utility beyond its original context, we translated the scale into Spanish.

This aim of this work is to address the first analysis of the internal structure of the BRIEF scores, and also to obtain evidence of their reliability and the degree of its association with other behavioral measures, specifically the scale "Evaluación del Déficit de Atención con Hiperactividad" (EDAH; Farré & Narbona, 1997; in English, Attention Deficit with Hyperactivity Assessment). For this purpose, we conducted a non normative study with a heterogeneous clinical sample of 125 Spanish children and adolescents, aged from 5 to 18 years, whose families completed the translated scale. The results obtained were contrasted with those of the original validation study in a clinical sample (Goia et al., 2000).

Method

Participants

We used a non probabilistic clinical sample, made up of 125 participants (*range* = 5-18, M = 12.68, SD = 5.22), 54 females (43%) 71 and males (57%). The main clinical groups were: ADHD (N = 112, 89%), Intellectual Disability (N = 27, 21.6%), RD (N = 63; 50.4%), Anxiety (N = 49, 39.5%), and Emotional Maladaptation (N = 12, 9.6%). The participants were initially identified by the Pediatric Unit following the criteria of the *Diagnostic and Statistical Manual of Mental Disorders-IV-TR* (American Psychiatric Association [APA], 2002), and referred to a clinic for more extensive assessment. The final diagnosis was based on the information provided by the interviews, behavior rating questionnaires administered to the families and children, as well as diverse psychometric and neuropsychological tests. Assessment included the semi-structured *Diagnostic Interview Schedule for Children* (DISC-IV; Shaffer, Fisher, Lucas, Dulcan, & Schwab, 2000) adapted to Spanish, and also data from the developmental history, direct observation of the child, and the prior neuropediatric examination.

The IQ of the sample, assessed with the Wechsler Intelligence Scale for Children (WISC-IV; Wechsler; 1974/2005), included scores ranging from 54 to 125 (M = 92.58, SD = 13.75). Some participants presented various associated disorders. No participant was receiving medication at the time of assessment.

Instruments

The main assessment instrument and object of research was the BRIEF scale (Gioia et al., 2000) in its form for families translated into Spanish (hereafter, BRIEF-E). In view of the importance of adapting linguistic and cultural aspects during the translation process, and following the guidelines on the use of psychological and educational tests (Elosúa, 2003) and those of the International Test Commission (ITC), described by Muñiz, Elosua, and Hambleton (2013), this process began with two independent translations carried out by members of the research team. The translations were then reviewed by a committee consisting of three translators specialized in the field of Psychology and assessment. After reviewing the translated version and making some linguistic corrections, prior to the study, the scale was administered to a reduced sample of families to verify their comprehension of the statements.

We also used the "Evaluación del Déficit de Atención con Hiperactividad" (EDAH) scale (Farré & Narbona, 1997). This scale assesses the presence of a series of behaviors related to the symptoms of hyperactivity and attention deficit. Like the BRIEF, it was administered in the family form.

Procedure

The study was conducted according to The Helsinki Declaration of the World Medical Association (available in Williams, 2008), which presents the ethical principles for research with human beings. All the participants' families gave written informed consent after receiving a complete description of the study. After an initial interview, the families completed the EDAH and BRIEF scales. The children and adolescents underwent a broad psychoeducational assessment, performed at a specialized clinical center, during two 30-minute sessions that took place in the afternoon. Participation in the study was voluntary, and the anonymity and ethical treatment of the data were guaranteed. The participants did not receive any incentive for their collaboration.

Administration of the BRIEF scale was supervised by a member of the research team, and these scores were not taken into account to establish the diagnosis. After obtaining the data, they were analyzed and compared with the results provided by the study with the original English version of the test. Following the authors' recommendations, scales with more than two missing values on the same subscale were excluded from the study, whereas the remaining missing values were substituted by the value 1 (Gioia et al., 2000, p. 7). Although in the original work, the authors' recommend treating reports with scores higher than 7 on the Negativity scales and higher than 9 on Inconsistency with caution (Gioia et al., 2000, pp. 10-15), they were also eliminated from this study.

Data analysis

Firstly, to obtain evidence of the internal structure of the questionnaire scores, as in the original study, we carried out Principal Component Analysis (PCA), selecting the components with Eigenvalues higher than 1. In view of the probability of correlations among components, we used direct Oblimin as the rotation method (Ferrando & Anguiano-Carrasco, 2010). Next, we analyzed the internal consistency of the scores for each subscale and index with Cronbach's alpha. Lastly, to explore convergent validity with other variables, we calculated the Pearson correlations among the subscales and indexes of the BRIEF scale and the EDAH (Farré & Narbona, 1997). All the analyses were performed with the SPSS-17.0 software.

Results

Analysis of the internal structure

Table 1 presents the results of the analyses of the internal structure of the scores carried out with the original version of the BRIEF scale and with the version of the present study (BRIEF-E).

In both studies, the same two-component structure was obtained: the first component (Metacognition) is made up of Initiate, Working Memory, Plan/Organizing, Organization of Materials, and Monitor, and the second one (Behavioral Regulation) is made up of Inhibit, Shift, and Emotional Control. These components explain 76% of the cumulative variance in the BRIEF, versus 62.48% in the BRIEF-E. Although in the original study with the BRIEF, these data were not provided, in this study with the BRIEF-E, the first component explained 43.35% of the variance, and the second one explained 19.13%. The correlations between the components were .74 in the BRIEF versus .35 in the BRIEF-E.

Regarding the weight of each subscale on its component, a similar pattern was obtained although some differences between the studies were also found. Thus, in both scales, Plan/Organizing and Working Memory were the subscales with the highest weight on the Metacognition component (for Plan/Organize, .96 in the BRIEF versus .87 in the BRIEF-E, and for Working Memory (.81 versus .82, respectively). Regarding the Behavioral Regulation component, the variable with the highest weight on the BRIEF was Emotional Control (.93), whereas Inhibit (.90) obtained the highest weight on the BRIEF-E, followed by Emotional Control (.89).

Analysis of internal consistency

The internal consistency coefficients (Cronbach's alpha) for the subscales and indexes of the BRIEF and BRIEF-E are presented in Table 2. The coefficients were generally higher in the BRIEF,

Table 1
Results of the analysis of the internal structure of the BRIEF and BRIEF-E scores

	BRIEF		BRIEF-E	
Subscales and indexes	**Component 1**	**Component 2**	**Component 1**	**Component 2**
Inhibit		.68		.90
Shift		.77		.56
Emotional control		.93		.89
Initiate	.71		.78	
Working memory	.81		.82	
Plan/Organize	.96		.87	
Organization of materials	.68		.54	
Monitor	.58		.70	
Correlation between factors	.71		.35	
% Explained variance			43.35	19.13
% Cumulative variance	76%		62.48%	

Note: Component 1 = Metacognition; Component 2 = Behavioral Regulation. Clinical Sample: BRIEF (N = 852), BRIEF-E (N = 125) albeit following the same tendency. Thus, Initiate obtained the lowest values in both cases (α = .82 in BRIEF versus .57 in BRIEF-E), and the highest values were found in the GEC (α = .98 versus .92, respectively).

The alpha coefficients obtained were very similar to those of the original study for Emotional Control and Organization of Materials (α = .92 in the BRIEF versus .89 in the BRIEF-E for Emotional Control, and .88 versus .85, respectively, for Organization of Materials).

Regarding the main indexes (BRI and MI) and the GEC, the Cronbach alpha coefficients of the BRIEF-E were high, around .90 in all cases, although lower than those of the BRIEF.

Convergent validity with other variables (EDAH Scale)

Although, in the original study, the ADHD Rating Scale-IV (ADHD-IV; DuPaul, Power, Anastopoulos, & Reid, 1996) was used, in this case, we administered the EDAH scale (Farré & Narbona, 1997). However, both scales assess behavioral problems related to attention deficit and hyperactivity.

As seen in Table 3, in both studies, statistically significant correlations were found between the BRIEF and BRIEF-E subscales and the ADHD-IV and EDAH scales, respectively, although with some differences.

Firstly, in the case of BRIEF-E, Inhibit, Shift, Emotional Control, and the BRI were significantly related only to the Hyperactivity subscale of the EDAH, whereas in the original study, they were related to both the ADHD-IV subscales, although more so to Hyperactivity. In both studies, the Inhibit subscale had the highest correlation with Hyperactivity (.73 in the BRIEF versus .71 in the BRIEF-E), followed by the BRI (.70 versus .68, respectively).

Secondly, Initiate, Working Memory, Plan/Organize, Organization of Materials, Monitor, and the MI all correlated significantly with both subscales, although more so with Attention Deficit in both studies. Organization of Materials was the only variable that correlated exclusively with Attention Deficit in the BRIEF, whereas in the BRIEF-E while in this study was initiate which showed an exclusive correlation with this variable. In both studies, Plan/Organize and Working Memory obtained the highest correlations with the Attention Deficit subscales (.63 and .60, respectively, in the original study versus .55 and .47 in this study).

Lastly, in both studies, the GEC had statistically significant correlations with both subscales, but its correlation was higher with Attention Deficit than with Hyperactivity in the BRIEF (.58 and .47, respectively), and they were practically the same in the BRIEF-E (.60 and .63, respectively).

Table 2
Internal consistency coefficients (Cronbach's alpha) of the subscales and indexes of the BRIEF and BRIEF-E scales

	Alpha coefficient		
Subscales and indexes	**BRIEF**	**BRIEF-E**	**Elements**
Inhibit	.94	.85	10
Shift	.88	.72	8
Emotional control	.92	.89	10
Initiate	.82	.57	8
Working memory	.92	.81	10
Plan/Organize	.91	.68	12
Organization of materials	.88	.85	6
Monitor	.85	.60	8
BRI	.96	.91	28
MI	.96	.89	44
GEC	.98	.92	72

Note: BRI = Behavioral Regulation Index; MI = Metacognition Index; GEC = Global Executive Composite. Clinical Sample: BRIEF (N = 852), BRIEF-E (N = 125).

Table 3
Correlations of the BRIEF and BRIEF-E scales with the ADHD rating scale-IV and the EDAH scale

	BRIEF – ADHD-IV		BRIEF-E - EDAH	
Subscales and indexes	**Attention deficit**	**Hyperactivity**	**Attention deficit**	**Hyperactivity**
Inhibit	.42**	.73**	.05	.71**
Shift	.39**	.59**	.07	.34**
Emotional control	.39**	.56**	.12	.59**
Initiate	.55**	.36**	.47**	.12
Working memory	.60**	.44**	.47**	.21*
Plan/Organize	.63**	.33**	.55**	.22*
Organization of materials	.49**	.15	.26**	.29**
Monitor	.54**	.45**	.39**	.39**
BRI	.44**	.70**	.10	.68**
MI	.67**	.38**	.62**	.32**
GEC	.63**	.60**	.47**	.58**

Note: BRI = Behavioral Regulation Index; MI = Metacognition Index; GEC = Global Executive Composite. Clinical Sample. BRIEF (N = 100), BRIEF-E (N = 125)
* p<.05; ** p<.01.

Discussion and conclusions

The goal of this study was to obtain preliminary data in a convenience clinical sample of 125 Spanish children and adolescents about the potential utility of the BRIEF scale as an instrument to assess executive functions. For this purpose, the internal structure and consistency of its scores were analyzed, as well as its convergent validity with behavioral variables by analyzing its correlations with the EDAH scale (Farré & Narbona, 1997). The results were compared with those obtained in the original validation study with a clinical sample (Gioia et al., 2000).

In general terms, the results obtained suggest good, albeit improvable, psychometric properties of the scale, coinciding to a great extent with those obtained in the original study. For instance, we obtained the same factor structure, made up of two components (Metacognition and Behavioral Regulation), each one in turn, comprising the same subscales.

Regarding reliability of the scores, the internal consistency coefficients were generally higher in the original study. Nevertheless, in both studies, a similar pattern was obtained, with Initiate being the subscale with the lowest coefficients, and the GEC presenting the highest. The BRI and MI scores presented high internal consistency, very similar to that of the original study.

Lastly, evidence of convergent validity with other behavioral variables was obtained, with statistically significant correlations found between the components of the BRIEF-E scale and the Hyperactivity and Attention Deficit subscales of the EDAH scale. These results coincide with those obtained by Gioia et al. (2000) in the original study, although the results of the present study suggest a higher discriminatory capacity of the Behavioral Regulation component, which only correlated with Hyperactivity. Similar results were found by McCandless and O'Laughlin (2007) using the Behavior Assessment System for Children (BASC: Reynolds & Kamphaus, 1992).

However, although these results provide preliminary data supporting the potential utility of the BRIEF scale in different cultural settings from the original one, some aspects of the study, as well as of the instrument itself, should be taken into account.

One on them is the small sample size and the broad age range used in this study, as well as the heterogeneity of the clinical groups that make up the sample, and the high rate of ADHD in the groups. All this makes it necessary to perform new studies, with larger and more representative samples. This would allow analysis by age group, thereby obtaining a developmental profile of deficits in executive functioning. Also, in view of the diversity of the cognitive and behavioral manifestations of these disorders, it would be interesting to establish differences among them, thereby providing new data about the utility of the scale. In this sense, two prior studies were carried out with Spanish samples (García et al., 2013; García et al., in press), providing evidence of the existence of differential profiles of the ADHD subtypes, as well as of their comorbidity with RD.

The large number of items that make up the scale should also be taken into account. Lejeune et al. (2010) recently conducted a study with a brief version of the scale, made up of 24 items. Although in view of its recent appearance, there are still no data about its diagnostic capacity, a shorter questionnaire would be particularly useful in cases where families are requested to complete a broad range of reports or when administering it for epidemiological studies.

Although our results contribute preliminary evidence of the utility of the BRIEF scale in a Spanish clinical sample, following Muñiz et al. (2013), if the final goal is to achieve maximum correspondence and adequacy between both instruments, it is necessary to continue to analyze and to study in more depth the adequacy of the methodological, cultural, linguistic, conceptual, and metric aspects of the scale. This would open up new lines of research in this direction.

Acknowledgements

This study was carried out with funds from the "Severo Ochoa" program with reference nr. BP: 11-067, of the Foundation for Promotion of Scientific and Applied Research and Technology in Asturias (FICYT).

References

Anderson, P.J., & Reidy, N. (2012). Assessing executive function in preschoolers. *Neuropsychological Review, 22*, 345-360.

Baym, C.L., Corbett, B.A., Wright, S.B., & Bunge, S.A. (2008). Neural correlates of tic severity and cognitive control in children with Tourette Syndrome. *Brain, 131*, 165-179.

Bental, B., & Tirosh, E. (2007). The relationship between attention, executive functions and reading domain abilities in attention deficit hyperactivity disorder and reading disorder: A comparative study. *Journal of Children Psychology and Psychiatry, 48*(5), 455-463.

Brock, L.L., Rimm-Kaufman, S.E., Nathanson, L., & Grimm, K.J. (2009). The contribution of 'hot' and 'cool' executive function to children's academic achievement and classroom behavior. *Early Childhood Research Quarterly, 24*, 337-349.

Di Trani, M., Casini, M.P., Capuzzo, F., Gentile, S., Bianco, G., Menghini, D., & Vicari, S. (2011). Executive and intellectual functions in attention-deficit/hyperactivity disorder with and without comorbidity. *Brain and Development, 33*(6), 462-469.

Donders, J., Den Braber, D., & Vos, L. (2010). Construct and criterion validity of the Behaviour Rating Inventory of Executive Function (BRIEF) in children referred for neuropsychological assessment after pediatric traumatic brain injury. *Journal of Neuropsychology, 4*(2), 197-209.

DuPaul, G.J., Power, T.J., Anastopoulus, A.D., & Reid, R. (1996). *ADHD Rating Scale-IV: Checklist, norms and clinical interpretation*. New York: Guilford Press.

Elosua, P. (2003). On test validity. *Psicothema, 15*(2), 315-321.

Farré, A., & Narbona, J. (1997). *Escala de Déficit de Atención e Hiper-actividad (E.D.A.H.)* [Scale of Attention Deficit and Hyperactivity (E.D.A.H.)]. Madrid: TEA Ediciones.

Ferrando, P.J., & Anguiano-Carrasco, C. (2010). El análisis factorial como técnica de investigación en Psicología [Factor analysis as a research technique in Psychology]. *Papeles del Psicólogo, 31*(1), 18-33.

García, T., González-Castro, P., Rodríguez, C., Álvarez, D., Cueli, M., & Álvarez, L. (in press). Application of the BRIEF scale in the delimitation of clinical subgroups of ADHD: An exploratory study. *Aula Abierta*.

García, T., Rodríguez, C., González-Castro, P., Álvarez, D., Cueli, M., & González-Pienda, J.A. (2013). Executive functioning in attention deficit hyperactivity disorder and reading disabilities. *International Journal of Psychology & Psychological Therapy, 13*(2), 179-194.

Garon, N., Bryson, S.E., & Smith, I.M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin, 134*, 31-60.

Gioia, G.A., Isquith, P.K., Guy, S., & Kenworthy, L. (2000). *BRIEF: Behavior Rating Inventory of Executive Function professional manual*. Lutz, FL: Psychological Assessment Resources.

Gioia, G.A., Kenworthy, L., & Isquith, P.K. (2010). Executive function in the real world: BRIEF lesson from Mark Ylvisaker. *Journal of Head Trauma Rehabilitation, 25*(6), 433-439.

Hale, J.B., Reddy, L.A., Semrud-Clikeman, M., Hain, L.A., Whitaker, J., Morley, J., ..., & Jones, N. (2011). Executive impairment determines ADHD medication response: Implications for academic achievement. *Journal of Learning Disabilities, 44*(2), 196-212.

Han, Y.M.Y., Leunga, W.W., Wong, C.K., Lamc, J.M.K., Cheung, M., & Chan, A.S. (2011). Lymphocyte subset alterations related to executive function deficits and repetitive stereotyped behavior in autism. *Research in Autism Spectrum Disorders, 5*, 486-494.

Henry, L.A., & Bettenay, C. (2010). The assessment of executive functioning in children. *Child and Adolescent Mental Health, 15*(2), 110-119.

Huizinga, M., & Smidts, D.P. (2011). Age-related changes in executive function: A normative study with the Dutch version of the Behavior Rating Inventory of Executive Function (BRIEF). *Child Neuropsychology, 17*(1), 51-66.

Jarratt, K.P, Riccio, C.A., & Siekierski, B.M. (2005). Assessment of Attention Deficit Hyperactivity Disorder (ADHD) using the BASC and BRIEF. *Applied Neuropsychology, 12*(2), 83-93.

Kenworthy, L., Black, D.O., Harrison, B., Della Rosa, A., & Wallace, G.L. (2009). Are executive control functions related to autism symptoms in high-functioning children? *Child Neuropsychology, 15*(5), 425-440.

Korzeniowski, C.G. (2011). Developmental evolution of executive functioning and its relationship with academic learning. *Revista de Psicología, 7*(13), 7-26.

Latzman, R.D., Elkovitch, N., Young, J., & Clark, L. (2010). The contribution of executive functioning to academic achievement among male adolescents. *Journal of Clinical and Experimental Neuropsychology, 32*(5), 455-462.

Lee, N.R., Fidler, D.J., Blakeley-Smith, A., Daunhauer, L., Robinson, C., & Hepburn, S.L. (2011). Caregiver report of executive functioning in a population-based sample of young children with Down syndrome. *American Journal of Intellectual Development Disabilities, 116*(4), 290-304.

LeJeune, B., Beebe, D., Noll, J., Kenealy, L., Isquith, P., & Gioia, G. (2010). Psychometric support for an abbreviated version of the Behavior Rating Inventory of Executive Function (BRIEF) Parent Form. *Child Neuropsychology, 16*(2), 182-201.

Lezak,M.,Howieson,D.,Bigler,E.,&Tranel,D.(2012).*Neuropsychological assessment* (5th ed.). New York: Oxford University Press.

Makris, N., Biederman, J., Monuteaux, M.C., & Seidman, L.J. (2009). Towards conceptualizing a neural systems-based anatomy of attention-deficit/hyperactivity disorder. *Developmental Neuroscience, 31*, 36-49.

McCandless, S., & O'Laughlin, L. (2007). The clinical utility of the Behavior Rating Inventory of Executive Function (BRIEF) in the diagnosis of ADHD. *Journal of Attention Disorders, 4*(10), 381-389.

McClelland, M.M., Cameron, C.E., Connor, C.M., Farris, C.L., Jewkes, A.M., & Morrison, F.J. (2007). Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. *Developmental Psychology, 43*, 947-959.

Muñiz, J., Elosua, P., & Hambleton, R.K. (2013). Guidelines for the translation and adaptation of tests: Second edition. *Psicothema, 25*(2), 151-157.

Neri, M.L., Guimaraes, C.A., Oliveira, E.P., Duran, M.H., Medeiros, L.L., Montenegro, M.A., ..., & Guerreiro, M.M. (2012). Neuropsychological assessment of children with rolandic epilepsy: Executive functions. *Epilepsy & Behavior, 24*(4), 403-407.

Pratt, B.M. (2000). The comparative development of executive functions in elementary school children with reading disorder and attention-deficit/hyperactivity disorder. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 60*(9-B), 4933-4944.

Raghubar, K.P., Barnes, M.A., & Hecht, S.A. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and Individual Differences, 20*, 110-122.

Reynolds, C.R., & Kamphaus, R.W. (1992). *Behavior Assessment System for Children*. Circle Pines, MN: American Guidance Service.

Riccio, C.A., Homack, S.P., Jarratt, K., & Wolfe, M. (2006). Differences in academic and executive function domains among children with ADHD predominantly inattentive and combined types. *Archives of Clinical Neuropsychology, 21*, 657-667.

Shaffer, D., Fisher, P., Lucas, C., Dulcan, M., & Schwab, M. (2000). NIMH Diagnostic Interview Schedule for Children Version IV (NIMH DISC-IV): Description, differences from previous versions and reliability of some common diagnoses. *Journal of the American Academy of Child and Adolescent Psychiatry, 39*, 28-38.

Shimoni, M., Engel-Yeger, B., & Tirosh, E. (2012). Executive dysfunctions among boys with Attention Deficit Hyperactivity Disorder (ADHD): Performance-based test and parents' report. *Research in Developmental Disabilities, 33*(3), 858-865.

Smidts, D.P., & Huizinga, M. (2009). *BRIEF Executieve Functies Gedragsvragenlijst: Handleiding*. Amsterdam: Hogrefe.

Stelzer, F., & Cervigni, M.A. (2011). Academic achievement and executive functions in childhood and adolescence. A review of the literature. *Revista de Investigación en Educación, 9*(1), 148-156.

Toplak, M.E., Bucciarelli, S.M., Jain, U., & Tannock, R. (2009). Executive functions: Performance-based measures and the Behavior Rating Inventory of Executive Function (BRIEF) in adolescents with attention deficit/hyperactivity disorder (ADHD). *Child Neuropsychology, 15*(1), 53-72.

Van der Ven, S.H., Kroesbergen, E.H., Boom, J., & Leseman, P.P. (2013). The structure of executive functions in children: A closer examination of inhibition, shifting, and updating. *British Journal of Developmental Psychology, 31*(1), 70-87.

Van De Voorde, S., Roeyers, H., Verté, S., & Wiersema, J.R. (2010). Working memory, response inhibition, and within-subject variability in children with attention-deficit/hyperactivity disorder or reading disorder. *Journal of Clinical and Experimental Neuropsychology, 32*(4), 366-379.

Verdejo-García, A., & Bechara, A. (2010). Neuropsychology of the executive Functions. *Psicothema, 22*(2), 227-235.

Vinţan, M.A., Palade, S., Cristea, A., Benga, I., & Muresanu, D.F. (2012). A neuropsychological assessment using computerized battery tests (CANTAB), in children with benign rolandic epilepsy before AED therapy. *Journal of Medical Life, 5*(1), 114-119.

Wechsler, D. (1974). Manual for the Wechsler Intelligence Scale for Children-Revised. NY: Psychological Corporation [Spanish translation: Escala de Inteligencia Wechsler para niños-IV edición (WISC-IV). Madrid: TEA Ediciones, 2005].

Willcutt, E.G., Betjemann, R.S., McGrath, L.M., Chabildas, N.A., Olson, R.K., DeFries, J.C., & Pennington, B.F. (2010). Etiology and neuropsychology of comorbidity between RD and ADHD: The case for multiple-deficit models. *Cortex, 46*(10), 1345-1361.

Williams, J.R. (2008). Revising the Declaration of Helsinki. *World Medical Journal, 54*(4), 120-122.

Wilson, K.R., Donders, J., & Nguyen, L. (2011). Self and parent ratings of executive functioning after adolescent traumatic brain injury. *Rehabilitation Psychology, 56*(2), 100-106.

METACOGNICIÓN Y FUNCIONAMIENTO EJECUTIVO EN EDUCACIÓN PRIMARIA

TRINIDAD GARCÍA, CELESTINO RODRÍGUEZ, PALOMA GONZÁLEZ-CASTRO, DAVID ÁLVAREZ-GARCÍA Y JULIO ANTONIO GONZÁLEZ-PIENDA

UNIVERSIDAD DE OVIEDO (ESPAÑA)

IN PRESS

RESUMEN. Este trabajo analizó las diferencias en habilidades metacognitivas y de funcionamiento ejecutivo entre dos grupos de estudiantes (10-12 años) con diferentes niveles de conocimiento metacognitivo (Alto = 50, Bajo = 64). El conocimiento metacognitivo fue evaluado mediante en un test de reconocimiento de estrategias, mientras que las habilidades metacognitivas (o aplicación de dicho conocimiento) se evaluaron mediante auto-informe. Los estudiantes la frecuencia de empleo de estas habilidades en las fases de Planificación, Ejecución y Evaluación del aprendizaje. Por último, la información sobre el funcionamiento ejecutivo fue proporcionada por familias y profesorado, los cuales cumplimentaron dos formas paralelas de un inventario de calificación de la conducta. Los resultados indicaron que: a) los estudiantes con alto conocimiento metacognitivo señalaron emplear más frecuentemente las habilidades metacognitivas que el grupo con bajo conocimiento metacognitivo, principalmente en las fases de Planificación y Ejecución del aprendizaje; b) tanto familias como profesorado informaron acerca de unas mejores habilidades de funcionamiento ejecutivo en los estudiantes con alto conocimiento metacognitivo, siendo estadísticamente significativas las diferencias en planificación, memoria funcional, focalización de la atención y atención sostenida. Estos resultados muestran la existencia de una asociación entre diferentes niveles de conocimiento metacognitivo y las habilidades metacognitivas y de funcionamiento ejecutivo, y sugieren la necesidad de incidir en el conjunto de variables estudiadas con el fin de promover entre los estudiantes niveles crecientes de control sobre el proceso de aprendizaje.

Palabras clave: metacognición, funciones ejecutivas, conocimiento, habilidades, aprendizaje.

ABSTRACT. This study analyzes differences in metacognitive skills and executive functioning between two groups of students (10-12 years) with different levels of metacognitive knowledge (High = 50; Low = 64). Metacognitive knowledge was assessed by means of a strategy recognition test. Metacognitive skills (or application of this knowledge) were evaluated though self-report. Students reported the frequency with which they applied these skills during the phases of Planning, Execution and Evaluation of learning. Finally, information about student executive functioning was provided by families and teachers, who completed two parallel forms of a behavior rating scale. The results indicated that: a) the group with high levels of metacognitive knowledge reported to use their metacognitive skills more frequently than their peers in the other group. These differences were statistically significant in the phases of Planning and Execution; b) both family and teachers informed about best levels of executive functioning in the students with high metacognitive knowledge. Statistically significant differences were found in planning, functional memory, focus and sustained attention. These results show the existence of an association between different levels of metacognitive knowledge, and differences in metacognitive skills and executive functions, and suggest the need to make emphasis in this set of variables in order to encourage students to acquire increasing levels of control over their learning process.

Keywords: metacognition, executive functions, knowledge, skills, learning.

1. INTRODUCCIÓN

El desempeño académico depende en gran medida de la capacidad de los estudiantes para llevar a cabo procesos de control ejecutivo, también denominados *Funciones Ejecutivas* (en adelante FE). De este modo, desde los primeros años de escolaridad, a los estudiantes se les requiere adquirir grados de responsabilidad crecientes, así como aprender a organizar, integrar y manejar cantidades cada vez mayores de información. Responder exitosamente a las demandas del contexto escolar implica la intervención de procesos como la habilidad para priorizar los objetivos a alcanzar; organizar el tiempo, la información y los materiales; comportarse de un modo flexible y acorde a la situación; juzgar la conveniencia de un posible curso de acción; o monitorizar el propio progreso en la realización de una tarea. Las FE hacen referencia, por tanto, a un amplio rango de procesos o habilidades responsables de la conducta orientada a un objetivo (Anderson, Jacobs, y Anderson, 2008; Flores-Lázaro, Castillo-Preciado, y Jiménez-Miramonte, 2014; García, González-Castro, Areces, Cueli, y Rodríguez, 2014a; Meltzer, 2013), e implica componentes como la memoria de trabajo, la organización y planificación, la inhibición de respuestas, la flexibilidad cognitiva, la capacidad atencional o el control del propio estado emocional (Diamond, 2013; Korzeniowski, 2011; Van De Voorde, Roeyers, Verté, y Wiersema, 2010).

Las dificultades en las FE se manifiestan en el contexto escolar a menudo en problemas relacionados con la escritura, la lectura y el razonamiento matemático, entre otros (García et al. 2013a; Lee, Lynn, y Fong, 2009; Toll, Van der Ven, Kroesbergen, y Van Luit, 2012; Van der Ven, Kroesbergen, Boom, y Leseman, 2013). Estas dificultades se hacen aún más evidentes conforme los estudiantes progresan hacia niveles superiores de escolarización, en parte debido al aumento del volumen y la complejidad de la información a procesar. Como resultado, muchos estudiantes no rinden lo esperado dada su capacidad real, lo cual puede ser extremadamente frustrante, afectando su motivación así como su desempeño académico y social.

No obstante, otra parte importante en el rendimiento académico es la capacidad de los estudiantes para evaluar su propio proceso de aprendizaje y diferenciar aquellas estrategias que son útiles para ellos, sabiendo reconocer porqué, cómo y cuándo aplicarlas. Esta capacidad para reflexionar acerca del propio pensamiento y aprendizaje ha sido definida como *Metacognición* (Flavel, 1979).

El término Metacognición es un constructo multidimensional. Tradicionalmente se han diferenciado dos componentes: conocimiento metacognitivo y habilidades metacognitivas (Flavell, 1979; Lucangeli y Cabrele, 2006; Pennequin, Sorel, y Mainguy, 2010a). El *conocimiento metacognitivo* hace referencia a varios aspectos: el conocimiento declarativo sobre estrategias de aprendizaje y sobre uno mismo como aprendiz; el conocimiento procedimental sobre como usar esas estrategias; y el conocimiento condicional a cerca de cuándo y porqué emplearlas. Este conocimiento se basa en una interrelación entre la persona, las características de la tarea y las estrategias disponibles en una situación de aprendizaje. Por su parte, las *habilidades metacognitivas* hacen referencia a habilidades de orden superior e implican un componente de regulación de la propia cognición y la conducta. De este modo, las habilidades metacognitivas implicarían componentes como el análisis, la planificación, la monitorización, y la reflexión y evaluación de la ejecución de una tarea (Throndsen, 2011; Pennequin et al., 2010a). Estas habilidades permitirían un aprendizaje profundo y transferible (Panadero y Alonso-Tapia, 2014).

Una evidencia a cerca de esta distinción reside en el hecho de que ambos componentes parecen desarrollarse a diferentes ritmos, siendo el conocimiento metacognitivo anterior a las habilidades metacognitivas y base sobre la cual se sustentan estas últimas (Blöte, Van Otterloo, Stevenson, y Veenman, 2004; Pennequin, Sorel, Nanty, y Fontaine, 2010b; Weil et al., 2013). En este sentido, Pennequin et al. (2010b) señalan que, mientras el conocimiento metacognitivo se empezaría a desarrollar a la edad de seis años, la propia aplicación de este conocimiento (es decir, las habilidades metacognitivas) no parece alcanzar la madurez hasta los once o doce años de edad. De este modo, como señalan Valle et al. (2009), para que los estudiantes sean exitosos en su aprendizaje no es suficiente con que entiendan y conozcan qué estrategias deben ser aplicadas, sino que es necesario que sepan aplicarlas eficazmente en situaciones de aprendizaje.

Ambos tipos de componentes, FE y Metacognición, guardan una estrecha relación. En este sentido, la revisión de la literatura ha confirmado el papel de diversas funciones ejecutivas como la planificación, memoria, organización, o flexibilidad cognitiva, en el control del proceso de aprendizaje, así como del esfuerzo y la persistencia en la tarea (Corso, Sperb, Inchausti de Jou, y Fumagalli, 2013; Meltzer, 2013; Lyons y Zelano, 2011). No obstante, a pesar de la relación conceptual y práctica entre ambos componentes, la mayoría de los estudios centrados en estos aspectos han sido realizados desde orientaciones muy diferentes. En este sentido, mientras que los estudios sobre Metacognición se habrían llevado a cabo desde el campo de la Psicología y la Educación, la mayoría de los estudios sobre FE parecen haberse desarrollado dentro de contextos más clínicos, fundamentalmente desde la Neuropsicología (Corso et al., 2013; Pennequin et al., 2010a).

Son pocos los estudios, por tanto, que han analizado la relación entre ambos aspectos desde un punto de vista comprensivo (por ejemplo, Garner, 2009; Kuhn, y Pease, 2010; Pennequin et al., 2010a; Roebers, Cimeli, Röthlisberger, y Neuenschwande, 2012; Schneider, 2010). Estos estudios han demostrado la importancia de la relación entre ambos componentes en diferentes edades y etapas educativas. Concretamente, en el estudio de Garner (2009), con 108 estudiantes universitarios a los que se les aplicaron cuestionarios de FE y estrategias de aprendizaje, se encontró que las habilidades de planificación predijeron significativamente el uso de estrategias cognitivas y metacognitivas, así como la regulación del esfuerzo académico. En esta misma línea, el trabajo de Pennequin et al. (2010a) ha evidenciado la relevancia de aspectos como la memoria de trabajo y la flexibilidad cognitiva como soporte tanto del conocimiento metacognitivo como de las habilidades metacognitivas en adultos. Por su parte, en el trabajo de Roebers et al. (2012), con una muestra de 209 estudiantes en los primeros cursos de Educación Primaria, los resultados indicaron que el rendimiento en varias tareas de FE (inhibición, flexibilidad cognitiva y fluidez verbal) se relacionó significativamente con los niveles de control metacognitivo de los estudiantes, siendo un importante predictor en el rendimiento en matemáticas, lectura y escritura. Finalmente Kuhn y Pease (2010) y Schneider (2010) han demostrado la importancia de las habilidades de inhibición de respuestas y memoria de trabajo en el desarrollo metacognitivo en la infancia y adolescencia.

Si bien estos estudios no son numerosos, una de las características comunes entre ellos es el tipo de instrumentos de evaluación de las FE que emplean. En este caso, la mayoría de ellos se basan en la aplicación de pruebas neuropsicológicas o medidas basadas en la ejecución. También denominadas Test Neuropsicológicos, este tipo de pruebas se aplican generalmente en contextos clínicos y consisten en pruebas individuales o baterías que miden una serie de indicadores objetivos relacionados con la ejecución de los sujetos, como tiempos de respuesta, número de errores y omisiones. Un ejemplo de estas pruebas serían el Test Stroop (Stroop, 1935; Martín et al., 2012) para la medida de la inhibición de respuestas, o la Torre de Hanoi (Borys, Spitz, y Dorans, 1892; Díaz et al., 2012) para la medida de la planificación. Estas pruebas han sido ampliamente utilizadas, mostrando en líneas generales su utilidad en la evaluación del funcionamiento ejecutivo. Sin embargo, han sido criticadas por mostrarse poco específicas, demasiado estructuradas y constituir modelos poco representativos del mundo real. De ahí que se les atribuya una baja validez ecológica (Burin, Drake, y Harris, 2007; Chevignard, Catroppa, Galvin, y Anderson, 2010; Gioia, Kenworthy, e Isquith, 2010; Lee, 2011; Lezak, Howieson, Bigler, y Tranel, 2012).

Una alternativa a este tipo de medidas es el uso de cuestionarios de calificación de la conducta. Éstos permiten valorar una gran variedad de componentes desde el punto de vista de la observación de la conducta de niños y adolescentes en el hogar y centro educativo. Se basan en la información proporcionada por familias y profesorado sobre la frecuencia o intensidad de ciertas conductas problemáticas, que serían indicativas de dificultades en las funciones ejecutivas. Este tipo de medidas han sido ampliamente desarrolladas en los últimos años, en parte debido al reconocimiento del hecho de las FE no sólo implican aspectos cognitivos, sino también conductuales y emocionales, muchos de los cuales son solamente evidenciados en contextos habituales (Egeland y Fallmyr, 2010; Mares, McLuckie, Schwartz, y Saini, 2007). Entre los instrumentos de evaluación estandarizados más conocidos se encontrarían el Child Behavior Checklist (CBCL; Achenbach, 1991), el Children Executive Function Inventory (CHEXI; Thorell y Nyberg, 2008), el Behavior Rating Inventory of Executive Functions (BRIEF; Gioia, Isquith, Guy, y Kenworthy, 2000), y más recientemente el Barkley Deficits in Executive Functioning Scale - Children and Adolescents (BDEFS-CA: Barkley, 2012). No obstante, ninguno de estos instrumentos se encuentra disponible en español. En este sentido, se ha elaborado en nuestro país la Escala de Funcionamiento Ejecutivo para Profesorado (EFE-P; García, Álvarez-García, González-Castro, Álvarez y Segurola, 2014a) y la Escala de Funcionamiento Ejecutivo para Familias (EFE-F; García, Álvarez-García, Cueli, González-Castro, y Álvarez, 2013b). Diseñadas como dos escalas paralelas, los primeros datos disponibles evidencian su potencial utilidad en la evaluación de estos componentes en nuestra población.

No obstante, a pesar de las ventajas que el uso de medidas basadas en la observación puede tener como método de evaluación de las FE en situaciones diarias, otros factores deben ser tenidos en cuenta. En este sentido, numerosos estudios han mostrado una baja correspondencia entre el rendimiento de niños y adolescentes en las pruebas basadas en la ejecución y las dificultades observadas en diversos ámbitos de la vida diaria (Bishop, 2011; Lezak et al., 2012; McAuley, Chen, Goos, Schachar, y Crosbie, 2010), así como un acuerdo de bajo a moderado entre diferentes informantes, fundamentalmente familias y profesorado (Rettew et al., 2011; Salbach-Andrae, Lenz, y Lehmkuhl, 2009). Estos aspectos llevan a menudo a resultados contradictorios, de ahí que no exista un único patrón de funcionamiento ejecutivo relacionado con los componentes metacognitivos del aprendizaje. En este sentido, contar con varios tipos de medidas o en su defecto con diferentes informantes se hace esencial con el fin de delimitar con mayor exactitud esta relación.

En resumen, si bien los estudios que han analizado la relación entre diferentes FE y la Metacognición son escasos en el contexto actual, éstos sugieren la existencia de una asociación entre ambos componentes, destacando la memoria de trabajo, la planificación, la flexibilidad cognitiva o la inhibición de respuestas entre los factores más relevantes. No obstante, aspectos como la distinción conceptual y evolutiva entre conocimiento y habilidades metacognitivas, o el tipo de medidas de las FE empleadas (principalmente basadas en la ejecución, con las implicaciones que este aspecto tiene para la validez ecológica de las medidas), imponen en cierta medida restricciones para la generalización de los resultados obtenidos en estudios previos.

De este modo, y partiendo de la concepción de que el conocimiento metacognitivo es anterior a las habilidades metacognitivas y clave para su desarrollo, el objetivo del presente estudio ha sido analizar las diferencias en habilidades metacognitivas y de funcionamiento ejecutivo en dos grupos de estudiantes de tercer ciclo de Educación Primaria (10 a 12 años) que presentan diferentes niveles de conocimiento metacognitivo (Alto vs. Bajo). Se empleó el cuestionario de Conocimiento de Estrategias de Aprendizaje- CEA (Rosário, Mourão, Núñez, González-Pienda, y Solano-Pizarro, 2006) para evaluar el conocimiento metacognitivo, mientras que las habilidades metacognitivas se evaluaron mediante el Inventario de Procesos de Autorregulación del Aprendizaje- IPAA (Rosário et al., 2010). Este cuestionario evalúa el uso de estrategias metacognitivas durante las fases de Planificación, Ejecución y Evaluación del aprendizaje. Por último, con el objetivo de evitar un posible sesgo en la evaluación de las FE se han administrado dos formas paralelas de la escala EFE a familias y profesorado (García et al., 2013b, 2014b), previamente comentadas.

Los objetivos específicos y las hipótesis del presente estudio se presentan a continuación:

1) Determinar si los estudiantes con un buen conocimiento metacognitivo, comparados con aquellos que muestran un bajo conocimiento, muestran unas mejores habilidades metacognitivas (es decir, emplean con más frecuencia estrategias metacognitivas en las fases de Planificación, Ejecución y Evaluación del aprendizaje). Puesto que el

conocimiento metacognitivo funcionaría como base para el desarrollo de las habilidades metacognitivas, se espera que los estudiantes con un buen conocimiento metacognitivo muestren de una forma general puntuaciones significativamente superiores en el cuestionario de habilidades metacognitivas en comparación con el grupo con bajo conocimiento metacognitivo.

2) Analizar si los estudiantes con diferente conocimiento de estrategias metacognitivas se diferencian en los diferentes componentes de las FE, evaluados éstos mediante las informaciones proporcionadas por familias y profesorado. Teniendo en cuenta los estudios previos, se esperan encontrar diferencias estadísticamente significativas entre ambos grupos, principalmente en los compontes de planificación, memoria, organización, inhibición y flexibilidad cognitiva. Los estudiantes con alto conocimiento metacognitivo mostrarán mejores niveles de funcionamiento ejecutivo que sus compañeros con bajo conocimiento metacognitivo.

2. MÉTODO

2.1. Participantes

La muestra estuvo compuesta por 114 estudiantes de Tercer Ciclo de Educación Primaria pertenecientes a 9 centros educativos en Asturias. De la muestra total, 49 estudiantes (43%) fueron mujeres y 65 (57%) varones. El rango de edad de los participantes fue de 10 a 12 años (*M* = 11.25, *DT* = .686). Un total de 31 estudiantes (27.2%) asistía a quinto, mientras que 83 (72.8%) asistía a sexto de Educación Primaria. La muestra total fue dividida en dos grupos en función del nivel de conocimiento de estrategias metacognitivas (alto vs. bajo), evaluadas mediante el cuestionario Conocimiento de Estrategias de Aprendizaje- CEA (Núñez et al., 2011; Rosário et al., 2006). Se eligió el percentil 50 como criterio para la asignación de los estudiantes a los grupos. Los estudiantes se distribuyeron de la siguiente forma:

El grupo 1 (alto conocimiento metacognitivo) estuvo formado por 50 estudiantes, de los cuales 28 (43.8%) fueron mujeres y 36 (56.3%) varones. La edad media de los participantes fue de 11.10 años (*DT* = .647). A este grupo fueron asignados aquellos estudiantes que obtuvieron una puntuación superior al percentil 50 en el cuestionario CEA (Rosário et al., 2006).

El grupo 2 (bajo conocimiento metacognitivo) estuvo formado por 64 estudiantes, de los cuales 21 (42%) fueron mujeres y 29 (58%) varones. La media de edad de los participantes fue 11.36 años (*DT* = .698). Este grupo lo conformaron los estudiantes con una puntuación menor o igual al percentil 50 en el cuestionario anteriormente comentado.

Los estudiantes participaron voluntariamente en el estudio, previo consentimiento informado de las familias. La elección de la muestra fue realizada por accesibilidad (Casal y Mateu, 2003). Los estudiantes con algún diagnóstico previo fueron excluidos de los análisis. No hubo diferencias significativas entre los dos grupos en edad ($F(1,112) = 3.128$, $p \geq 0.05$), ni en la proporción de mujeres [$\chi^2(1) = 1.000$, $p \geq 0.05$] y varones [$\chi^2(1) = .754$, $p \geq 0.05$].

2.2. Medidas

Conocimiento metacognitivo: esta variable fue evaluada a través de la aplicación del cuestionario de Conocimiento de Estrategias de Aprendizaje- CEA (Núñez et al., 2011; Rosário et al., 2006). Esta prueba se basa en el reconocimiento de 10 estrategias metacognitivas generales, relacionadas tanto a aspectos cognitivos, como afectivos y motivacionales. De este modo, presenta 10 cuestiones con 3 alternativas de respuesta, de las cuales solo una de ellas es verdadera. El estudiante tiene que indicar cuál de las opciones es la correcta. La puntuación máxima en esta escala es 10. Puntuaciones elevadas son indicativas de un buen conocimiento metacognitivo. La fiabilidad para la escala, estimada a partir del estadístico Alfa de Cronbach fue de .89.

Habilidades metacognitivas: se empleó el *Inventario de Procesos de Autorregulación del Aprendizaje- IPAA* (Rosário et al., 2010). Este cuestionario está compuesto por 12 ítems y evalúa el uso de varias estrategias metacognitivas en las diferentes fases del proceso de Autorregulación del Aprendizaje (Planificación, Ejecución y Evaluación; Zimmerman, 2000, 2008). Cada fase o dimensión se mide mediante 4 ítems, a través de una escala tipo Likert con 5 alternativas de respuesta (desde 1 = nunca, hasta 5 = siempre). Se evalúa por tanto la frecuencia con que el estudiante emplea estas estrategias, durante estas fases, en situaciones de aprendizaje. La puntuación máxima en cada componente es 20. A mayor puntuación en cada uno de estos ítems y dimensiones mejores habilidades metacognitivas. La fiabilidad de la escala, estimada a través del estadístico Alfa de Cronbach, fue de .80 para la dimensión de Planificación, de .85 para Ejecución y de .87 para Evaluación.

Funciones Ejecutivas: se empleó la *Escala de Funcionamiento Ejecutivo* en sus formas para profesorado (EFE-P; García et al., 2013b) y familias (EFE-F; García et al., 2014b).

Diseñadas como dos formas paralelas, son aplicables desde los 6 a los 18 años de edad. Estas escalas evalúan la frecuencia con la que niños y adolescentes muestran una serie de conductas, indicativas de posibles déficits en FE, en el contexto educativo o en el hogar. Para ello, se basan en las informaciones proporcionadas por familias y profesorado. Formadas por 27 ítems cada una, evalúan un total de 9 componentes (Control de la Impulsividad, Control de la Hiperactividad, Control Emocional, Capacidad de Concentración, Focalización de la Atención, Planificación, Organización, Memoria Funcional y Flexibilidad Cognitiva). Los ítems siguen una escala tipo Likert con 5 opciones de respuesta, desde 1 = Nunca, hasta 5 = Siempre. La puntuación máxima en cada componente es 15. Puntuaciones elevadas en estos componentes serían indicativas de dificultades en las FE. Ambas escalas han sido puestas a prueba en una muestra de 1019 y 616 estudiantes respectivamente, pertenecientes a 41 centros educativos de Asturias (España). Los análisis factoriales confirmatorios mostraron un buen ajuste de los datos al modelo compuesto por nueve factores, con unos índices de fiabilidad que varían entre .77 y .89 para los diferentes componentes en ambas formas de la escala.

La correspondencia entre las informaciones proporcionadas por familias y profesorado en el presente estudio se muestra en la Tabla 1. Como se puede observar, se encontraron correlaciones estadísticamente significativas entre pares de componentes, con la excepción de la variable Flexibilidad. Estas correlaciones fueron en todos los casos positivas, variando desde correlaciones bajas a moderadas. Las correlaciones más bajas (cercanas a .25) fueron encontradas en los componentes de Control Emocional, Memoria Funcional y Control de la Impulsividad, mientras que las más elevadas (superiores a .37) se encontraron en Planificación, Atención Sostenida, Focalización de la Atención y Organización.

TABLA 1

Correlaciones bivariadas entre las puntuaciones en las escalas EFE-F y EFE-P

	EFE-P								
EFE-F	1	2	3	4	5	6	7	8	9
1.Control de la Impulsividad	.282**	.261**	.207*	.276**	.084	.097	.208*	.233*	.112
2.Control de la Hiperactividad	.302**	.312**	.241**	.332**	.181	.135	.219*	.247**	.137
3.Control Emocional	.166	.155	.218*	.192*	.098	.026	.137	.204*	.101
4.Focalización	.251**	.293**	.247**	.390**	.256**	.190*	.281**	.313**	.093
4.Atención Sostenida	.223*	.306**	.215*	.428**	.397**	.316**	.413**	.402**	.177
6.Memoria	.243**	.294**	.226*	.338**	.277**	.232*	.324**	.387**	.164*
7.Planificación	.206*	.281**	.177	.377**	.338**	.243**	.417**	.370**	.146
8.Organización	.166	.244**	.161	.333**	.271**	.197*	.314**	.371**	.125
9.Flexibilidad	.227*	.192	.268*	.018	.053	.002	.094	.119	.035

Nota. N = 114; EFE-F = Escala de Funcionamiento Ejecutivo para Familias; EFE-P = Escala de Funcionamiento Ejecutivo para Profesorado

** $p < .01$; * $p < .05$

2.3. Procedimiento

Este estudio fue llevado a cabo de acuerdo con la Declaración de Helsinki (Williams, 2008), la cual recoge los principios éticos para la investigación con seres humanos. Una vez presentado el estudio a los equipos directivos, se obtuvo la autorización de los mismos y el consentimiento informado de las familias, tras lo cual se realizó la recogida de la información. Los cuestionarios administrados a alumnado se aplicaron colectivamente en una sesión regular de clase (aproximadamente 50 minutos). La participación del alumnado fue voluntaria, teniendo presente en todo momento el anonimato y la garantía de confidencialidad de los datos. Los cuestionarios de FE fueron entregados a los tutores, que además de cumplimentar sus propios cuestionarios, se encargaron de la entrega y recogida de los cuestionarios administrados a las familias. Éstas respondieron los cuestionarios en sus casas.

Las instrucciones sobre cómo cumplimentar las escalas fueron proporcionadas junto con los propios cuestionarios. Solo se tuvieron en cuenta en este estudio aquellos cuestionarios en los que se contaba con las dos formas cumplimentadas (familias y profesorado).

2.4. Análisis de datos

En este trabajo se empleó un diseño comparativo de tipo transversal (Ato, López, y Benavente, 2013). Dados los objetivos del mismo, se optó por llevar a cabo Análisis Multivariados de la Varianza (MANOVA) con el fin de analizar las diferencias entre grupos con diferente conocimiento metacognitivo en las variables de estudio: habilidades metacognitivas (fases de Planificación, Ejecución y Evaluación), y los nueve componentes de las FE evaluados por familias y profesorado.

Se empleó el programa SPSS 19.0 para el tratamiento estadístico de los datos. Las diferencias se consideraron estadísticamente significativas a un nivel de $p \leq .05$. Teniendo en cuenta la necesidad de conocer la significación práctica, además de la significación estadística de las diferencias, se incluyó un indicador de la magnitud del efecto (Ato et al., 2013). Se empleó para ello el criterio de Cohen (1988), según el cual el efecto es pequeño cuando $\eta p^2 = .01$ ($d = .20$), medio cuando $\eta p^2 = .059$ ($d = .50$), y alto cuando $\eta p^2 = .138$ ($d = .80$). Finalmente, dado que el presente estudio se basa en el análisis de las diferencias entre dos grupos, junto con las medias y el análisis de éstas diferencias, se obtuvieron los valores de asimetría y curtosis por separado para cada grupo y variable analizada (Tablas 2 y 3). El criterio de Finney y Di Stefano (2006) fue empleado para comprobar la adecuación de estos valores. Según este criterio, ±2 y ±7 son los valores máximos aceptables en simetría y curtosis para llevar a cabo análisis paramétricos. Las variables en este estudio cumplieron con esta condición.

3. Resultados

3.1. Diferencias en las habilidades metacognitivas entre los grupos con alto y bajo conocimiento metacognitivo

En la Tabla 2 se pueden observar las medias de los estudiantes con diferente conocimiento metacognitivo (alto vs. bajo) en cuanto al uso de estrategias metacognitivas en las diferentes fases del Aprendizaje Autorregulado. Estas medias indicaron como los estudiantes en el grupo con alto conocimiento metacognitivo señalaron emplear con más frecuencia esas estrategias de forma sistemática en todas las fases. En este sentido, el

MANOVA llevado a cabo confirmó la existencia de diferencias estadísticamente significativas entre ambos grupos en el conjunto de variables mencionadas (*Lambda de Wilks* = .873, $F(3,110) = 5.333$, $p < .001$, $\eta p^2 = .127$). Este análisis reveló asimismo Planificación ($p < .001$) y Ejecución ($p < .01$) como las fases donde se hallaron estas diferencias, fundamentalmente en la primera, con un tamaño del efecto considerablemente mayor. En cuanto a la fase de Evaluación, no se hallaron diferencias estadísticamente significativas entre los grupos ($p = .25$).

Teniendo en cuenta las medias de ambos grupos en estas variables, se observó un uso infrecuente de las habilidades metacognitivas en las fases de Evaluación, y sobre todo en Planificación (con una media cercana a 8 en esta fase, cuando la puntuación máxima para esta sub-escala iría hasta 20), mientras que el uso de estas estrategias fue mucho más frecuente en la fase de Ejecución.

TABLA 2

Medias *(M)*, Desviaciones Típicas *(DT)* y diferencias entre los grupos en habilidades metacognitivas

Habilidades Metacognitivas (IPAA)	Grupo 1 (N = 50)			Grupo 2 (N = 64)			Diferencias	
	M(DT)	*Asimetría*	*Curtosis*	*M(DT)*	*Asimetría*	*Curtosis*	$F(1,112)$	ηp^2
Planificación	8.560 (1.296)	-.460	-.875	7.677 (1.299)	-.986	2.559	12.965 ***	.104
Ejecución	15.780 (2.565)	-1.210	2.982	14.479 (2.739)	-.501	.405	6.688 **	.056
Evaluación	11.960 (1.947)	-.183	-.222	11.544 (1.922)	-.306	-.209	4.851	.011

Nota. Grupo 1: Alto conocimiento metacognitivo; Grupo 2: Bajo conocimiento metacognitivo; IPAA = Inventario de Procesos de Autorregulación del Aprendizaje.

*** $p < .001$; ** $p < .01$

3.2. Diferencias en funcionamiento ejecutivo (evaluado por familias y profesorado) entre los grupos con alto y bajo conocimiento metacognitivo

Las medias de ambos grupos en las variables de funcionamiento ejecutivo evaluadas por familias y profesorado se muestran en las Tabla 3 y 4, respectivamente. De forma general, los estudiantes con alto conocimiento metacognitivo fueron evaluados por sus familias y profesorado de una forma más positiva (informando acerca de un menor déficit ejecutivo), en comparación con el grupo con bajo conocimiento metacognitivo. Esto se puede observar atendiendo a las medias obtenidas en los componentes de las FE, menores en el primero de los grupos en la mayoría de los casos. Solamente en los componentes de Control de la Impulsividad, Control Emocional y Flexibilidad evaluados por el profesorado se encontró un patrón diferente, con las medias fueron ligeramente superiores en el grupo con alto conocimiento de estrategias metacognitivas.

Los MANOVA realizados revelaron la existencia de diferencias estadísticamente significativas en el conjunto de las FE evaluadas entre ambos grupos de estudiantes, tanto cuando los informantes fueron las familias (*Lambda de Wilks* = .826, $F(9,104) = 2.438$, $p = .015$, $\eta p^2 = .174$) como cuando fue el profesorado (*Lambda de Wilks* = .807, $F(9,104) = 2.766$, $p = .006$, $\eta p^2 = .193$). Se encontraron diferencias estadísticamente significativas en ambas formas de la escala en los componentes de Focalización de la atención, Atención Sostenida, Memoria y Planificación. Aunque con un tamaño del efecto menor, también se encontraron diferencias estadísticamente significativas en el componente de Organización en la escala para familias. El componente de Planificación fue el que presentó un tamaño del efecto empleando ambas escalas.

Teniendo nuevamente en cuenta las medias de ambos grupos, se puede observar como las medias en el cuestionario administrado a profesorado fueron inferiores a las registradas en familias en la totalidad de los componentes evaluados, lo que indica que éstos primeros informaron a cerca de un mejor funcionamiento ejecutivo (menor déficit) en el alumnado. Las medias en ambos casos señalaron no obstante la ausencia de dificultades importantes en las conductas evaluadas. En este sentido, si bien la puntuación máxima en cada componente sería 15, las medias variaron de 4.82 y 7.84 a en el cuestionario administrado a familias, y de 4.14 a 6.66 en el administrado a profesorado.

TABLA 3

Medias *(M)*, Desviaciones Típicas *(DT)* y diferencias entre los grupos en los componentes de Funciones Ejecutivas evaluados por familias

Funciones Ejecutivas	Grupo 1 (N = 50)			Grupo 2 (N = 64)			Diferencias	
	M(DT)	*Asimetría*	*Curtosis*	*M(DT)*	*Asimetría*	*Curtosis*	*F*(1,112)	ηp^2
Control de la Impulsividad	6.78 (2.243)	.999	.575	7.27 (2.674)	.647	.196	1.064	.009
Control de la Hiperactividad	5.76 (1.985)	1.081	2.110	6.20 (2.297)	.790	.680	1.175	.010
Control Emocional	6.68 (2.817)	.564	-.674	6.34 (2.869)	1.187	1.313	.392	.003
Focalización	6.48 (2.288)	.751	.504	7.66 (2.692)	.571	-.436	6.101 *	.052
Atención Sostenida	6.38 (2.230)	.660	-.323	7.55 (2.594)	.418	-.390	6.412 *	.054
Memoria	4.82 (1.662)	1.131	1.155	5.94 (2.390)	.595	-.067	7.929 **	.066
Planificación	6.16 (2.566)	.788	.128	7.84 (2.940)	.352	-.440	10.278 **	.084
Organización	5.76 (2.568)	1.121	1.170	6.86 (2.828)	.409	-.915	4.595 *	.039
Flexibilidad	6.08 (2.156)	.937	.634	6.56 (2.349)	.889	1.639	1.272	.011

Nota. Grupo 1: Alto conocimiento metacognitivo; Grupo 2: Bajo conocimiento metacognitivo. Puntuaciones elevadas en EFE-P y EFE-F son indicativas de dificultades en las funciones ejecutivas.

*** $p < .001$; ** $p < .01$; * $p < .05$

TABLA 4

Medias *(M)*, Desviaciones Típicas *(DT)* y diferencias entre los grupos en los componentes de Funciones Ejecutivas evaluados por profesorado

Funciones Ejecutivas	Grupo 1 (N = 50)			Grupo 2 (N = 64)			Diferencias	
	M(DT)	*Asimetría*	*Curtosis*	*M(DT)*	*Asimetría*	*Curtosis*	*F*(1,112)	ηp^2
Control de la Impulsividad	5.36 (2.724)	1.541	2.472	5.25 (2.777)	1.697	2.910	.045	.000
Control de la Hiperactividad	4.92 (2.554)	1.448	1.535	5.14 (2.636)	1.770	3.216	.202	.002
Control Emocional	4.88 (2.471)	1.402	1.123	4.63 (2.640)	.652	2.020	.277	.002
Focalización	5.54 (2.443)	1.096	.774	6.58 (2.224)	.849	.262	5.609 *	.048
Atención Sostenida	4.70 (1.940)	1.456	1.873	6.20 (2.644)	.592	-.624	11.368 ***	.092
Memoria	4.14 (1.841)	1.729	2.248	5.14 (2.403)	1.055	.710	5.942 *	.050
Planificación	5.08 (2.415)	1.307	1.065	6.66 (2.540)	.556	-.460	11.286 ***	.092
Organización	4.80 (2.286)	1.239	.730	5.59 (2.683)	1.143	.716	2.792	.024
Flexibilidad	5.30 (2.742)	1.787	2.852	5.05 (2.134)	1.020	.683	.307	.003

Nota. Grupo 1: Alto conocimiento metacognitivo; Grupo 2: Bajo conocimiento metacognitivo. Puntuaciones elevadas en EFE-F son indicativas de dificultades en las funciones ejecutivas.

*** $p < .001$; ** $p < .01$; * $p < .05$

4. DISCUSIÓN Y CONCLUSIONES

Este trabajo ha tenido como objetivo analizar las diferencias en las habilidades metacognitivas y de funcionamiento ejecutivo entre dos grupos de estudiantes de Tercer Ciclo de Primaria (10-12 años) con diferente conocimiento metacognitivo (alto vs. bajo). Los resultados obtenidos se discuten a continuación en relación con los objetivos e hipótesis planteados:

4.1. Diferencias en las habilidades metacognitivas entre los grupos con alto y bajo conocimiento metacognitivo

En primer lugar, los estudiantes con un alto conocimiento metacognitivo señalaron emplear hacer un uso más de las estrategias metacognitivas evaluadas en las fases de Planificación, Ejecución y Evaluación. En este sentido, un mayor conocimiento metacognitivo se relacionó con unas mejores habilidades metacognitivas, lo que es coherente con estudios previos como el de Blöte et al. (2004), Pennequin et al. (2010a) y Weil et al. (2013), que indican como el conocimiento metacognitivo sería previo a las habilidades metacognitivas y un importante precursor para su desarrollo. Estas diferencias fueron estadísticamente significativas en las fases de Planificación y Ejecución, pero no en la de Evaluación del aprendizaje.

Relacionado con este último resultado, un aspecto a reseñar es el hecho de que, si se consideran las puntuaciones de ambos grupos en la escala de habilidades metacognitivas, se observa como los estudiantes de forma general señalaron emplear con mayor frecuencia estas habilidades metacognitivas en la fase de Ejecución, en contraste con las fases de Evaluación del aprendizaje y sobretodo Planificación. Este resultado podría estar relacionado con el hecho de que los estudiantes en estas etapas educativas (últimos años de Educación Primaria) tienden a dar una mayor prominencia a la ejecución de la tarea, frente a la elaboración de un plan previo y la evaluación del progreso y de los resultados obtenidos (Cleary y Chen, 2009; Kramarski y Gutman, 2006; Montague, Enders, y Dietz, 2011). Por tanto, es en cierto modo esperable que los estudiantes muestren un mejor control metacognitivo sobre los procesos de ejecución, en parte porque les resultan más familiares.

Quizás este perfil observado en la muestra total explique por qué pese a haberse observado ciertas diferencias entre los grupos con diferente conocimiento metacognitivo en el uso de estrategias metacognitivas en la fase de Evaluación (si bien dichas diferencias no fueron tan marcadas como en las otras fases), éstas no alcanzaron la significación estadística. En este sentido, los estudiantes en general podrían haber mostrado ciertas dificultades para reconocer el tipo de estrategias propias de esta fase, lo que haría más difícil para ellos evaluar hasta qué punto las aplican o no. No obstante, puesto que el instrumento de evaluación del conocimiento metacognitivo empleado en este estudio para la asignación de los estudiantes a los grupos no permite establecer diferencias entre diferentes fases, esta hipótesis no ha podido ser comprobada. Sería necesario por tanto incidir en este aspecto en futuros estudios.

Estos resultados en su conjunto indicarían la existencia de una relación entre el conocimiento y las habilidades metacognitivas, así como la necesidad de fomentar el desarrollo de ambos componentes desde edades tempranas. Esto podría hacerse del siguiente modo: atendiendo a la promoción de un buen conocimiento metacognitivo como base para el desarrollo posterior de las habilidades metacognitivas en primer lugar, e incidiendo directamente sobre la aplicación de este conocimiento en situaciones de aprendizaje reales. La intervención en estrategias de Aprendizaje Autorregulado ha mostrado ser beneficiosa en ambos sentidos, fundamentalmente en aquellos estudiantes que muestran dificultades del aprendizaje o bajo rendimiento académico (González-Pienda, Fernández, Bernardo, Núñez, y Rosário, 2014; Moos y Ringdal, 2012; Stoeger y Ziegler, 2008).

4.2. *Diferencias en funcionamiento ejecutivo (evaluado por familias y profesorado) entre los grupos con alto y bajo conocimiento metacognitivo*

En cuanto a la segunda cuestión planteada en este estudio, los resultados indicaron que si bien los estudiantes no presentaron dificultades ejecutivas importantes tanto cuando fueron evaluados por familias como por profesorado, ambos informantes atribuyeron un mejor funcionamiento ejecutivo de forma general al grupo con alto conocimiento metacognitivo. Las diferencias entre los grupos se encontraron principalmente en Focalización de la atención, Atención sostenida, Memoria funcional y Planificación en ambas versiones de la escala, así como en las capacidades de Organización en el caso de la forma administrada a familias. En este sentido, un alto conocimiento metacognitivo estuvo relacionado con mejores habilidades de control ejecutivo (o menor déficit) en los componentes evaluados.

Un aspecto a destacar no obstante, es el hecho de que no se observaron diferencias estadísticamente significativas en inhibición de respuestas ni flexibilidad cognitiva, las cuales habían sido encontradas en estudios previos (Corso et al., 2013; Garner, 2009; Kuhn y Pease, 2010; Meltzer, 2013; Pennequin et al., 2010a; Schneider, 2010). Una posible explicación a este hecho estaría relacionada con el tipo de medidas de evaluación de las FE mayoritariamente empleadas en estudios previos, basadas en la ejecución de diferentes tareas ejecutivas. Estas medidas han sido descritas por algunos autores por ser poco específicas en el sentido de que diferentes funciones pueden intervenir en la correcta ejecución de una misma tarea (Burin et al., 2007; Chevignard, et al., 2010; Lee, 2011; Lezak et al., 2012). En el presente estudio, el uso de cuestionarios de calificación de la conducta parece abogar a favor de la distinción propuesta por varios autores entre las denominadas *funciones cálidas* (hot) y *frías* (cool) (Brock, Rimm-Kaufman, y Nathanson, 2009; Zelazo y Carlson, 2012). Las funciones cálidas estarían implicadas en el tratamiento de la información emocional, incluyendo componentes como el control de impulsos, la interpretación de señales corporales, la toma de decisiones y el reconocimiento de la perspectiva del otro, y estarían más relacionadas con el desempeño social del individuo en su entorno. Por otra parte, las funciones frías permitirían un tratamiento más racional de la información, y estarían relacionadas con el razonamiento y el procesamiento de información abstracta, como por ejemplo la memoria de trabajo, la planificación, conceptualización y categorización entre otras. Estas funciones estarían más relacionadas, por tanto, con aspectos del aprendizaje, y serían las que habrían generado diferencias estadísticamente significativas entre los grupos en el presente estudio, siendo las que relacionaron con en el conocimiento de estrategias metacognitivas.

Estos resultados muestran nuevamente la existencia de una relación entre habilidades específicas de funcionamiento ejecutivo y la Metacognición, al menos con respecto al componente de conocimiento metacognitivo. En este sentido, ambos componentes se relacionarían en tanto que implican componentes de planificación, organización, memoria y atencionales. Si bien para autores como Roebers et al. (2012) esta relación podría explicarse en parte desde un punto de vista evolutivo - estos autores sostienen que el progreso en las FE durante la infancia y la adolescencia se ve acompañado del desarrollo de un mayor conocimiento y control de los propios procesos de aprendizaje-, los resultados obtenidos en el presente estudio no permiten conocer la dirección de esta relación. En cualquier caso, se entiende que el hecho de promover avances en un componente (Metacognición o FE), resultaría beneficioso para el desarrollo del otro. No obstante, si bien existen numerosos trabajos centrados en la promoción del desarrollo de los componentes metacognitivos en la edad escolar, lo cierto es que la intervención en FE se ha hecho principalmente en contextos clínicos. Teniendo en cuenta las implicaciones que funciones como las analizadas en el presente estudio tienen en el aprendizaje escolar y en el control de la propia conducta (García, González-Castro, Areces, Cueli, y Rodríguez, 2014b; Marcovitch y Zelazo, 2009, Van De Voorde et al., 2010), se hace cada vez más necesario la generalización de la intervención en estos aspectos a nuevos contextos, como el educativo o el familiar.

Por último, si bien la correspondencia entre la información proporcionada por familias y profesorado en las escalas de FE fue moderada (Tabla 1), en la línea de estudios previos en este campo (Papageorgiou, Kalyva, Dafoulis, y Vostanis, 2008; Rettew et al., 2011; Salbach-Andrae et al., 2009), los datos del presente estudio han revelado como cuando los estudiantes son clasificados basándose en un criterio externo (en este caso haber sido asignados a un grupo según el número de respuestas correctas dadas en el cuestionario de conocimiento metacognitivo-CEA), las informaciones proporcionadas por familias y profesorado tienden a converger. Estos resultados indicarían, por tanto, la conveniencia de contar con varios informantes cuando la evaluación se realiza de acuerdo a la observación de la conducta.

Limitaciones

Para finalizar, hay una serie de limitaciones en el presente estudio que deben ser tenidas en cuenta: En primer lugar, el tamaño muestral impone ciertas limitaciones para la generalización de los resultados obtenidos. Así mismo, junto con una muestra más amplia, sería interesante establecer niveles más extremos para la asignación de los participantes a los diferentes grupos. Si bien en este estudio se empleó el percentil 50 como punto de corte, niveles como el percentil 25 o 75 quizás permitirían obtener un patrón más claro de diferencias significativas; en segundo lugar, el tipo de evaluación de las habilidades metacognitivas empleado debe ser tenido en cuenta. Si bien estos componentes suelen ser valorados mediante métodos de auto-informe, como en el presente estudio, son numerosos los autores que señalan que el empleo de otro tipo de medidas, basadas en el análisis del proceso, serían más adecuadas cuando se trata de la evaluación de este tipo de componentes procedimentales y estratégicos (Azevedo y Aleven, 2013; Lazakidou y Retalis 2010; Veenman, 2011). Estos métodos, como los protocolos Think-aloud (Montague et al., 2011) o la Triple Tarea y sus variantes (Piolat, Kellogg y Farioli, 2001; Piolat, Olive, y Kellogg, 2005) están basados en la recogida de información concurrente a la realización de una tarea cognitiva y suponen un complemento al uso de cuestionarios; en tercer lugar, y una vez salvadas las anteriores limitaciones, otro de los aspectos a considerar sería la conveniencia de ampliar este estudio mediante el análisis de la relación entre FE y el componente de habilidades metacognitivas. Esto permitiría conocer con mayor exactitud la relación entre FE y Metacognición; en cuarto lugar, si bien este estudio ha mostrado la relevancia de ciertas habilidades de control ejecutivo como la atención, planificación, la memoria o la capacidad de organización en relación con el conocimiento metacognitivo, los resultados obtenidos no permiten determinar si las FE explican el desarrollo de tal conocimiento o viceversa, o si otras variables como puede ser el tipo de método instruccional afectan a esta relación; finalmente, y puesto que estudios previos han señalado como los componentes de Metacognición y FE se relacionan con el rendimiento en diversas áreas como las matemáticas o el lenguaje, incluir algún indicador de este tipo en estudios futuros sería de gran interés. Esto permitiría delimitar con mayor exactitud la relación entre ambos componentes, así como si influencia en el rendimiento en diferentes áreas curriculares.

Referencias

Achenbach, T. M. (1991). *Integrative guide for the 1991 CBCL/4-18, YSR and TRF profiles.* Burlington: Department of Psychiatry, University of Vermont.

Anderson, V. M., Jacobs, R. y Anderson, P. (2008). *Executive functions and the frontal lobes: A lifespan perspective.* New York, NY: Taylor & Francis.

Ato, M., López, J. J. y Benavente, A. (2013). Un sistema de clasificación de los diseños de investigación en psicología. *Anales de Psicología, 29*(3), 1038-1059. doi: 10.6018/analesps.29.3.178511

Azevedo, R. y Aleven, V. (2013). *International Handbook of Metacognition and Learning Technologies*. Springer International Handbooks of Education, 28. New York, NY: Springer.

Barkley, R. A. (2012). *Barkley Deficits in Executive Functioning Scale - Children and Adolescents (BDEFS-CA)*. New York, NY: The Guilford Press.

Bishop, T. L. (2011). Relationship between performance-based measures of executive function and the Behavior Rating Inventory of Executive Function (BRIEF), a parent rating measure. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 72*(1-B), 522.

Blöte, A. W., Van Otterloo, S. G., van Stevenson, C. E. y Veenman, M. V. J. (2004). Discovery and maintenance of the many-to-one counting strategy in 4-year-olds: A microgenetic study. *British Journal of Developmental Psychology, 22,* 83–102.

Borys, S. V., Spitz, H. H. y Dorans, B. A. (1982). Tower of Hanoi performance of retarded young adults and nonretarded children as a function of solution length and goal state. *Journal of Experimental Child Psychology, 33,* 87-110.

Brock, L. L., Rimm-Kaufman, S. E. y Nathanson, L. (2009). The contributions of 'hot' and 'cool' executive function to children's academic achievement and learning-related behaviors, and engagement in kindergarten. *Early Childhood Research Quarterly, 24*(3), 337-349.

Burin, D. I., Drake, M. y Harris, P. (2007). Evaluación de la atención. En D.I. Burin, M. Drake, y P. Harris (Eds.), *Evaluación Neuropsicológica en adultos* (pp. 146-147) Buenos Aires: Paidós.

Casal, J. y Mateu, E. (2003). Tipos de muestreo. *Revista de Epidemiología y Medicina preventiva, 1*, 3-7.

Chevignard, M. P., Catroppa, C., Galvin, J. y Anderson, V. (2010). Development and evaluation of an ecological task to assess executive functioning post childhood TBI: The Children's Cooking Task. *Brain Impairment, 11*(2), 125–143. doi: 10.1375/brim.11.2.125.

Cleary, T. J. y Chen, P. P. (2009). Self-regulation, motivation, and math achievement in middle school: Variations across grade level and math context. *Journal of School Psychology, 47*(5), 291-314.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.).* Hillsdale, NJ: Lawrence Erlbaum.

Corso, H. V, Sperb, T. M., Inchausti de Jou, G. y Fumagalli, J. (2013). Metacognition and executive functions: relationships between concepts and implications for learning. *Psicologia: Teoria e Pesquisa, 29*(1), 21-29. doi: 10.1590/S0102-37722013000100004

Diamond, A. (2013). Executive Functions. *Annual Review of Psychology, 64*, 135-168. doi: 10.1146/annurev-psych-113011-143750

Díaz, A., Martín, R., Jiménez, J. E., García, E., Hernández, S. y Rodríguez, C. (2012). Torre de Hanoi: datos normativos y desarrollo evolutivo de la planificación. *European Journal of Education and Psychology, 5*(1), 79-91.

Egeland, J. y Fallmyr, Ø. (2010). Confirmatory Factor Analysis of the Behavior Rating Inventory of Executive Function (BRIEF): support for a distinction between emotional and behavioral regulation. *Child Neuropsychology, 16*(4), 326-37. doi: 10.1080/09297041003601462

Finney, S. J. y DiStefano, C. (2006). Non-normal and categorical data in structural equation modeling. En G.R. Hancock y R.O. Muller (Eds.), *Structural equation modeling: A second course* (pp. 269-314). Greenwich, CT: Information Age.

Flavell, J. H. (1979). Metacognition and cognitive monitoring. *American Psychologist, 34,* 906-911.

Flores-Lázaro, J. C., Castillo-Preciado, R. E. y Jiménez-Miramonte, N. A. (2014). Desarrollo de funciones ejecutivas, de la niñez a la juventud. *Anales de Psicología, 30*(2), 463-473. doi: 10.6018/analesps.30.2.155471

García, T., Álvarez-García, D., Cueli, M., González-Castro, P. y Álvarez, L. (2013b). Propiedades psicométricas de la Escala de Funcionamiento Ejecutivo para Profesorado. *Enviado para publicación.*

García, T., Álvarez-García, D., González-Castro, P., Álvarez, L. y Segurola, L. A. (2014b). Propiedades psicométricas de la Escala de Funcionamiento Ejecutivo para Profesorado. *Revista de Psicología y Educación, 9*(1), 49-69.

García, T., González-Castro, P., Areces, D., Cueli, M. y Rodríguez, C. (2014a). Executive functions in children and adolescents: the types of assessment measures used and implications for their validity in clinical and educational contexts. *Papeles del Psicólogo, 35*(3), 215-223.

García, T., Rodríguez, C., González-Castro, P., Álvarez, D., Cueli, M. y González-Pienda, J. A. (2013a). Executive Functioning in children and adolescents with Attention Deficit Hyperactivity Disorder and Reading Disabilities. *International Journal of Psychology y Psychological Therapy, 13*(2), 179-194.

Garner, J. K. (2009). Conceptualizing the Relations between Executive Functions and Self-Regulated Learning. *Journal of Psychology, 143*(4), 405-426. doi: 10.3200/JRLP.143.4.405-426

Gioia, G. A., Isquith, P. K., Guy, S. y Kenworthy, L. (2000). *BRIEF: Behavior Rating Inventory of Executive Function professional manual.* Lutz, FL: Psychological Assessment Resources.

Gioia, G. A., Kenworthy, L. y Isquith, P. K. (2010). Executive function in the real world: BRIEF lessons from Mark Ylvisaker. *Journal of Head Trauma Rehabilitation, 25*(6), 433–439. doi: 10.1097/HTR.0b013e3181fbc272

González-Pienda, J. A., Fernández, E., Bernardo, A. B., Núñez, J. C. y Rosário, P. (2014). Assessment of a self-regulated learning intervention. *The Spanish Journal of Psychology, 17*, 1-9. doi: 10.1017/sjp.2014.12

Korzeniowski, C. G. (2011). Desarrollo evolutivo del funcionamiento ejecutivo y su relación con el aprendizaje escolar. *Revista de Psicología, 7*(13), 7-26.

Kramarski, B. y Gutman, M. (2006). How can self-regulated learning be supported in mathematical E-learning environments? *Journal of Computer Assisted Learning 22*, 24–33. doi: 10.1111/j.1365-2729.2006.00157.x

Kuhn, D. y Pease, M. (2010). The dual components of developing strategy use: Production and inhibition. In H. S. Waters y W. Schneider (Eds.). *Metacognition strategy use and instruction* (pp. 135–159). New York, NY: Guilford Press.

Lazakidou, G. y Retalis, S. (2010). Using computer supported collaborative learning strategies for helping students acquire self-regulated problem-solving skills in mathematics. *Computers y Education 54*, 3-13. doi: 10.1016/j.compedu.2009.02.020

Lee, K., Lynn, N. y Fong, S. (2009). The Contributions of Working Memory and Executive Functioning to Problem Representation and Solution Generation in Algebraic Word Problems. *Journal of Educational Psychology, 101*(2), 373–387.

Lee, E. (2011). An exploratory analysis of the ecological validity of a performance-based assessment of attention. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 71*(9-B), 5777.

Lezak, M. D., Howieson D. B., Bigler, E. D. y Tranel, D. (2012). *Neuropsychological assessment 5th edition.* Oxford: Oxford University Press Chapters.

Lucangeli, D. y Cabrele, S. (2006). The relationship of metacognitive knowledge, skills and beliefs in children with and without mathematical learning disabilities. En A. Desoete y M. V. Veenman (Eds.), *Metacognition in mathematics education* (pp. 103–133). New York, NY: Nova Science Publishers.

Lyons, K. E. y Zelazo, P. D. (2011). Monitoring, metacognition, and executive function: elucidating the role of self-reflection in the development of self-regulation. En J. Benson (Ed.), *Advances in Child Development and Behavior* (pp.379-412). Burlington: Academic Press.

Marcovitch, S. y Zelazo, P. D. (2009). A hierarchical competing systems model of the emergence and early development of executive function. *Development Science, 12*(1), 1-18.

Mares, D., McLuckie, A., Schwartz, M. y Saini, M. (2007). Executive function impairments in children with attention-deficit hyperactivity disorder: do they differ between school and home environments? *Canadian Journal of Psychiatry, 52*(8), 527-534.

Martín, R., Hernández, S., Rodríguez, C., García, E., Díaz, A. y Jiménez, J. E. (2012). Datos normativos para el Test de Stroop: patrón de desarrollo de la inhibición y formas alternativas para su evaluación. *European Journal of Education and Psychology, 5*(1), 39-51.

McAuley, T., Chen, S., Goos, L., Schachar, R. y Crosbie, J. (2010). Is the Behavior Rating Inventory of Executive Functions more strongly associated with measures of impairment or executive function? *Journal of the International Neuropsychological Society, 16*(3), 495-505. doi: 10.1017/S1355617710000093.

Meltzer, L. (2013). Executive Function and Metacognition in Students with Learning Disabilities: New Approaches to Assessment and Intervention. *International Journal for Research in Learning Disabilities, 1*(2), 31-63.

Montague, M., Enders, G. y 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. doi: 10.1017/i0.ii77/073i9487M421762

Moos, D. C. y Ringdal, A. (2012). Self-Regulated Learning in the classroom: A literature review on the teacher’s role. *Education Research International, 2012,* Article ID 423284. doi: 10.1155/2012/423284

Núñez, J. C., Cerezo, R., Bernardo, A., Rosário, P., Valle, A., Fernández, E. y Suárez, N. (2011). Implementation of training programs in self-regulated learning strategies in moodle format: Results of an experience in higher education. *Psicothema, 23*(2), 274-281.

Panadero, E. y Alonso-Tapia, J. (2014). ¿Cómo autorregulan nuestros alumnos? Revisión del modelo cíclico de Zimmerman sobre autorregulación del aprendizaje. *Anales de Psicología, 30*(2), 450-462. doi: 10.6018/analesps.30.2.167221

Papageorgiou, V., Kalyva, E., Dafoulis, V. y Vostanis, P. (2008). Differences in Parents' and Teachers' Ratings of ADHD Symptoms and Other Mental Health Problems. *The European Journal of Psychiatry, 22*(4), 200-210. doi: 10.4321/S0213-61632008000400003

Pennequin, V., Sorel, O. y Mainguy, M. (2010a). Metacognition, Executive Functions and Aging: The Effect of Training in the Use of Metacognitive Skills to Solve Mathematical Word Problems. *Journal of Adult Development*, *17*(3), 168-176. doi: 10.1007/s10804-010-9098-3

Pennequin, V., Sorel, O., Nanty, I. y Fontaine, R. (2010b). Metacognition and low achievement in mathematics: The effect of training in the use of metacognitive skills to solve mathematical word problems. *Thinking & Reasoning, 16*(3), 198-220.

Piolat, A., Kellogg, R. T. y Farioli, F. (2001). The triple task technique for studying writing processes: on which task is attention focused? Current Psychology Letters. *Brain, Behavior and Cognition, 4,* 67–83.

Piolat, A., Olive, T. y Kellogg, R. T. (2005). Cognitive effort during note taking. *Applied Cognitive Psychology, 19*, 291-312. doi: 10.1002/acp.1086

Rettew, D. C., van Oort, F. V. A., Verhulst, F. C., Buitelaar, J. C., Ormel, J., Hartman, C. A., …, y Hudziak, J. H. (2011). When Parent and Teacher Ratings don't agree: The Tracking Adolescents' Individual Lives Survey (TRAILS). *Journal of Child and Adolescent Psychopharmacology, 21*(5), 389–397. doi: 10.1089/cap.2010.0153

Roebers, C. M., Cimeli, P., Röthlisberger, M. y 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 y Learning, 7,* 151-173. doi: 10.1007/s11409-012-9089-9

Rosário, P., González-Pienda, J. A., Pinto, R., Ferreira, P., Lourenço, A. y Paiva, O. (2010). Efficacy of the program "Testas's (mis)adventures" to promote the deep approach to learning. *Psicothema, 22,* 828-834.

Rosário, P., Mourão, R., Núñez, J. C., González-Pienda, J. A. y Solano, P. (2006). School-family: Is a reciprocal and positive relationship possible? *Papeles del Psicólogo, 27*(3), 174-182.

Salbach-Andrae, H., Lenz, K. y Lehmkuhl, U. (2009). Patterns of agreement among parent, teacher and youth ratings in a referred sample. *European Psychiatry, 24*(5), *345–351*. doi: 10.1016/j.eurpsy.2008.07.008

Schneider, W. (2010). Metacognition and memory development in childhood and adolescence. En H. S.Waters y W. Schneider (Eds.). *Metacognition, strategy use and instruction* (pp. 54–81). New York, NY: Guildford Press.

Stoeger, H., y Ziegler, A. (2008). Evaluation of a classroom based training to improve self-regulation in time management tasks during homework activities with fourth graders. Metacognition and Learning, 3(3), 207-230. doi: 10.1007/s11409-008-9027-z

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18,* 643-622.

Thorell, L. B. y Nyberg, L. (2008). The Childhood Executive Functioning Inventory (CHEXI): A new rating instrument for parents and teachers. *Developmental Neuropsychology, 33,* 536–552. doi: 10.1080/87565640802101516

Throndsen, I. (2011). Self-regulated learning of basic arithmetic skills: A longitudinal study. *British Journal of Educational Psychology, 81,* 558–578.

Toll, S. W. M., Van der Ven, S. H. G., Kroesbergen, E. H. y Van Luit, J. E. H. (2011). Executive functions as predictors of math learning disabilities. *Journal of Learning Disabilities, 44*(6), 521 – 532. doi: 10.1177/0022219410387302

Valle, A., Rodríguez, S., Cabanach, R. G., Núñez, J. C., González-Pienda, J. A. y Rosário, P. (2009). Diferencias en rendimiento académico según los niveles de las estrategias cognitivas y de las estrategias de autorregulación. *SUMMA Psicológica UST, 6*(2), 31-42.

Van der Ven, S. H., Kroesbergen, E. H., Boom, J. y Leseman, P. P. (2013). The structure of executive functions in children: a closer examination of inhibition, shifting, and updating. *British Journal of Developmental Psychology, 31*(1), 70-87. doi: 10.1111/j.2044-835X.2012.02079.x

Van De Voorde, S., Roeyers, H., Verté, S. y Wiersema, J. R. (2010). Working memory, response inhibition, and within-subject variability in children with attention-deficit/hyperactivity disorder or reading disorder. *Journal of Clinical and Experimental Neuropsychology, 32*, 366-79. doi: 10.1080/13803390903066865

Veenman, M. V. J. (2011). Learning to self-monitor and self-regulate. En R. Mayer y P. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 197-218). New York, NY: Routledge.

Weil, L. G., Flemingc, S. M., Dumontheil, I., Kilforda, E. J., Weil, R. S., Rees, G., … y Blakemorea. S. (2013). The development of metacognitive ability in adolescence. *Consciousness and Cognition, 22*, 264–271. doi: 10.1016/j.concog.2013.01.004

Williams, J. R. (2008). Revising the Declaration of Helsinki. *World Medical Journal, 54*, 120-125.

Zelazo, P. D. y Carlson, S. M. (2012). Hot and Cool Executive Function in Childhood and Adolescence: Development and Plasticity. *Child Development Perspectives, 6(*4), 354-360. doi: 10.1111/j.1750-8606.2012.00246.x

Zimmerman, B. J. (2000). Attaining self-regulation. A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, y M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13-39). San Diego, CA: Academic Press.

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical, background, methodological developments and future prospects. *American Educational Research Journal*, *45*, 166-183.

Informe del Factor de Impacto de las publicaciones

A continuación se presenta la información refieren al Factor de Impacto de las revistas donde se han aceptado las anteriores publicaciones. Estas revistas se encuentran entre aquellas incluidas en JCR Social Sciences Edition. Para obtener el Factor de Impacto de las mismas se tuvo en cuenta la información proporcionada por la Web of Sciences. Se empleó el año 2013 como referencia, al tratarse éste de la última anualidad recogida hasta el momento.

1) *Revista Latinoamericana de Investigación en Matemática Educativa (RELIME):* esta revista presenta un Factor de Impacto de 0.120, lo que la sitúa en la posición 210 de 219 en la categoría de Educación e Investigación Educativa. Correspondiente con un Cuartil 4.

ISI Web of Knowledge℠

Journal Citation Reports®

WELCOME HELP RETURN TO JOURNAL

Rank in Category: Revista Latinoamericana de Investigacion en Matema...

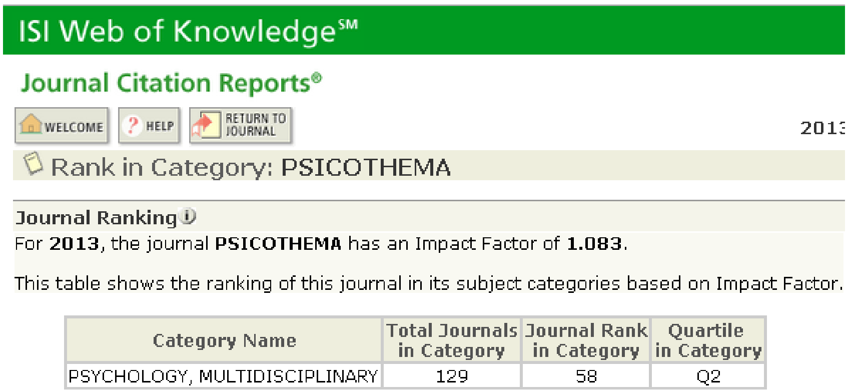
Journal Ranking

For **2013**, the journal **Revista Latinoamericana de Investigacion en Matema...** has an Impact Factor of **0.120**.

This table shows the ranking of this journal in its subject categories based on Impact Factor.

Category Name	Total Journals in Category	Journal Rank in Category	Quartile in Category
EDUCATION & EDUCATIONAL RESEARCH	219	210	Q4

2) *Psicothema:* esta revista tiene in Factor de Impacto de 1.083 y figura en la posición 58 de las 129 revistas incluidas en la categoría de Psicología, Multidisciplinar. Este Factor de Impacto se corresponde con un Cuartil 2.



ISI Web of Knowledge℠

Journal Citation Reports®

WELCOME HELP RETURN TO JOURNAL 2013

Rank in Category: PSICOTHEMA

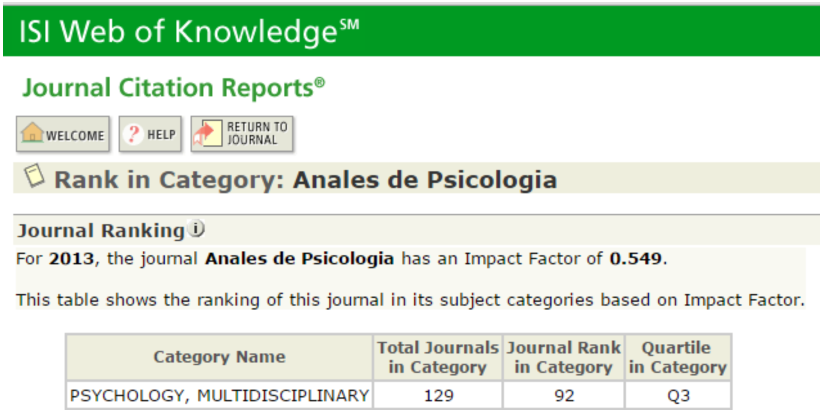
Journal Ranking

For **2013**, the journal **PSICOTHEMA** has an Impact Factor of **1.083**.

This table shows the ranking of this journal in its subject categories based on Impact Factor.

Category Name	Total Journals in Category	Journal Rank in Category	Quartile in Category
PSYCHOLOGY, MULTIDISCIPLINARY	129	58	Q2

3) *Anales de Psicología:* el Factor de Impacto de esta revista es de 0.549. Esta revista ocupa la posición 92 de las 129 revistas incluidas dentro de la categoría Psicología, Multidisciplinar. Se corresponde con un Cuartil 3.



ISI Web of Knowledge℠

Journal Citation Reports®

WELCOME | HELP | RETURN TO JOURNAL

Rank in Category: Anales de Psicologia

Journal Ranking

For **2013**, the journal **Anales de Psicologia** has an Impact Factor of **0.549**.

This table shows the ranking of this journal in its subject categories based on Impact Factor.

Category Name	Total Journals in Category	Journal Rank in Category	Quartile in Category
PSYCHOLOGY, MULTIDISCIPLINARY	129	92	Q3

Trabajos Complementarios

Trabajo Complementario 1:

García, T., Rodríguez, C., González-Castro, P., González-Pienda, J. A. & Torrance, M. (2013). Elementary students´ metacognitive process and post-performance calibration in mathematical problem solving. *Submitted for publication.*

Trabajo Complementario 2:

García, T., Cueli, M., Rodríguez, C., Krawec, J. & González-Castro, P. (2014). Metacognitive knowledge and skills in students with Deep approach to learning. Evidence from mathematical problem solving. *Submitted for publication.*

Trabajo Complementario 3:

García, T., Kroesbergen, E. H., Rodríguez, C., González-Castro, P. & González-Pienda, J. A. (2014). Analysis of the factors involved in making post-performance judgments in mathematical problem solving. *Submitted for publication.*

ELEMENTARY STUDENTS´ METACOGNITIVE PROCESSES AND POST-PERFORMANCE CALIBRATION ON MATHEMATICAL PROBLEM SOLVING

TRINIDAD GARCÍA, CELESTINO RODRÍGUEZ, PALOMA GONZÁLEZ-CASTRO, JULIO ANTONIO GONZÁLEZ-PIENDA & MARK TORRANCE*

UNIVERSITY OF OVIEDO (SPAIN), * NOTTINGHAM TRENT UNIVERSITY (UNITED KINGDOM)

SUBMITTED FOR PUBLICATION

ABSTRACT. Calibration, or the correspondence between one's perception of performance and one's actual performance, is an important metacognitive mechanism linked to learning, Making students more aware of the quality of their performance is important in elementary school settings, even more so when mathematical problems are involved. However, students seem to be poorly calibrated, with a tendency towards over-confidence. The present study analyzes the relationship between post-performance calibration accuracy and the metacognitive process shown by 524 fifth- and sixth-grade students while solving two mathematical problems. Once an index of calibration was calculated and the stability of students´ judgments and actual performance was established, differences in the metacognitive process exhibited by students with different calibration accuracy (Accurate vs. Inaccurate) were analyzed. The emergence of different patterns of calibration and differences in the metacognitive process as function of mathematics achievement and grade level was also examined. Results indicated that: 1) students in the general group were little calibrated and over-confident, showing high stability in their judgments and actual performance; 2) inaccurate students reported using information representation sub-processes (drawing/summarizing) less frequently, but writing and reviewing (and also correcting mistakes) more frequently than the accurate group, which could be related to the use of "trial and error" mechanisms; 3) differences in calibration patterns and the metacognitive process were found when achievement level was considered, while grade level did not generate any important effect. These findings suggest the usefulness of process-based measures to examine the metacognitive processes involved in making post-performance judgments, considering achievement and its mediating role in this relationship.

Keywords: Calibration, Mathematics, Metacognition, Problem solving, Process, Self-regulation.

1. BACKGROUND

Solving mathematical problems is a very common activity in elementary school. However, it is also a complex cognitive activity that involves multiple processes. In order to successfully perform these tasks, students have to integrate cognitive, metacognitive and self-regulatory mechanisms (Cleary and Chen 2009; Montague et al. 2011). In this way, they have to accurately regulate and monitor their learning processes. An important aspect of monitoring is *calibration*, or the degree to which students' judgments about the correctness or adequacy of their performance correspond to their actual performance, the latter determined on the basis of an objective measure (Hacker et al. 2008a). These judgments can be expressed in terms of *predictions* or *postdictions*, depending on whether they are made prior to or after completing a task, and both involve different processes. Specifically, predictions would be related to a sort of self-confidence judgments, whereas postdictions (or post-performance judgments) would be informative of control processes during the task (McCormick 2003; Hacker et al. 2008a). As will be discussed later on this paper, the distinction between predictions and postdictions is especially relevant in the context of the present study as it focuses on analysing how the metacognitive process involved in solving mathematical problems is related to students´ judgments of accuracy after completing the task.

The relevance of calibration mechanisms in mathematics has been substantially demonstrated, suggesting a direct and positive relationship between these metacognitive control processes and academic achievement in the subject. Within the context of mathematical problem solving, it has also been demonstrated that highly calibrated students tend to perform more successfully than lower calibrated students on these tasks, presumably because they exercise a better control over their problem-solving processes (Desoete and Roeyers 2006, Jacobse and Harskamp 2012; Lipko et al. 2009; Özsoy 2012; Rinne and Mazzocco 2014; Hadwin and Webster 2013). However, previous literature has shown that students are commonly poorly calibrated, showing a tendency towards over-confidence, and that these pattern of inaccuracy tend to remain persistent (Bol et al. 2012; Dinsmore and Parkinson 2013; Hacker et al. 2008b; Stolp and Zabrucky 2009). Given the implications that different patterns of calibration accuracy may have on students´ learning, it is necessary to determine which factors influence these judgments. In this sense, a revision of the current literature shows that: 1) most research has focused on the study of these mechanisms in undergraduate or secondary schools students, giving less attention to elementary school ages; 2) most studies have been conducted in laboratory settings, which may be an important constraint for results generalization; 3) whereas both predictions and postdictions have been substantially studied regarding mathematics problems, the amount of research conducted from the perspective of the analysis of the process involved in these tasks is almost inexistent to date, even more if we focus on the metacognitive nature of this process.

According to that, the main objective of the present study was to analyze the relationship between students' post-performance calibration accuracy and the metacognitive process they showed while solving two mathematical problems, having a broad sample of fifth- and sixth-grade students for this purpose. To address this objective, five hundred and twenty-four students were assessed in a regular mathematics class, using the Triple Task Procedure in Mathematics (TTPM; García and González-Pienda 2012; García et al. in press) as a measure of the process. Students were divided into two groups based on their judgments accuracy (accurate vs. inaccurate) after solving each problem, and differences between groups in the metacognitive process during TTPM were analysed. Prior to analysing these differences, an index of calibration accuracy was calculated, and the stability of students´ judgments and actual performance was examined. The metacognitive process shown by the students in the general group was also described. Finally, the possible emergence of different calibration accuracy patterns and differences in the metacognitive process as function of mathematics achievement and grade level were analyzed. In order to accomplish these goals, this paper is organized in two different parts:

The paper begins by setting out the research context for the present study. This first half of the paper focuses on defining calibration and the most commonly used measures, followed by a description of the main findings from previous literature regarding the relationship among calibration, metacognition, self-regulation, and mathematics problem solving. The main assessment tool (TTPM; García and González-Pienda 2012; García et al. in press) is also exhaustively described in this part, paying special attention to the specific models that serve as a basis for its design and some methodological and technical issues related to the assessment procedure.

The second half of the paper presents the methodology used in the present study, as well as the main results and their discussion. An example of students' artifacts during TTPM is provided in order to better illustrate the differences in the metacognitive process evidenced by the statistical analyses. The paper ends with a discussion of the main limitations of the present study and some suggestions for future research. Some practical implications of the results obtained in the present study are also discussed.

On the whole, the present study is an attempt to provide the study of the relationship among calibration accuracy, metacognition and mathematical problem solving in elementary school with some additional evidence through the analysis of the process involved in solving mathematical problems, using the TTPM for this purpose. Even recognizing its possible limitations, the emphasis given to the process makes this study a novel contribution to this field. Further research is necessary, however, in order to better determine the potentiality of this approach.

1.1. Definition of calibration

Calibration has been characterized as an important metacognitive process involved in the development of self-regulatory competence (Dinsmore and Parkinson 2013; Hadwin and Webster 2013; Labuhn et al. 2010; Zimmerman and Schunk 2011). Because of its implications in students' motivation, metacognitive control and self-regulation, it is an essential process affecting academic success and task completion (Alexander 2013; Cleary 2009; Efklides and Misailidi 2010; Schunk and Pajares 2009). In this sense, as Alexander pointed out, students' ability to accurately gauge their progress and performance plays an important role in their subsequent effort and strategy use in problem-solving situations. In this context, mathematical problem solving is an important issue in current scientific literature. Characterized as a complex cognitive activity, students tend to show poor metacognitive skills when they are engaged in mathematical problem-solving situations. Many students (even those without learning disabilities) forge ahead without considering alternative approaches, jumping immediately into calculations, giving impulsive responses, and using trial and error as a strategy while trying to solve mathematical problems. They commonly get stuck in irrelevant details of the task or fail to verify solution paths and evaluate the suitability of their responses (Cleary and Chen 2009; Kramarski and Gutman 2006; Montague et al. 2011; Pennequin et al. 2010; Pereis et al. 2009). Literature also indicates that students tend to be inaccurate when they are asked to predict or judge the results of their performance either in mathematics and other academic areas, showing a tendency towards over-confidence (Bol and Hacker 2001; Bol et al. 2005, 2010; Hacker et al. 2008a; Lipko et al. 2009; Özsoy 2012). This is an important aspect, as whether or not students' metacognitions are calibrated to their performance has important consequences for learning outcomes (Dunlosky and Rawson 2012; Dunlosky and Thiede 2013; Finn and Metcalfe 2014; Hacker et al. 2008a).

1.2. Calibration measurement

Whereas the relevance of calibration in learning has been substantially demonstrated, an important issue that must still be addressed is the distinction between *absolute* and *relative calibration*. The term *calibration* used in the present study refers to the *absolute accuracy* or the degree of correspondence between one's judged level of performance and one's actual performance (Hacker et al. 2008a; Maki et al. 2005). Absolute accuracy (or calibration) differs from the concept of *resolution or discrimination*, referred to as *relative accuracy,* and traditionally defined as the degree to which a person's judgments can predict the likelihood of correct performance of one item relative to another (Hacker et al. 2008a). Both types of calibration are different aspects of metacognitive monitoring, showing low correlations with each other (Maki et al. 2005).

In calibration studies, learners are asked to make a confidence judgment, which consists of judging the degree to which they think they will be able to learn some knowledge or perform a task (prediction) or judge their performance once the task has been completed (postdiction). This judgment is then compared with an objective measure of that learning (their actual performance), such as the score in an exam or test, after they complete the task (Bol et al. 2010; Winne 2004). The more closely a student's judgment of performance matches his/her actual performance, the better calibrated he/she is (Hacker et al. 2008a). In the context of the present study, the distinction between prediction and postdiction acquires a special meaning. In fact, authors such as McCormick (2003) and Hacker et al. (2008a) suggest that both involve different processes. Thus, predictions (or prospective monitoring judgments) could be thought of as a type of self-confidence judgment, which reflects a person's belief in his or her mastery of some learning or task, whereas a postdiction judgment (or retrospective monitoring judgment) could inform about control processes, providing learners with more accurate feedback on their monitoring proficiency. Postdictions have also been found to be more accurate than predictions (Ackerman and Wolman 2007). Although postdictions have been less studied than predictions, many authors incorporate these measures in their studies (Bol and Hacker 2001; Bol et al. 2005, 2010; Hacker et al. 2008b; Nietfeld et al. 2006; Sheldrake et al. 2014).

Among the methods to calculate calibration, Parkinson et al. (2010) note dichotomous ratings, categorical ratings (such as Likert-type scales), data count, and 100-mm scales as the most commonly used measures. In the case of studies using dichotomous ratings, students are commonly asked to complete a multiple-choice recall measure of whether they feel "confident" or "not confident" about each answered item (Schraw et al. 2012), whereas in the case of Likert-type scales, they are asked to rate the degree to which they feel confident, for instance, ranging from "not confident" to "very confident" (Hattie 2013). Within the numerous indexes used to express judgment accuracy, different measures are feasible, from Pearson correlations to more complex measures, such as the rho or the Gamma coefficient (Dinsmore and Parkinson 2013; Schraw 2009; Winne and Muis 2011). However, the simplest index consists of expressing calibration in terms of percentages (Bol et al. 2005; Hacker et al. 2008a). For instance, if two students judge they will respond correctly to 90% and 80%, respectively, of the questions on an exam, but they finally respond correctly to 70%, both of them were over-confident in their judgments. However, the second student was more highly calibrated. In the present study, a dichotomous measure of calibration was used. Specifically, Phi coefficient was calculated (Schraw et al. 2012). As these authors point out, although continuous measures may be more sensitive, dichotomous judgments are most common in the literature. This may be motivated by the fact that a dichotomous scoring scheme may be easier for participants to understand than a more complex scheme, which is especially relevant when studies are conducted with young samples. Although there are several dichotomous measures of calibration, such as the G index, the Odds ratio, Gamma, Kappa, Phi, or the Sokal distance measure, Schraw and colleagues demonstrated in their study that all of them show high and similar sensitivity and specificity.

As mentioned above, an important issue in calibration studies is the fact that students tend to be inaccurate in their performance judgments, with a tendency to over-confidence (Bouffard et al. 2011; Butler 2011; Dinsmore and Parkinson 2013; Hadwin and Webster 2013). This tendency has been found in studies at different educational stages, across different subjects areas, and in mathematics in particular (Bol and Hacker 2001; Bol et al. 2005, 2010; Rinne and Mazzocco 2014; Sheldrake et al. 2014). Furthermore, there is a good deal of evidence suggesting that calibration judgments tend to be stable (Bouffard et al. 2011; Hacker et al. 2008a). These studies indicate that performance judgments are resistant to improvement. This has led to an important line of research on the determining factors behind students´ judgments and how to improve these processes. In this context, many studies have focused on exploring students' explanations for their confidence ratings. Aspects such as students' general cognitive ability or prior knowledge, academic achievement or their level of expertise in making such judgments; task or item characteristics; affective-motivational components such as goal orientation setting or self-efficacy beliefs; as well as more stable and persistent traits, such as attribution styles or personality, have been proposed as possible explanatory factors (Alexander 2013; Bol et al. 2012; Dinsmore and Parkinson 2013; Hacker et al. 2008b; Hadwin and Webster 2013; Stolp and Zabrucky 2009).

1.3. Metacognition, Self-Regulated Learning and Calibration in Mathematics

Within this perspective, another important determining factor of students' judgment accuracy is the link between calibration and metacognitive and self-regulatory mechanisms. Although both terms (metacognition and self-regulation) have been used as interchangeable on some occasions (Dinsmore et al. 2008), it is important to note that there is an important difference between them. In this sense, self-regulation could be defined in the context of learning as the control that students exert over their cognition, behaviour, emotion, and motivation in order to achieve established goals, whereas metacognition (or the ability to think about thinking) would refer to the cognitive component of self-regulation, in charge of processes such as self-awareness of problem-solving, monitoring and controlling one's mental processing (Dinsmore et al. 2008; Panadero and Alonso-Tapia 2014). Nevertheless, both metacognitive and self-regulatory components are articulated under the paradigm of Self-regulated Learning (SRL: Zimmerman 2000, 2008). SRL explores the acquisition, evaluation, and regulation of knowledge, with a conception of learners as being capable of monitoring their own learning, setting goals, evaluating, and regulating their own progress (Stolp and Zabrucky 2009). This perspective is intimately linked to the concept of calibration, or one's ability to accurately judge one's performance on a task. Calibration is relevant in this context, as it consists of a metacognitive monitoring process that provides information about the status of one´s knowledge and strategies at a cognitive level and sets the stage for effective self-regulation by generating the internal feedback that students use to control their learning and performance (Bol et al. 2012; Dunlosky and Rawson 2012). In this sense, given that Zimmerman's (2000) model establishes that self-regulated learners are more aware of what they do and do not know, it is expected that they exhibit more accurate perceptions of performance. This statement is supported by studies showing that self-regulated learners are commonly higher achieving students, and also more calibrated, than their lower achieving peers. These studies also showed that high-achieving students tend to be somewhat under-confident, whereas lower-achieving students tend to be inaccurate and over-confident when they are asked both to "predict" or "postdict" their performance on a task (Bol and Hacker 2001; Bol et al. 2005, Hacker et al. 2008b; Özsoy 2012). These results support the role of calibration judgments as a metacognitive and self-regulatory process, as well as their relationship with student achievement.

Additionally, the impact of metacognition and calibration mechanisms on mathematics, and problem solving in particular, has been substantially demonstrated (Desoete and Roeyers 2006; Jacobse and Harskamp 2012; Özsoy 2012; Rinne and Mazzocco 2014). On the one hand, calibration accuracy has been shown to explain about 16 to 36% of the variance of mathematics achievement in different studies (see Jacobse and Harskamp 2012). In this sense, a recent study conducted by Rinne and Mazzocco (2014) supported this association, suggesting that calibration accuracy on mental arithmetic judgments may represent an important developmental predictor of future mathematics performance in Elementary School. On the other hand, highly calibrated students have been shown to perform more successfully on mathematical problem-solving tasks. Specifically, previous studies suggest that highly calibrated students would be able to solve more complex problems by using strategies such as separating the tasks into simpler pieces, or building different ways to represent information, concepts, and relationships. They would also ask themselves questions to clarify their thoughts and would tend to evaluate the result of their performance more frequently than their low calibrated peers (Özsoy 2012). Hence, it would seem that highly calibrated students show more self-regulatory and metacognitive skills, which supports the consideration of calibration as an important metacognitive monitoring process (Boekaerts and Rozendaal 2010; Dinsmore and Parkinson 2013; Hadwin and Webster 2013; Stolp and Zabrucky 2009; Winne and Muis 2011). However, previous research has also revealed that many students rarely use self-regulatory and metacognitive mechanisms spontaneously during mathematical problem-solving tasks, which leads to poor performance and inaccurate self-evaluations (Kramarski and Gutman 2006; Pennequin et al. 2010; Pereis et al. 2009). In this context, one of the greatest challenges for researchers is to obtain evidence about the cognitive, metacognitive, and strategic processes exhibited by students during these tasks, and about how these processes relate to their ability of making accuracy judgments. This information may be gathered by means of on-line measures, such as Think-Aloud (Jacobse and Harskamp 2012; Montague et al. 2011) or Triple Task protocols (Olive and Piolat 2002; Piolat et al. 2001, 2005).

1.4. Process-based measures: the Triple Task Procedure in Mathematics (TTPM) and its predecessor

Self-regulatory and metacognitive mechanisms have been traditionally assessed by means of questionnaires or structured interviews. These kinds of measures are based on students' self-reports about how they tackle a problem, or the extent to which they use different strategies. Students are also asked about different situations in which they would (or would not) use a specific strategy, or which approach would be the best in each case. Students' responses are scored depending on the quality of their responses, and a total score is calculated. However, this kind of assessment (although widely applied) may yield inconsistent or incorrect information due to inaccurate memory or response biases, such as social desirability. Hence, while these tools have proved to be useful to assess students' declarative and situational knowledge, they do not provide information about how learners transfer this knowledge to regulate their problem-solving process (Cleary and Chen 2009; Veenman 2011). In order to address this issue, an alternative is the use of on-line methods (i.e., process-based measures). These measures are taken concurrently with task performance and have been shown to be especially useful providing information about students' cognitive processes during different activities (Azevedo and Aleven 2013; Throndsen 2011; Tillema et al. 2011).

An example of these measures is the *Triple Task* technique (Piolat et al. 2001). This procedure has a long tradition in the study of the processes involved in composition writing (Olive and Piolat 2002; Piolat et al. 2005). In traditional Triple Task studies, participants are required to perform three tasks simultaneously: a primary task (e.g., the composition of a text); a second probe task (based on response time-RT); and a third task, in which they are asked to verbalize or categorize the actions or thoughts that are interrupted by the probe. This method uses directed retrospection. In this sense, probes are presented at a certain time interval, and after each reaction to a probe, writers are asked to categorize their actions or thoughts according to a given category system regarding different writing sub-processes. Aimed at providing information about the activation of different writing sub-processes and cognitive demand as the written task progresses, this protocol is based on two measures: 1) retrospection data, which provide information about how the writing sub-processes are activated over time; and 2) RT´s to the probes, which allows studying the cognitive effort involved in these sub-processes. Specifically, RTs are taken before the task (baseline condition) as well as during the task. RTs taken during performance are called "secondary

RTs" and are linked to a reported action of sub-process. Differences between RTs in both conditions are interpreted as an estimation of the cognitive effort associated with each writing sub-process or category. Before starting the tasks, participants are trained to recognize each of the sub-processes of the given category system. As a measure of the process, the Triple Task procedure differs from Think-aloud protocols (Montague et al. 2011) in two main aspects: 1) Triple Task procedure uses directed retrospection, as opposed to the undirected retrospection used in Think-aloud. In this sense, providing a system of categories facilitates students' categorization of their thoughts or actions and allows the experimenter to limit the number of sub-processes to study, exploring those of special interest; 2) response transcription and coding are not required. This aspect makes the Triple Task technique less costly than Think-aloud, which makes it more suitable to conduct studies with larger sample sizes.

Within this context, an adaptation of the Triple Task method has been recently proposed for application in mathematical problem solving (García and González-Pienda 2012; García et al. in press). The *Triple Task Procedure in Mathematics* (TTPM) is a modification of the traditional Triple Task. It has been designed to study the process underlying solving mathematical problems from the viewpoint of the articulation between the main SRL phases of planning, execution, and evaluation (Rosário et al. 2008; Zimmerman 2000), and Bransford and Stein´s (1993) IDEAL model. García et al. administered this protocol to 510 fifth- and sixth-grade students from Northern Spain, who carried out mathematical tasks of varying difficulty. Students in this study showed a lack of revision mechanisms and ineffective planning strategies. However, students who successfully solved the problems showed significantly better metacognitive mechanisms. These differences gained prominence as task-difficulty seemed to increase, suggesting a relationship among problem-solving performance and metacognitive processes, and maybe task difficulty. A brief review of the SRL and problem-solving models used in TTPM will be made before the proposed evaluation system is described in this section.

1.4.1. Self-Regulated Learning model

There are different theories and models that attempt to explain SRL mechanisms. All of them share the common ground that self-regulation involves different processes (e.g., planning, monitoring, etc.) and is cyclical (i.e., previous performance provides feedback for future tasks: Panadero and Alonso-Tapia 2014). These authors suggest the models purposed by Zimmerman (2000, 2008), Winne (2001), and Boekaerts (1999) as the most representative of the last few years. Within this context, Zimmerman's model (2000, 2008) has been the most frequently cited in the literature, and extensively used to guide interventions (Moos and Ringdal 2012; Stoeger and Ziegler 2008). These practical implications explain why this model was elected as a framework for the TTPM design. This model comprises three cyclical phases (forethought, performance, and self-reflection), which correspond to the processes that occur before, during and after SRL takes place. In the forethought phase, students analyze the task and assess their capacity to perform it, establishing goals and plans about how to complete it. The second phase is performance, in which two types of processes are carried out: self-control and self-observation. This phase is aimed at keeping track of one's progress during the task and involves employing the strategies identified during the previous phase of forethought. These strategies can be classified as metacognitive strategies (e.g., self-instruction, imagery, time management, etc.) and motivational strategies (e.g., self-given messages to remind one of the goals, self-rewards, etc.). The last phase is self-reflection, in which students judge their performance and formulate reasons for their results (Panadero and Alonso-Tapia 2014).

1.4.2. Problem-solving model

Several problem-solving models have been proposed from different conceptual perspectives. This study focuses on the perspective of cognitive psychology, which emphasizes the mental processes underlying problem solving. These models conceive the application of general problem-solving strategies as the basis for problem-solving expertise and performance. These strategies are called "heuristics" and involve different processes (e.g., identifying the problem, drawing a graph or diagram, thinking about structurally similar problems, simplifying the problem, reviewing, or generalizing), which are commonly expressed as a series of stages through which problem-solving progresses. The models of Polya (1954), Gick (1986), and the IDEAL Model of Bransford and Stein (1993) have been the most studied. The IDEAL Model describes five stages: identifying potential problems, defining and representing the problem, exploring possible strategies, acting on those strategies, and looking back and evaluating the effects of those activities. It is one of the most extended models today (Newton et al. 2009, 2012) and, in contrast to the other mentioned models, it establishes that the problem-solving process is cyclical, according to Zimmerman's SRL Model (2000, 2008). As it also establishes a comprehensive system to examine the metacognitive and self-regulatory processes involved in problem solving during performance, this model makes up the other basis for TTPM.

The combination of both Zimmerman's (2000) SRL model and the IDEAL model (Bransford and Stein 1993) produces a system with eight categories or sub-processes, organized into three higher-level categories, corresponding to the SRL phases (Rosário et al. 2008; Zimmerman 2000): Planning, Execution , and Evaluation. Table 1 shows the category system used, obtained from the combination of the two aforementioned models. In accordance with previous studies, an additional category ("other") has been included to gather all the thoughts or activities unrelated to the mathematical problems.

TABLE 1

Category system. Based on the Self-Regulation Model (Rosário et al. 2008; Zimmerman 2000) and the IDEAL Model (Bransford and Stein 1993)

SRL Model	IDEAL Model	Process categories (I am ...)
Planning	Identification of the problem	Reading
	Definition and representation	Drawing or summarizing Recalling similar problems
	Exploration of possible strategies	Thinking about a solution
Execution	Action based on the strategy	Calculating Writing a response
Evaluation	Look at effects of solutions	Reviewing Correcting mistakes
"Other"		Doing something unrelated

An important issue in this category system is the allocation of drawing/summarizing sub-process in the SRL planning phase. In this sense, imagery, or the construction of mental images to organize the information, would be similar in essence to drawing/summarizing and falls under the performance phase in Zimmerman's (2000, 2008) SRL model. Specifically, it consists of a metacognitive strategy under self-control, as mentioned above. The location of this strategy under the planning phase is due to the problem-solving model adopted (DEAL model), in which the Definition and representation stage is prior to execution (or Action based on the strategy). Previous studies using TTPM (García and González-Pienda 2012; García et al. in press) confirmed the allocation of this sub-process under the SRL planning phase. This last study analyzed the temporal sequence of the process shown by a sample of 350 elementary students during TTPM. This sequence was based on the percentage of students who reported each of the eight TTPM categories or sub-processes across 20 different moments while solving several mathematical problems. Results indicated that, although drawing/summarizing tends to be present both in planning and execution phases, this process is mainly activated during the first phase, while its frequency decreases dramatically as the frequency of the calculation sub-process (execution phase) increases. The same pattern occurs in the case of thinking about solutions and recalling similar problems (to a lesser extent), both processes identified under the SRL planning phase.

One of the potentialities of this assessment procedure relies on its design and implementation features (see Procedure section in the present study), which facilitate its applicability to broad sample sizes, and more important, in classroom settings (García et al. in press). In this sense, one of the most discussed aspects in calibration studies is the fact that they are commonly conducted in laboratory settings (Dinsmore and Parkinson 2013; Hadwin and Webster 2013; Winne and Muis 2011). As Hacker et al. (2008a) pointed out, although these contexts often provide important information on calibration accuracy mechanisms, generalizing these findings to different contexts, especially to classroom contexts, can sometimes be difficult. Laboratory study components, material and procedures, such as highly structured context, tasks with little meaning for students, or presented over short periods of time, differ considerably from those used in classrooms settings, where students are required to deal with multiple requirements from the environment and the task. Additionally, although a large amount of research has focused on the study of the relation between calibration, metacognitive and self-regulatory mechanisms in mathematics (Bol et al. 2010; Hacker et al. 2008a, 2008b; Rinne and Mozzocco 2014; Sheldrake et al. 2014), and mathematical problem-solving in particular (Jacobse and Harskamp 2012; Zimmerman et al. 2011), none of these studies have been conducted from the perspective of the analysis of the relation between students' judgments of accuracy and the metacognitive process involved in solving mathematical problems.

1.5. The present study

This study examined the metacognitive process shown by a sample of 524 fifth- and sixth-grade students while solving two mathematical problems, and how this process may explain differences in students' accuracy judgments once the tasks are finished. For this purpose, TTPM (García and González-Pienda 2012; García et al. in press) will be used as a measure of the process. Additionally, the possible influence that variables such as mathematics achievement and grade level may have both on metacognitive processes and calibration accuracy will be analyzed. This was done because, whereas the relation between students' achievement in calibration has been substantially demonstrated (Bol et al. 2005, 2010; Hacker et al 2008b; Özsoy 2012), it is not clear whether grade level may follow the same pattern. However, as different grade levels may be indicating different cognitive or developmental levels (Boston and Smith 2009); it is possible that this variable will exert some degree of influence on the variables of this study.

Overall, this study aimed to answer the following questions: 1) Do students with different accuracy in their post-performance judgments show differences in their problem-solving processes?; 2) Which SRL phases or sub-processes could better explain these differences?; and 3) Will different patterns of calibration accuracy emerge as function of students' achievement in mathematics and grade level, and could these variables account for differences in the metacognitive processes? Prior to these analyses, an estimation of the correspondence between students' post-performance judgment and actual performance was calculated in order to establish the degree to which the students were calibrated and the tendency of their judgments (over- or under-confidence) in both problems. In addition, post-performance judgments and actual performance stability across problems was analyzed. Previous laboratory and classroom studies have shown that calibration tends to be relatively stable across time and tasks. Specifically, previous studies indicated that students tend to be more stable making performance predictions than they are in their actual performance (Bol et al. 2005; Bouffard et al. 2011; Hacker and Bol 2004; Hacker et al. 2008b). Given these assumptions and according to previous research results, we expect that:

1) Students will be imprecise in their judgments, showing a tendency towards over-confidence.
2) Students will show stability in their judgments and actual performance across problems. This stability may be higher in judgments than in actual performance.
3) Differences in the metacognitive process (TTPM phases and sub-processes) are expected to be found between students who show different accuracy in their post-performance judgments (accurate group = when students' judgments and actual performance match; or inaccurate group = when they do not). Students in the first group will show different and more efficient metacognitive strategies during TTPM, such as organization and representation of the information. They will also evaluate the result of their performance more frequently than their peers in the other group.

4) It is possible that different patterns of calibration accuracy emerge when students' mathematics achievement is considered. High achievers are expected to be more accurate and maybe more under-confident than their low-achieving peers. This variable could also explain differences in the metacognitive process during TTPM, mainly in those activities related to planning.

5) Differences in grade level could lead to different patterns of calibration accuracy, as well as differences in the metacognitive process during TTPM. Sixth-grade students may show higher levels of calibration accuracy, as well as better mechanisms of planning and revision.

2. METHOD

2.1. Participants

Five-hundred and twenty-four students from fifth and sixth grade of Primary school took part in this study. They were recruited from 11 schools in Northern Spain. Ages ranged between 10 and 13 years (M = 10.991, SD = 0.716). Of these students, 260 (49.6%) were female, and 264 (50.4%) were male. The sample comprised 220 students from the fifth grade (42%; male = 108, female = 112) and 304 students from the sixth grade of elementary school (58%; male = 156, female = 148).

Sample selection was made through convenience or accessibility procedures. Students volunteered for the study and presented informed consent from their parents. Children with a diagnosis of severe learning disabilities or those who showed reading difficulties were excluded from the analyses.

Given that data analyses were conducted separately per each mathematics problem, students were distributed differentially based on the accuracy of their post-performance judgments in each problem (accurate versus inaccurate). Both post-performance judgments and actual performance were expressed in dichotomous terms (success =1, failure = 0). Students were assigned to the accurate group when there was a correspondence between their post-performance judgments and actual performance and to the inaccurate group when there was a mismatch. Sample characteristic are shown in Table 2.

TABLE 2

Characteristics of the sample

Demographic variables	Problem 1		Problem 2		Total Sample
	Accurate	Inaccurate	Accurate	Inaccurate	
N	221/42.2%	303/57.8%	278/53.1%	246/46.9%	524
Age (*M*/*SD*)	10.991/.729	10.993/.707	10.992/.731	10.994/.700	10.991/.716
Gender (female)	109/49.32%	151/49.83%	136/48.9%	124/50.4%	260/49.62%
Gender (male)	112/50.68%	152/50.17%	142/51.1%	122/49.6%	264/50.48%
Level (fifth)	97/43.89%	123/40.59%	111/49.3%	109/42.68%	220/42%
Level (sixth)	124/56.11%	180/59.41%	167/50.7%	137/57.32%	304/58%

Note. *M* = Mean, *SD* = Standard Deviation

There were no statistically significant differences between groups with different calibration accuracy either in age (p = .954), in grade level (p = .489) and gender distribution (p = .253) in Task 1. Regarding Task 2, gender (p = .177) and grade level (p = .405) were equally distributed in the different groups. There were no age differences between accurate and inaccurate students (p = .921).

2.2. Measures

In accordance with the objectives proposed in this study, three sets of measures were used: those related to the problem-solving process (TTPM), those applied to calculate post-performance calibration, and a measure of students' achievement in mathematics.

2.2.1. Problem-solving process

Evidence from students' problem-solving processes was obtained using the *Triple Task Procedure in Mathematics* (TTPM; García and González-Pienda 2012; García et al. 2013), previously described in the present study. The main characteristics of TTPM are related to the designed category system, based on Bransford and Stein's (1993) IDEAL Model and the SRL perspective (Rosário et al. 2008; Zimmerman 2000), and comprised of eight sub-processes (reading, drawing or summarizing, recalling similar problems, thinking about a solution, mental calculation, writing a response, reviewing, correcting mistakes) and three main phases (Planning, Execution and Evaluation). Dependent variables were students' relative frequency of election of each category and phase, expressed in percentages.

2.2.2. Post-performance calibration

Three measures were established (actual performance, post-performance judgments and a calibration index):

Actual Performance: it was obtained by asking students to complete two mathematical word problems taken from the book *"Problem-solving and comprehension"* (Whimbey and Lochhead 1993) translated into Spanish. This book is a systematic review of the characteristics that define a good problem solver, while proposing strategies and activities to develop these skills, mainly based on mathematical problem-solving situations. Actual performance was expressed as success (1) or failure (0). Although the second problem involves two different questions (number of dogs and their breeds), no partial scores were awarded in order to unify the measurement system. Prior to commencing the study, these problems were reviewed by teachers to ensure that the mathematical problems were developmentally appropriate for the age of the students. The two mathematical problems are shown below:

Problem 1*: "Beatriz lends €700 to Susana. But Susana borrows €1500 from Ester and €300 from Juana. In addition, Juana owes Ester €300 and Beatriz €700. One day they meet at Beatriz*'*s home to settle their debts. Who went back home with €1800 more than she brought?"*

Problem 2: *"Paula, Mari, and Juana have a total of 16 dogs, 3 of which are poodles, 6 are hounds, and the rest of them are German shepherds and Pekinese dogs. Juana does not like poodles and Pekinese dogs, but she has 4 hounds and 2 German shepherds, giving a total of 6 dogs. Paula has a poodle and 2 more dogs, which are German shepherds. Mari has 3 Pekinese dogs and several dogs of other breeds. Which breeds, and how many dogs of each breed, does Mari have?"*

Post-performance judgments: they were obtained by asking students after each problem whether they considered that they had solved the problem successfully. They had to respond in dichotomous terms (yes/no), and the values of 1/0 were assigned, respectively. Post-performance judgments were expressed in dichotomous terms based on the characteristics of the problems (i.e., actual performance was established as success or failure), but also on the sample itself (i.e., fifth- and sixth-grade students), in an attempt to provide them with an easier scoring scheme.

Calibration index: an index of calibration was calculated in terms of the correspondence between post-performance judgments and actual performance. As both variables were dichotomous, Phi correlation coefficient ($\rho\phi$) was used as a calibration index.

2.2.3. Mathematics achievement

Students' final academic grades in mathematics were used as indicator of achievement in this subject. This information was given by the teachers in charge of the subject, once permission from parents had been obtained. Academic grades on a scale of 0 to 10 points were used at first to analyze differences in achievement between groups with different accuracy in their post-performance judgments. Then, three levels of achievement were distinguished: low (grades of 5 or below), medium (grades between 5 and 7.5) and high achievement (grades over 7.5). This distinction was made according to the grading system commonly used at Spanish schools. Different levels of achievement were established in order to examine how differences in this variable may alternatively explain different patterns of calibration accuracy, as well as differences in the metacognitive process during TTPM.

2.3. Procedure

The study was conducted in accordance with The Helsinki Declaration of the World Medical Association (Williams 2008), which reflects the ethical principles for research involving humans. The evaluation was collectively administered during a regular class, in the schedule devoted to mathematics. Students worked with their personal computers in their regular classroom. This session was part of a broader research project, and this was the third time that examiners and students worked together. Although two trained examiners carried out the evaluations, different groups of students were assigned to each one, so that the children had only one examiner as a reference. Teachers were not present in this evaluation session.

Participants in this study were not given any incentive to take part. However, both the materials and the evaluation procedure in the present study were designed so that they were perceived by students as appealing and motivating. Figure 1 shows an example of the interface used to present the category system, where each category is represented both through text and graphic symbols. In addition, the hypothetical case of a boy of their age (Alex), who tried to solve a mathematical problem, was used to introduce the TTPM procedure and guide the training phase. Prior to starting the evaluation, students were told that it was not an exam, but they should do their best because the main goal of this study was to know how they solved mathematical problems in order to help them to improve their problem-solving skills and get better grades.

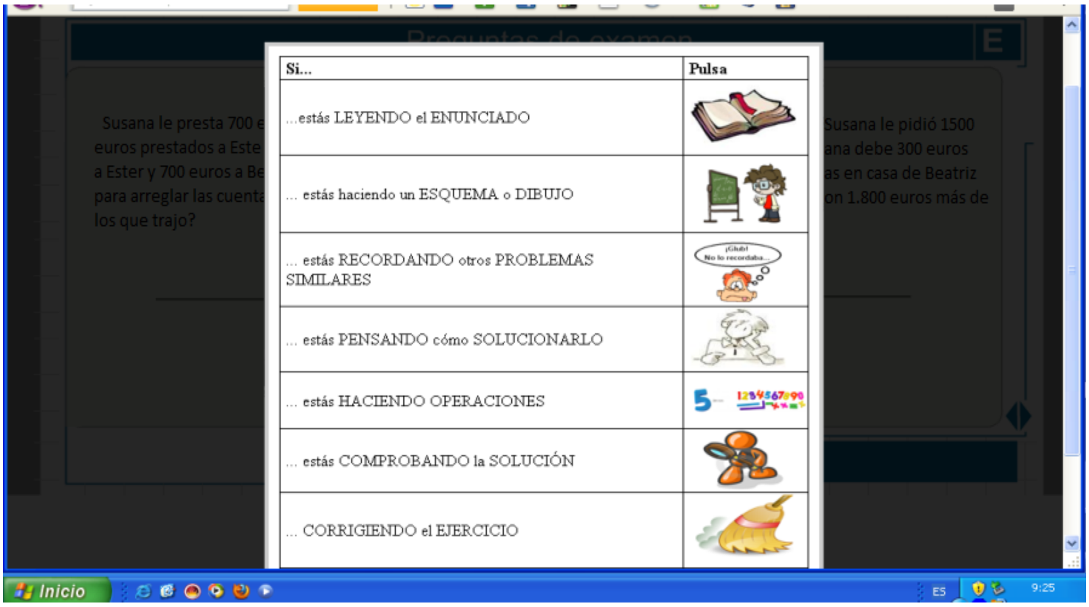


Figure 1. Example of category system presentation during TTPM.

This figure shows seven of the nine categories (reading, drawing/summarizing, recalling, thinking, calculation, reviewing, and correcting)

The first phase consisted of training in order to help students to familiarize themselves with the system of categories and the assessment procedure. This phase started by presenting Alex and the different actions he performs while solving a problem. After training, students performed a category-recognition test consisting of 18 multiple-choice items (2 per category) with four response alternatives. Students had to indicate the category that best expressed each proposed activity in order to test their understanding of the nine categories, including the "something unrelated". For instance, for the statement, "Alex realized he made a mistake, so he is erasing it," the alternatives were: "Alex is thinking about a solution," "Alex is writing," "Alex is reviewing," or "Alex is correcting mistakes." The length of this phase is variable and depends on students' involvement (comments, questions, etc.). However, it must never extend for more than 20-25 minutes (García et al., in press). Students with scores below 16 points on the category-recognition test (around 90% of correct responses) were excluded from the analyses. Originally, the sample was comprised of 567 students, of which 43 (7.58%) scored below the cut point. The remaining 524 students (included in this study) showed a mean score of 16.984 (*SD* = 0.493) on this test.

The reason why only two mathematical problems were used in the present study was motivated by the design of the TTPM, which requires that training and evaluation phases should be conducted in the same session. As a training phase is needed to ensure that the children understand the assessment procedure and category system, this phase acquires special relevance and is an important part of the session. Additionally, this method is designed so that it matches the length of a regular class session (40-45 minutes). Thus, time must be carefully assigned to the different components of TTPM, taking into consideration that students must perform the problems without time pressure.

Once the system of categories was understood, students completed the Response Time base-line task. However, as it extended the aims of the present study, a measure of cognitive effort was not calculated. Students were informed that they would occasionally hear a beep from the computer at variable intervals, and were asked to quickly react to every probe by a mouse-click. Beeps were presented at random intervals of 10-15 seconds. A total of 30 beeps were presented. Then, students were informed that this same sound would appear while performing the mathematical problems, after which a box with the category system would be displayed on the computer screen. They would have to choose the category that best represented what they were doing in each moment. Auditory probes (or beeps) during performance were presented in intervals of 40-45 seconds. Their presentation lasted until students indicated they had finished each problem by clicking on a button set up for this purpose. This task was designed so that, regardless of each student's response speed and the moment they started or finished each problem, the time intervals among probes were the same for all the participants. Although RT´s were taken before and during TTPM administration, the analysis of this variable was not incorporated as it exceeds the purpose of the present study.

An important constraint in the Triple Task technique is the choice of the time interval with which the probe interrupts the primary task. Piolat et al. (1996) examined whether the choice of a particular time interval affects text production, cognitive effort, and temporal organization of the writing process. In this study, three groups completed the Triple Task protocol under different time intervals: a high-rate group (auditory probe every 15 s, varying between 10 and 20 s), average-rate group (auditory probe every 30 s, varying between 15 and 45 s), and low-rate group (auditory probe every 45 s, varying between 30 and 60 s). The results showed that there were no differences either in the temporal organization of the writing processes or in text production as function of time interval. However, secondary RTs (cognitive effort) were affected by this variation, with longer RT´s both in the low- and high-rate groups. Subsequently, García and Rodríguez (2007) found in their study that the longer the time interval, the better the performance in writing composition. Results from both studies suggest that high-rate patterns may be more intrusive than average- and low-rate patterns in terms of cognitive effort and even performance. These studies have focused on writing composition. However, the number of phases and sub-processes analyzed, and the time devoted to them, may vary considerably depending on the cognitive task under investigation. In addition, excessively long time intervals may result in a loss of information, as certain processes may occur between probes and could not be registered. In order to reach a balance between informative capacity and the degree of interference caused by the assessment procedure, a time interval of 40-45 seconds was established in the present study.

As data were collectively gathered, headphones were given to the students to prevent them disturbing each other. Data were collected and stored in a Moodle platform for later analyses. Students accessed this platform through an individual username and password in order to guarantee their anonymity. Mathematical problems were presented by computer screen and also on paper. Students could use the paper to write whatever they needed with the condition that they had to write their answer on the paper when they finished each problem. Once students reported having finished the problem by pressing the finish button, a new box appeared on the computer screen displaying the following question: “Do you think that you have solved the problem successfully?” Responses were coded as yes (1) or no (0). Students´ achievement in the mathematical problems was established in terms of success (1) or failure (0) according to their written answers.

Once this information was coded, students were assigned to groups as function of the correspondence between post-performance judgment and actual performance in each problem. Students who judged their performance accurately were assigned to Group 1 (accurate), while those who failed in their post-performance judgments were assigned to Group 2 (inaccurate).The small number of students who reported that they had failed to solve the problem when they actually did solve the problem successfully (under-confident students) was about 1-2% in both problems, which made it implausible to consider four groups for further analyses. Thus, only the correspondence between post-performance judgments and actual performance was considered. Group composition varied in both problems, and separate analyses were conducted for each one.

Process variables were based on frequency counts. In order to minimize the effect of students' differences in the total number of reported categories, the frequency of each sub-process or category was established by dividing the election frequency of that category by the total number of elections across categories. In addition, and given the high variability shown by students in their process profiles, with large differences in frequencies across students and categories (see means and standard deviations in Table 5), frequency counts were then transformed into percentages by multiplying the quotient by 100 in order to simplify data analysis and interpretation. This high variability also made it necessary to conduct non-parametric analyses, as described below.

2.4. Data analysis

To address the proposed goals, the data were analyzed in five steps:

Fist, calibration, defined as the correspondence between post-performance judgments and actual performance, was calculated by using Phi coefficient (ρ_{ϕ}) given the dichotomous nature of these variables (Adeyemi 2011, Schraw et al. 2012). Although more dichotomous measures of calibration are available, showing similar specificity and sensibility, Phi coefficient was used as it is provided by SPSS statistical package and easily interpreted in terms of association or correspondence.

Second, also considering the dichotomous nature of the variables *McNemar's* χ^2 statistic was calculated in order to analyze the stability of judgments and real performance across problems, This statistic is based on a 2x2 cross classification of paired responses to a dichotomous item. Data from the concordant cells of the cross-classification table provide an indicator of stability, whereas χ^2 focuses on the change (i.e., non-concordant cells), and establishes whether this change is statistically significant (Adedokun and Burgess 2012). A *p* value $<.05$ indicates the existence of statistically significant differences between the responses given at two different moments (e.g., performance judgments in Problem 1 and 2). As it informs about two different aspects, it implies that high stability can be found even when the change is significant.

Third, prior to analyzing differences in the process between the groups with different accuracy, dependent variable distribution was examined, and students' metacognitive process during the mathematical problems was described. Dependent variables included the three SRL phases (planning, execution and evaluation), and the eight TTPM categories or sub-processes

(reading, drawing or summarizing, recalling similar problems, thinking about solutions, calculating, writing a response, reviewing and correcting mistakes)

Fourth, as variables distribution did not meet normality conditions (see Table 4); non-parametric analyses were conducted to examine group differences. Specifically, the Mann-Whitney U-test was applied, using Cliff's delta (δ) to measure effect size (Macbeth et al. 2011). The non-parametric nature of Cliff's delta reduces the influence of characteristics such as distribution shape, dispersion differences, and extreme values. It provides a measure of dominance, or the degree of overlapping between two distributions of scores. The value of this statistic ranges from –1 (if scores in Group 2 are larger than scores in Group 1) to +1 (if scores in Group 2 are smaller than scores in Group 1), and takes on the value of zero if the two distributions are similar (i.e., absence of significant differences between groups in the measured variables). Cohen (1988, in Romano et al. 2006) established a bridge between Cohen's *d* and Cliff´s statistic. In this sense, a δ value of .147 will have an effect size of d = .20 (small effect), a δ value of .330 will correspond to an effect size of d = .50 (medium effect), and a δ of .474 will have an effect size of d = .80 (large effect).

Finally, the existence of different calibration accuracy patterns and possible differences in the metacognitive process between students with different mathematics achievement were examined. Phi coefficient ($\rho\phi$) was calculated for each group of students (low, medium and high achievers/fifth and sixth grade students). Differences among groups with different levels of achievement in the metacognitive process were examined with the Kruskal-Wallis statistic. As the non-parametric nature of these statistics did not allow us to carry out post-hoc analyses, differences between pairs of groups were analyzed with the Mann-Whitney U-test. This statistic was also used to analyze differences between groups with different grade level. Cliff's delta (δ) was used as a measure of effect size.

SPSS v.19 (Arbuckle 2010) was used to carry out the statistical analyses. However, as this program does not provide any non-parametric alternative to obtain an estimation of effect size, Cliff's Delta Calculator (CDC: Macbeth et al. 2011) was used to calculate *Cliff's* δ. A p-value $\leq .05$ was established as criterion of statistical significance.

3. RESULTS

3.1. Correspondence between Post-performance Judgments and Actual Performance

Table 3 shows the frequency distribution of post-performance judgments and actual performance, as well as the Phi correlation coefficients and their statistical significance in the two mathematical problems.

In Problems 1 and 2, a total of 221 (42.2%) and 278 (53.1%) students, respectively, were accurate in their post-performance judgments, whereas 303 (57.8%) and 246 (46.9%), respectively, were not. The correlation between the two variables (judgments and actual performance) was .207 ($p < .001$) in Problem 1 and .272 ($p < .001$) in Problem 2. Thus, a low and directly proportional relationship between post-performance judgments and actual performance was found. Taking this relationship into consideration, there is a strong mismatch between students' post-performance judgments and their actual performance.

In this sense, students´ post-performance judgments were characterized by a tendency towards over-confidence. In this regard, 298 students (56.871%) in Problem 1 and 232 (44.274%) in Problem 2 reported that they had successfully solved the problem when, in fact, they gave an incorrect response. On the other hand, the number of students who erroneously reported that they had failed to solve the problem when they gave a correct solution (i.e., under-confident) accounted for less than 1% (n = 5) and 2.671% (n = 14) of the sample, respectively. At this point, it is important to note the students' low rates of success in solving both mathematical problems, with only 156 students (29.771%) giving a correct answer in Problem 1, and 185 (35.305%) in Problem 2. These data contrast with the high perception of success shown by students, with 449 (85.687%) and 403 (76.908%) students reporting having solved the problem successfully in Problems 1 and 2, respectively.

TABLE 3

Table of Cross-Classified Data Frequency according to Correspondence between Post-Performance Judgments and Actual Performance, and Phi Correlation (ρ_{ϕ}). Problems 1 and 2

Problem 1						
		Post-performance judgments			Phi Correlation	
		Failure	Success	Total	ρ_{ϕ}	*p*
Actual performance	Failure	70	298	368	.207	< .001
	Success	5	151	156		
	Total	75	449	524		
Problem 2						
		Post-performance judgments			Phi Correlation	
		Failure	Success	Total	ρ_{ϕ}	*p*
Actual performance	Failure	107	232	339	.272	< .001
	Success	14	171	185		
	Total	121	403	524		

3.2. Stability of Post-performance Judgments and Actual Performance

Data about students' stability in post-performance judgments and actual performance are provided in Table 4. Regarding post-performance judgments, 354 students reported being successful on Problems 1 and 2, while 26 students reported being unsuccessful on both problems. This means that a total of 380 students (72.5% of the sample) were stable in their post-performance judgments across problems. Thus, only 27.5% of the students showed a change in their judgments. As *McNemar's* χ^2 statistic indicated, this change was statistically significant (p < .001). In addition, data from actual performance seem to follow a similar pattern. In this sense, 252 students were unsuccessful on Problems 1 and 2, whereas 69 were successful on them. This indicates that 321 students (61.2% of the sample) were stable in their actual performance, whereas 38.8% were not. This change in actual performance across problems was still statistically significant (p = .039).

TABLE 4

Table of Cross-Classified Data Frequency according to Post-performance Judgments and Actual Performance Stability, and *McNemar's* χ^2 statistic. Problems 1 and 2

Post-performance judgments						
		Problem 1			McNemar's χ^2	
		Failure	Success	Total	χ^2	*p*
	Failure	26	49	75	14.063	< .001
Problem 2	Success	95	354	449		
	Total	121	403	524		
Actual performance						
		Problem 1			McNemar's χ^2	
		Failure	Success	Total	χ^2	*p*
	Failure	252	116	368	3.862	.039
Problem 2	Success	87	69	156		
	Total	339	185	524		

3.3. Differences in the metacognitive process shown by students with different calibration accuracy

Descriptive statistics for each TTPM sub-process and phases are shown in Table 5. High standard deviations indicated high variability within subjects in their problem-solving process. Kurtosis and skewness values confirmed the unsuitability of using parametric analyses. The presence of high variability within subjects was also confirmed by important differences in the total time that students spent on both problems. Specifically, students spent a mean of 376.794 seconds (*SD* = 151.491) on Problem 1 and 339.084 seconds (*SD* = 125.626) on Problem 2.

This table also provides information about students' metacognitive process before differences in calibration accuracy were considered. As can be observed, profiles are quite similar in both mathematical problems, with students reporting spending most time on calculation in comparison to the rest of activities. As expected, the amount of time they spent on calculation was significantly higher than the time they spent writing an answer. Students also reported spending a large amount of time thinking about solutions, in comparison to other sub-processes such as recalling similar problems (i.e., previous knowledge and experiences) or drawing/summarizing (organizing the information). Finally, reviewing and correcting sub-processes were the least frequently reported by students in both problems.

TABLE 5

Descriptive Statistic for Dependent Variables (TTPM variables).

Problems 1 and 2

Process categories and phases	Problem 1			Problem 2		
	M(*SD*)	*Kurtosis*	*Skewness*	*M*(*SD*)	*Kurtosis*	*Skewness*
Reading	16.438 (15.109)	5.871	1.952	13.454 (14.57)	8.040	2.232
Drawing or summarizing	11.984 (17.352)	2.507	1.663	13.383 (21.138)	3.043	1.868
Recalling similar problems	3.022 (6.697)	8.160	2.693	2.471 (6.510)	16.305	3.553
Thinking about a solution	19.175 (17.181)	1.250	1.111	18.614 (19.467)	2.110	1.439
Calculating	27.616 (22.832)	-.314	.646	30.973 (25.298)	-.847	.438
Writing	11.746 (12.066)	1.942	1.280	13.767 (14.270)	2.435	1.375
Reviewing	6.299 (9.230)	2.319	1.566	4.229 (7.683)	4.877	2.052
Correcting mistakes	3.763 (7.955)	8.431	2.658	3.251 (7.888)	17.255	3.498
Planning	50.637 (23.950)	-.632	.111	47.872 (27.247)	-.860	.327
Execution	39.351 (23.108)	-.449	.291	44.729 (26.452)	-.887	-.005
Evaluation	10.063 (13.121)	2.590	1.504	7.473 (11.936)	6.093	2.141

Note. *M* = Mean of the group in each process variable. It refers to the mean frequency of election of each sub-process and phase divided by the total number of elections and expressed in percentages

Regarding differences between groups with different calibration accuracy, Table 6 shows means and standard deviations in the process variables.

In problem 1, statistically significant differences were found in the sub-processes of drawing or summarizing (U = 29143.500, p = .006, *Cliff´s* δ = .129), writing (U = 29760.000, p = .026, *Cliff´s* δ = -.111) and reviewing (U = 29301.000, p = .006, *Cliff´s* δ = -.125), as well as in the phases of planning (U = 29807.500, p = .032, *Cliff´s* δ = .110) and evaluation (U = 28340.500, p = .001, *Cliff´s* δ = -.153). Means showed that students who were accurate in their port-performance judgments reported drawing or summarizing (and also planning) more frequently than their inaccurate peers. Accurate students also reported having spent less time on writing, reviewing, and on the evaluation phase, than inaccurate ones.

A similar pattern of results was observed in Problem 2. Specifically, statistically significant differences were found in the sub-processes of drawing or summarizing (U = 29101.000, p = .001, *Cliff´s* δ = .149) and writing (U = 28067.000, p = .001, *Cliff´s* δ = -.179), and in the phases of planning (U = 29453.000, p = .006, *Cliff´s* δ = .139) and evaluation (U = 31088.500, p = .042, *Cliff´s* δ = -.098). Although differences in this variable were not statistically significant, the accurate group also reported correcting mistakes in a lower proportion than students in the inaccurate group.

TABLE 6

Means *(M)*, Standard Deviations *(SD)* of the groups with different Calibration Accuracy in TTPM variables. Problems 1 and 2

Process categories and phases	Problem 1		Problem 2	
	Accurate N = 221	Inaccurate N = 303	Accurate N = 278	Inaccurate N = 246
	M(*SD*)	*M*(*SD*)	*M*(*SD*)	*M*(*SD*)
Reading	15.895 (15.354)	16.835 (14.941)	13.554 (15.376)	13.341 (13.653)
Drawing or summarizing	14.257 (18.276)	10.326 (16.480)	16.082 (22.906)	10.333 (18.521)
Recalling similar problems	3.601 (7.604)	2.600 (5.927)	2.399 (5.627)	2.552 (7.394)
Thinking about a solution	19.778 (18.213)	18.736 (16.405)	18.960 (19.642)	18.223 (19.300)
Calculating	27.977 (23.054)	27.353 (22.703)	31.370 (25.350)	30.524 (25.284)
Writing	10.479 (11.638)	12.670 (12.306)	11.352 (12.170)	16.495 (15.910)
Reviewing	4.972 (8.112)	7.267 (9.867)	3.712 (7.223)	4.813 (8.147)
Correcting mistakes	2.972 (6.329)	4.339 (8.924)	2.762 (7.055)	3.804 (8.717)
Planning	53.629 (23.955)	48.455 (23.749)	50.920 (27.872)	44.426 (26.153)
Execution	38.448 (23.357)	40.009 (22.941)	42.694 (26.842)	47.028 (25.866)
Evaluation	7.945 (11.547)	11.607 (13.978)	6.467 (10.991)	8.609 (12.850)

Note. *M* = Mean of the group in each process variable refers to the mean frequency of election of each sub-process and phase divided by the total number of elections and expressed in percentages

3.4. Differences in calibration and the metacognitive processes as function of achievement level

First, regarding the emergence of different patterns of calibration as function of achievement level (low-medium-high), Table 7 shows that calibration accuracy (ρ_ϕ) improved progressively as students' achievement level increased. This change was more evident in Problem 2, where ρ_ϕ reached values close to .50 in the group with high achievement, which supposes that almost 50% of the sample in the high achieving group would have made accurate judgments about their performance. This improvement in calibration accuracy is reflected in a decrease in over-confidence rates. Specifically, the number of students making over-confidence judgments in Problem 1 was 124 (61.083%) in the group of low achievers, and 64 (45.714%) in the group of high achievers. In Problem 2, the number of over-confident students ranged from 114 (56.157%) to 37 (26.428%) in the groups of low and high achievers, respectively.

TABLE 7

Patterns of Calibration Accuracy as function of Achievement Level in Mathematics. Problems 1 and 2

			Post-performance judgments			Phi Correlation	
			Failure	Success	Total	ρ_ϕ	*p*
Problem 1							
Low achievers	Actual performance	Failure	28	124	152	.151	.031
		Success	3	48	51		
		Total	31	172	203		
Medium achievers	Actual performance	Failure	21	110	131	.162	.030
		Success	2	48	50		
		Total	23	158	181		
High achievers	Actual performance	Failure	21	64	85	.338	<.001
		Success	0	55	55		
		Total	21	119	140		
Problem 2							
Low achievers	Actual performance	Failure	45	114	159	.162	.021
		Success	5	39	44		
		Total	50	153	203		
Medium achievers	Actual performance	Failure	32	81	113	.232	.002
		Success	6	62	68		
		Total	38	143	181		
High achievers	Actual performance	Failure	30	37	67	.479	<.001
		Success	3	70	73		
		Total	33	107	140		

Second, with respect to the possible differences in the process, Kruskal-Wallis statistic revealed the existence of statistically significant differences in the sub-processes of reading [*K-W* (χ^2) = 9.332, p = .009 in Problem 1; and *K-W* (χ^2) = 16.243, p <.001 in Problem 2], and drawing/summarizing [*K-W* (χ^2) = 6.904, p = .032 in Problem 1; and *K-W* (χ^2) = 15.636, p <.001 in Problem 2] as function of achievement level.

Means in Table 8 showed that students with higher levels of achievement in mathematics reported having read less frequently, but using representation strategies more frequently than their low-achieving peers (see Table 8). Mann-Whitney's U-test showed that the groups with low and medium achievement only differed in the sub-process of drawing/summarizing in Problem 1 (U = 16332.500, p = .037, *Cliff´s* δ = -.111), whereas significant differences in this sub-process were found between the groups with medium and high achievement in both problems (Problem 1: U = 11007.000, p = .032, *Cliff´s* δ = -.131; Problem 2: U = 10363.500, p = .003, *Cliff´s* δ = -.182). Regarding the groups with low and high achievement, differences in drawing/summarizing were found in both problems (Problem 1: U = 10938.500, p < .001, *Cliff´s* δ = -.230; Problem 2: U = 11128.000, p < .001, *Cliff´s* δ = -.216), as well as in reading (Problem 1: U = 11537.500, p = .003, *Cliff´s* δ = .188; Problem 2: U = 11961.500, p = .012, *Cliff´s* δ = .158). Effect sizes, although higher than in previous analyses, were low.

TABLE 8

Means *(M)*, Standard Deviations *(SD)* of the Groups with different Achievement Level in TTPM Variables.

Problems 1 and 2

	Problem1			Problem 2		
Process categories and phases	Low achievers	Medium achievers	High achievers	Low achievers	Medium achievers	High achievers
	M(*SD*)	*M*(*SD*)	*M*(*SD*)	*M*(*SD*)	*M*(*SD*))	*M*(*SD*)
Reading	19.492 (18.185)	15.243 (12.797)	13.557 (11.935)	16.049 (17.326)	12.314 (12.546)	11.164 (11.953)
Drawing or summarizing	8.536 (13.944)	11.966 (17.176)	17.007 (20.630)	11.640 (21.413)	11.668 (19.196)	18.128 (22.513)
Recalling similar problems	3.354 (7.285)	2.773 (6.541)	2.864 (5.999)	2.773 (7.256)	2.690 (6.720)	1.750 (4.874)
Thinking about a solution	20.182 (18.765)	18.060 (15.607)	19.157 (16.747)	19.335 (20.625)	18.729 (19.362)	17.421 (17.890)
Calculating	25.753 (23.003)	29.165 (22.451)	28.314 (23.044)	28.310 (25.044)	33.397 (25.542)	31.700 (25.163)
Writing	12.083 (12.241)	13.044 (13.661)	9.578 (9.014)	13.635 (14.618)	14.198 (15.929)	13.400 (11.256)
Reviewing	6.315 (8.827)	6.243 (9.368)	6.350 (9.675)	4.453 (8.131)	3.966 (7.363)	4.242 (7.454)
Correcting mistakes	4.305 (8.587)	3.591 (7.673)	3.2000 (7.344)	3.891 (9.785)	3.198 (6.469)	2.392 (6.281)
Planning	51.605 (24.799)	48.027 (22.930)	52.607 (23.864)	49.788 (27.295)	45.331 (27.229)	48.378 (27.137)
Execution	37.842 (23.703)	42.198 (22.288)	37.857 (23.092)	41.965 (26.025)	47.580 (26.694)	45.050 (26.529)
Evaluation	10.615 (13.851)	9.828 (12.910)	9.564 (12.347)	8.330 (13.659)	7.160 (10.517)	6.635 (10.946)

Note. *M* = Mean of the group in each process variable refers to the mean frequency of election of each sub-process and phase divided by the total number of elections and expressed in percentages. Low achievers (*N* = 203); Medium achievers (*N* = 181); High achievers (*N* = 140)

3.5. Differences in calibration and metacognitive processes as function of grade level

Table 9 shows patterns of calibration accuracy as function of grade level Problems 1 and 2. As can be observed, calibration accuracy was low, especially in the case of sixth-grade students in Problem 1 (ρ_{ϕ} = .114). In contrast to what happened with mathematics achievement, calibration accuracy does not seem to present a distinguishable pattern of change as function of grade level. As in previous analyses, calibration patterns were characterized by a marked tendency towards over-confidence. Specifically, 123 (55.909%) and 104 (47.272%) fifth-grade students were over-confident in Problems 1 and 2, respectively. Sixth-grade students were over-confident in 175 (57.565%) and 128 (42.105%) of the cases, respectively.

TABLE 9

Patterns of Calibration Accuracy as function of Grade Level.

Problems 1 and 2

			Problem 1				
			Post-performance judgments			Phi Correlation	
			Failure	Success	Total	ρ_{ϕ}	*p*
Fifth graders	Actual performance	Failure	47	123	170	.283	<.001
		Success	0	50	50		
		Total	47	173	220		
Sixth graders	Actual performance	Failure	23	175	198	.114	.047
		Success	5	101	106		
		Total	28	276	304		
			Problem 2				
			Post-performance judgments			Phi Correlation	
			Failure	Success	Total	ρ_{ϕ}	*p*
Fifth graders	Actual performance	Failure	57	104	161	.265	<.001
		Success	5	54	59		
		Total	62	158	220		
Sixth graders	Actual performance	Failure	50	128	178	.261	<.001
		Success	9	117	126		
		Total	59	245	304		

Regarding group differences in TTPM sub-processes and phases, means in the TTPM variables (Table 10) indicated that both groups tended to show a similar metacognitive process. This was confirmed by the results of Mann-Whitney U statistic. Specifically, statistically significant group differences were only found in Problem 1, in the sub-processes of calculating (U = 29533.500, p = .022, *Cliff´s* δ = -.117) and reviewing (U = 29414.500, p = .008, *Cliff´s* δ = -.120). In this way, sixth-grade students reported using calculating and reviewing sub-processes more frequently than their peers. However, similar results were not found in Problem 2. This result could be related to the finding that the sixth-grade students were considerably less accurate than fifth graders in Problem 1, reflecting the previously observed differences between accurate and inaccurate groups.

TABLE 10

Means (*M*), Standard Deviations (*SD*) and Differences between Fifth- and Sixth-grade Students in TTPM Variables.

Problems 1 and 2

	Problem 1		Problem 2	
Process categories and phases	Fifth graders	Sixth graders	Fifth graders	Sixth graders
	M(*SD*)	*M*(*SD*)	*M*(*SD*)	*M*(*SD*)
Reading	16.886 (14.961)	16.115 (15.232)	13.040 (13.960)	13.753 (15.026)
Drawing or summarizing	11.904 (18.368)	12.042 (16.609)	13.313 (22.482)	13.434 (20.148)
Recalling similar problems	3.668 (7.268)	2.555 (6.222)	2.700 (7.195)	2.305 (5.972)
Thinking about a solution	20.281 (19.040)	18.375 (15.684)	19.559 (19.222)	17.930 (19.646)
Calculating	25.677 (24.166)	29.019 (21.748)	29.568 (26.278)	31.990 (24.558)
Writing	11.600 (12.922)	11.852 (11.427)	13.727 (14.643)	13.796 (14.018)
Reviewing	5.231 (8.822)	7.072 (9.453)	4.509 (8.519)	4.026 (7.025)
Correcting mistakes	4.781 (9.766)	3.026 (6.246)	3.622 (8.812)	2.983 (7.150)
Planning	52.781 (25.611)	49.085 (22.589)	48.636 (28.196)	47.319 (26.572)
Execution	37.268 (24.562)	40.858 (21.913)	43.286 (27.024)	45.773 (26.025)
Evaluation	10.013 (14.509)	10.098 (12.043)	8.127 (13.447)	7.000 (10.709)

Note. *M* = Mean of the group in each process variable refers to the mean frequency of election of each sub-process and phase divided by the total number of elections and expressed in percentages. Fifth graders (*N* = 220); Sixth graders (*N* = 304)

4. Discussion and conclusions

The present study was aimed at examining patterns of post-performance calibration and their relationship with the metacognitive process shown by a broad sample of Elementary School students while solving two mathematical word problems. In order to evaluate the process, the *Triple Task Procedure in Mathematics* (TTPM: García and González-Pienda 2012; García et al. in press) was administered to 524 fifth- and sixth-grade students in a regular mathematics class. Once TTPM was completed, the students were asked to judge the results of their performance in terms of success or failure. These judgments were then compared with their actual performance, also expressed in dichotomous terms. The possible influence of students' mathematics achievement and grade level both on calibration accuracy patterns and the metacognitive process was analyzed. Separate analyses were conducted for each problem. Results are discussed according to the five initial predictions:

4.1. Correspondence between Post-performance Judgments and Actual Performance

As hypothesized, students were poorly calibrated in the present study, with a strong tendency towards over-confidence. Nearly 60% and 45% of the sample in Problems 1 and 2, respectively, reported having successfully solved the mathematical problems when they actually gave an incorrect answer. This finding is coherent with previous studies in mathematics and problem solving, conducted in Elementary School and higher educational stages (Bol et al. 2005, 2010; Hacker et al 2008b; Özsoy 2012).

4.2. Stability of Post-performance Judgments and Actual Performance

Previous studies suggest that performance judgments (and biases) may be stable over time, and more stable than actual performance (Bol et al. 2012; Dinsmore and Parkinson 2013; Hacker et al. 2008b; Stolp and Zabrucky 2009). In this sense, results from the present study tentatively support this statement, as well as the second hypothesis of the study. Specifically, post-performance judgments and actual performance were both quite stable across problems, with about 70% of students showing stability in their judgments, and 60% in their performance. However, as will be discussed later in the present study, the use of only two mathematics problems to estimate stability makes it difficult to generalize these findings.

4.3. Differences in the metacognitive process shown by students with different calibration accuracy

Descriptive analyses of the metacognitive process during TTPM revealed that it was characterized by students' spending a great amount of time planning. However, their planning strategies seemed to be ineffective, as indicated the low success rates performing the problems. In this sense, they spent a greater amount of time in this phase reading the statement of the problems and thinking about solutions, in comparison to the time they spent organizing information (drawing/summarizing), or recalling similar problems. Students also showed a tendency towards performing calculations as a method to solve the problems. This is, in fact, the most frequently reported sub-process. This pattern of results would suggest a use of ineffective problem-solving strategies by student, related to the presence of little metacognitive skills. This is coherent with some previous studies showing that students at these educational stages tend not to be self-regulated enough in mathematical problem-solving situations, which has demonstrated to lead to poor performance and inaccurate judgments (Cleary and Chen 2009; Desoete and Roeyers 2006; Kramarski and Gutman 2006).

However, differences between accurate and inaccurate students in the metacognitive process during TTPM were found, indicating that inaccurate students spent significantly less time planning (specially drawing and summarizing), but more time writing and reviewing than their accurate peers. They also reported correcting (and presumably making) more mistakes than the accurate students, although this variable was not statistically significant. Similar results were found in both problems. Although effect sizes were low in general terms, these findings initially suggest a relation between the use of more effective planning strategies (mainly regarding representation of the information) and making more accurate post-performance judgments. In fact, there seems to be an inverse relationship between the use of better planning strategies and the use of what can be defined as "trial-and-error" mechanisms, characterized as a tendency towards tackling a problem with various methods until a solution is found, and have been reported to be frequent during problem-solving tasks (Callander 2011; Clements et al. 2008; Dijk et al. 2003).

Evidence from students' process while solving the mathematical problems could be obtained by analyzing performance artifacts (e.g., drawings, notes or editing from students' drafts on paper). An example of these artifacts is provided below (Figure 2), which corresponds to the performance of two different children in Problem 1 (two 11-year-old boys, the first boy belonging to the accurate group and the second one to the inaccurate group). This figure represents what could be an example both of organization of the information and trial-and-error mechanisms. As can be observed below, the student who judged his performance accurately (left side of the image) used different strategies to organize the information, establishing clear relationships between specific data and facts. He also seems to have tried only one approximation to solve the problem, as a lack of mistakes or corrections indicate. Finally, he successfully solved the problem, giving a unique and correct answer. Regarding the student who was inaccurate in his performance judgment (right side of the image), it seems that the problem-solving process was less smooth than in the first case, with some signs of correcting and editing. A lack of information-organization strategies can be observed, and data and relationships seem to be rather chaotic. Although the student gave a unique answer, it was incorrect.

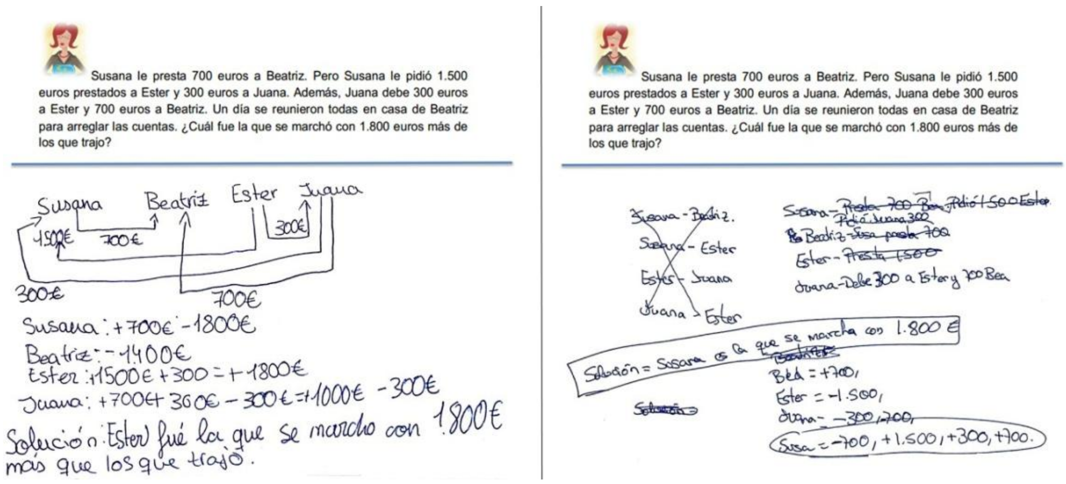


Figure 2. Example of students' artifacts during TTPM (Problem 1).

Two 11-year-old students: accurate (left) and inaccurate (right)

This pattern of results could indicate the presence of some degree of self-regulatory and metacognitive mechanisms in calibrated students. As Jacobse and Harskamp (2012) pointed out, an important cognitive action in mathematical problem-solving is building a representation of the problem situation, and few students in Elementary School use this strategy spontaneously. The use of schematic visualizations allows students to establish those much needed relations between variables that help to solve the problem successfully. For these authors, making these kinds of representations could be interpreted as an expression of sophisticated metacognitive regulation mechanisms in mathematical problem solving, which provide insight in the episodes of analyzing and exploring a problem (i.e., metacognitive regulation in the first episodes of the problem-solving process). Moreover, schematic representations have been shown to have good predictive validity for students' problem-solving performance in different studies (Fagnant and Vlassis 2013).

4.4. Differences in calibration and the metacognitive process as function of achievement level

An important issue that was addressed in the present study was the possibility that different patterns in calibration accuracy, as well as differences in the metacognitive process during TTPM, emerge as function of mathematics achievement level. In this sense, the low effect sizes found in the previous analyses may response partially to the influence of this variable. Differences in calibration accuracy as function of achievement level were found. Specifically, the higher the achievement level, the more precise the post-performance judgments, and the lower the tendency towards over-confidence shown by students. The same pattern of results was obtained in both problems. These findings are consistent with previous studies showing that high achieving students are commonly more accurate in their judgments of performance (and somewhat under-confident) than their low-achieving peers (Bol et al. 2005, 2010; Hacker et al. 2008b; Özsoy 2012). However, no conclusions about under-confidence were obtained in the present study, as the proportion of under-confident students was practically negligible (about 1-2%). This could be related to the fact that most students in the current sample were low achievers, as problem-solving scores evidenced. Another important finding in this case is that differences in the metacognitive process during TTPM also surfaced as function of achievement level, suggesting that previous differences in drawing/summarizing (and planning) may be at least partially explained by this variable. This is not surprising, as high achievers have been referred to as more highly calibrated and self-regulated than low achievers (Bol et al. 2005; Dunlosky and Rawson 2012; Nietfeld et al. 2005, 2006).

4.5. Differences in calibration and the metacognitive process as function of grade level

With regard to grade level, no distinguishable patterns of calibration accuracy were obtained as function of this variable. In general terms, students in fifth and sixth grades showed low calibration accuracy. However, sixth-grade students' calibration accuracy dropped considerably in Problem 1. This could explain why some significant differences between fifth and sixth graders were found, but only in this problem. In this sense, the differences found in reviewing would respond to previous differences between accurate and inaccurate groups. Thus, it seems that grade level did not have an influence on calibration accuracy or metacognitive processes in this sample. This finding could be related to the fact that both school grades are very close in time and constitute a unique educational cycle in the Spanish Educational System, making it difficult to observe developmental or cognitive differences between them. However, as there is a good deal of evidence suggesting that low calibration accuracy and over-confidence judgments seem to be present across different educational stages, an absence of differences can also be expected.

Implications

The main implication of the present study is related to the impact that inaccurate judgments of performance may have on learning processes and outcomes, and how these self-evaluation biases may be corrected. In this sense, as students seem to show a strong tendency towards over-confidence (also corroborated in the present study), most research has focused on studying the extent to which this positive bias is adaptive or maladaptive (Duperyrat et al. 2011). Results are mixed. On the one hand, based on Bandura´s social cognitive theory of human functioning (Bandura 1986), it has been argued that positive self-evaluation bias might be adaptive because it may engage motivation, persistence in the face of failures, and protect against negative emotions such as anxiety (Stolp and Zabrucky 2009). On the other hand, authors such as Bouffard and Narcis (2011), Narciss et al. (2011) and Schunk (2008) emphasize the negative effects of unrealistic judgments, whereas others argue that positive illusions might be adaptive or maladaptive depending on variables such as the domain or the temporariness of the related outcomes (Bouffard et al. 2011; Gonida and Leondari 2011; Gramzow et al. 2003).

This debate has prompted important efforts to improve calibration accuracy, many of them from the perspective of metacognition and SRL. Some of these studies have shown that gains are difficult to reach or are not durable (Bol and Hacker 2001; Bol et al. 2005; Nietfeld et al. 2005), whereas others have reported some improvement (DiGiacomo 2014; Hacker et al. 2008b; Nietfeld et al. 2006; Zimmerman et al. 2011). These studies agree that more comprehensive and adapted interventions must be designed and implemented. For this purpose, and as self-evaluation biases seem to be present at all academic stages, the first step would consist of analyzing the factors involved in making performance judgments at early educational stages. Results from the present study initially suggest the potential usefulness of process-based measures, specifically the Triple Task Procedure in mathematics (García and González-Pienda 2012; García et al. in press), in the study of the metacognitive mechanisms involved in making post-performance judgments in mathematical problem solving. This procedure also reflected the influence that the mathematics achievement level may have on calibration accuracy and the metacognitive process itself. This aspect is important because the relationship between these components may have implications for intervention. The study of Hacker et al. (2000) illustrates this aspect. These authors analyzed the effectiveness of an intervention based on feedback, practice tests, and course instruction to improve calibration accuracy (predictions and postdictions) and academic performance in a sample of undergraduate students. The results revealed that calibration accuracy improved, but only for higher achieving students. Similar results were obtained by Nietfeld et al. (2006) and Hacker et al. (2008b). These studies suggest that intervention strategies must be adapted to students' features, and achievement level is an important factor moderating the effects of the intervention.

Limitations

Finally, some limitations in the present study must be acknowledged: first, the low effect sizes found call for some degree of caution concerning the scope of our findings. This result could be related to the high inter-subject variability observed in the metacognitive process, where it would be necessary to establish more homogeneous groups. These low effect sizes may also indicate that other variables may be accounting for differences in the metacognitive process. These aspects should be addressed in further studies; second, the use of only two mathematical problems must also be considered. As explained before, the features of the assessment method used (TTPM) justified this decision. However, including more mathematical problems would be especially useful to obtain a clearer measure of judgments and performance stability. Similarly, and given the mentioned time constraints, the possibility of administering the problems across days must be considered. Taking into consideration that the dichotomous nature both of post-performance judgments and actual performance may have supposed a limitation in the present study, more continuous measures of calibration could be obtained by increasing the number of mathematical problems. In this sense, retaining the classical measure of cognitive effort (RT´s in the Triple Tasks procedure) in future studies would help to establish the appropriate number of problems by examining the extent to which they are challenging for students; third, another possible limitation in the present study is that the TTPM may result intrusive for students, leading to an effect of "reactivity" (Bowles and

Leow 2005), referred to as the possibility that the type of measure employed triggers changes in learners' cognitive and metacognitive processes while performing the task. This effect has been substantially studied regarding other on-line measures, mainly Think-aloud protocols (Bannert and Mengelkamp 2008; Ericsson and Simon 1996; Fox et al. 2011; Veenman et al. 1993). These authors concluded that, although the use of this sort of measure may be related to an increase in the time to complete the task, simply instructing participants to verbalize their thoughts during a task does not alter the sequence of the cognitive processes or task performance. In the same line, Kellog (1987, in Olive et al. 2001), using Triple Task procedures, reached similar results (e.i., this author did not find differences in written fluency or quality between writers engaged in directed retrospection and 30 control writers who made a written composition as a single task), which suggested that directed retrospection in Triple Task was not reactive. However, it is necessary to point out that the use of different tasks (i.e. mathematics problems in this study) must also be considered. In this sense, while there is wide tradition in the study of these processes though Think-aloud protocols (see Montague et al. 2011), the flexibility of the Triple Task technique and its possibility to be administered to larger samples motivated its application in the present study. Given its design, it is possible however, that it results more intrusive for students than Think-aloud. In order to reduce its possible effect on students' performance, the TTPM places special emphasis on the training phase so that the categorization task (directed retrospection) becomes more automatic, generating less interference and involving less cognitive effort. Additional research is being conducted to properly examine the possible reactive effect of this technique; fourth, students were not provided with any incentive for taking the study. This is relevant because students' performance was low. Providing incentives in the form of prizes or extra credits may increase motivation, improving performance. In fact, motivation has been demonstrated to be important for students to engage in and successfully complete a task, as well as to evaluate their performance (Sheldrake et al. 2014). This is relevant, as Zimmerman´s SRL model (2000, 2008) includes cognitive, behavioural, emotional and motivational components. Thus, exploring motivation and its relation to calibration accuracy and performance is an aspect to be considered in future studies; finally, although this study was initially aimed at increasing ecological validity, some contextual aspects such as the fact that teachers were not present during the evaluation, or the above-mentioned absence of incentives, might have compromised this attempt. Future studies will be conducted in order to address these issues.

References

Ackerman, P. L. & Wolman, S. D. (2007). Determinants and validity of self-estimates of abilities and self-concept measures. *Journal of Experimental Psychology: Applied, 13,* 57-78.

Adedokun, O. A. & Burgess, W. D. (2012). Analysis of paired dichotomous data: A gentle introduction to the McNemar´s Test in SPSS. *Journal of Multidisciplinary Evaluation, 8*(17), 125-131.

Adeyemi, O. (2011). Measures of association for research in educational planning and administration. *Research Journal of Mathematics and Statistics 3*(3), 82-90.

Alexander, P. A. (2013). Calibration: What is it and why it matters? An introduction to the special issue on calibrating calibration. *Learning & Instruction, 24,* 1-3, doi: 10.1016/j.learninstruc.2012.10.003

Arbuckle, J. L. (2010). *SPSS (Version 19.0) [Computer Program]*. Chicago: SPSS.

Azevedo, R. & Aleven, V. (2013). *International handbook of metacognition and learning technologies*. Springer International Handbooks of Education, 28, New York, NY: Springer. doi: 10.1007/978-1-4419-3_11.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice-Hall.

Bannert, M. & Mengelkamp, C. (2008). Assessment of metacognitive skills by means of instruction to think aloud and reflect when prompted. Does the verbalization method affect learning? *Metacognition and Learning, 3,* 39-58.

Boekaerts, M. (1999). Motivated learning: Studying student situation transactional units. *European Journal of Psychology of Education, 14*(1), 41-55. doi: 10.1007/bf03173110

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. doi: 10.1016/j.learninstruc.2009.03.002

Bol, L. & Hacker, D. J. (2001). A comparison of the effects of practice tests and traditional review on performance and calibration. *The Journal of Experimental Education, 69,* 133-151.

Bol, L., Hacker D. J., O'Shea P. & Allen D. (2005). The influence of overt practice, achievement level, and explanatory style on calibration accuracy and performance. *Journal of Experimental Education, 73,* 269-290. doi: 10.3200/JEXE.73.4.269-290

Bol, L., Hacker, D. J., Walck, C. C. & Nunnery, J. A. (2012). The effects of individual or group guidelines on the calibration accuracy and achievement of high school biology students. *Contemporary Educational Psychology, 37*, 280-287. doi: 10.1016/j.cedpsych.2012.02.004

Bol, L., Riggs R., Hacker D. J. & Nunnery J. (2010). The calibration accuracy of middle school students in math classes. *Journal of Research in Education, 21*, 81-96.

Boston, M. D. & Smith, M. S., (2009). Transforming secondary mathematics teaching: Increasing the cognitive demands of instructional tasks used in teachers' classrooms. *Journal for Research in Mathematics Education, 40*(2), 119-156.

Bouffard, T. & Narciss, S. (2011). (Guest Eds.). Benefits and risks of positive bias in self-evaluation of academic competence. *International Journal of Educational Research, 50*(4), 205-256.

Bouffard, T., Vezeau, C., Roy, M. & Lengelé, A. (2011). Stability of biases in self-evaluation and relations to well-being among elementary school children. *International Journal of Educational Research*, *50,* 221-229. doi: 10.1016/j.ijer.2011.08.003

Bowles, M. A. & Leow, R. P. (2005). Reactivity and type of verbal report in SLA research methodology. *Studies in Second Language Acquisition, 27,* 415-440.

Bransford, J. D. & Stein, B. S. (1993). *The ideal problem solver: A guide for improving thinking, learning and creativity* (2nd Ed.). New York, NY: W. H. Freeman.

Butler, R. (2011). Are positive illusions about academic competence always adaptive, under all circumstances? New results and future directions. *International Journal of Educational Research, 50,* 251-256. doi: 10.1016./j.ijer.2011.08.001

Callander, S. (2011). Searching and learning by trial and error. *American Economic Review, 101*(6), 2277-2308. doi: 10.1257/aer.101.6.2277

Cleary, T. J. (2009). Monitoring trends and accuracy of self-efficacy beliefs during interventions: Advantages and potential applications to school-based settings. *Psychology in the Schools, 46,* 154-171. doi: 10.1002/pits.20360

Cleary, T. J. & Chen, P. (2009). Self-regulation, motivation, and math achievement in middle school: Variations across grade level and math context. *Journal of School Psychology, 47,* 291-314. doi: 10.1016/j.jsp.2009.04.002

Clements, D. H., Sarama, J. & Liu, X. (2008). Development of a measure of early mathematics achievement using the Rasch model: The Research-based Early Maths Assessment. *Educational Psychology*, *28*(4), 457-482. doi: 10.1080/01443410701777272

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd Ed.). New York, NY: Academic Press.

Desoete, A. & Roeyers, H. (2006). Metacognitive macro-evaluations in mathematical problem-solving. *Learning and Instruction, 16,* 12-25.

DiGiacomo, G. (2014). Enhancing self-monitoring and self-reflection through a self-regulatory skills intervention embedded in a middle school mathematics curriculum. *Dissertations and Theses, 2014-Present.* Paper 201. http://works.gc.cuny.edu/etd/201.

Dijk, I. M. A. W., Van Oers, B. & Terwel, J. (2003). Providing or designing? Constructing models in primary maths education. *Learning and Instruction, 13,* 53-72. doi: 10.1016/S0959-4752(01)00037-8

Dinsmore, D., Alexander, P. & Loughlin, S. (2008). Focusing the conceptual lens on metacognition, self-regulation, and self-regulated learning. *Educational Psychology Review, 20,* 391-409.

Dinsmore, D. L. & Parkinson, M. M. (2013). What are confidence judgments made of? Students' explanations for their confidence ratings and what that means for calibration. *Learning and Instruction, 24,* 4-14. doi: 10.1016/j.learninstruc.2012.06.001

Dunlosky, J. & Rawson, K. A. (2012). Over-confidence produces underachievement: Inaccurate self-evaluations undermine students' learning and retention. *Learning and Instruction, 22,* 271-280. doi: 10.1016/j.learninstruc.2011.08.003

Dunlosky, J. & Thiede, K. W. (2013). Four cornerstones of calibration research: Why understanding students' judgments can improve their achievement. *Learning and Instruction, 24,* 58-61. doi: 10.1016/j.learninstruc.2012.05.002

Efklides, A. & Misailidi, P. (2010). *Trends and prospects in metacognition research*. New York, NY: Springer.

Ericsson, K. A. & Simon, H. A. (1996). *Protocol analysis: Verbal reports as data.* (Revised edition). Cambridge, MA: MIT Press.

Fagnant, A. & Vlassis, L. (2013). Schematic representations in arithmetical problem-solving: Analysis of their impact on grade 4 students. *Educational Studies in Mathematics, 84*(1), 149-168.

Finn, B. & Metcalfe, J. (2014). Over-confidence in children's multi-trial judgments of learning. *Learning and Instruction, 32,* 1-9. doi: 10.1016/j.learninstruc.2014.01.001

Fox, M. C., Ericsson, K. A. & Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychological Bulletin, 137,* 316-344. doi: 10.1037/a0021663

García, J. N. & Rodríguez, C. (2007). Influence of the recording interval and a graphic organizer on the writing process/product and on other psychological variables. *Psicothema, 19*(2), 198-205.

García, T., Betts, L., González-Castro, P., González-Pienda, J. A. & Rodríguez, C. (in press). On-line assessment of the process involved in Maths problem-solving in fifth and sixth grade students: self-regulation and achievement. *Revista Latinoamericana de Investigación en Matemática Educativa.*

García, T. & González-Pienda, J. A. (2012). Evaluación del proceso de aprendizaje autorregulado en el área de las matemáticas mediante pizarras digitales. [Assessment of the self-regulated learning process in Mathematics through electronic blackboards]. In J. Dulac & C. Alconada (Coords.), *III Congreso Pizarra Digital. Publicación de comunicaciones* (pp.105-117). Madrid: Pluma y Arroba.

Gick, M. L. (1986). Problem-solving strategies. *Educational Psychologist, 21,* 99-120.

Gonida, E. N. & Leondari, A. (2011). Patterns of motivation among adolescents with biased and accurate self-efficacy beliefs. *International Journal of Educational Research, 50*(4), 209-211. doi: 10.1016/j.ijer.2011.08.002.

Gramzow, R. H., Elliot, A. J., Asher, E. & McGregor, H. A. (2003). Self-evaluation bias and academic performance: Some ways and some reasons why. *Journal of Research in Personality, 37*, 41-61.

Hacker, D. J. & Bol, L. (2004). Metacognitive theory: Considering the social-cognitive influences. In D. M. McInerney & S. Van Etten (Eds.), *Big theories revisited, Vol. 4: Research on sociocultural influences on motivation and learning* (pp. 275-297). Greenwich, CT: Information Age.

Hacker D. J., Bol L. & Bahbahadi, K. (2008b). Explaining calibration accuracy in classroom contexts: The effects of incentives, reflection, and explanatory style. *Metacognition Learning, 3,* 101-121. doi: 10.1007/s11409-008-9021-5

Hacker, D. J., Bol, L., Horgan, D. D. & Rakow, E. A. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology, 92*, 160-170.

Hacker, D. J., Bol, L. & Keener, M. C. (2008a). Metacognition in education: A focus on calibration. In J. Dunlosky & R. Bjork (Eds.), *Handbook of memory and metacognition (*pp. 429-455). Mahwah, NJ: Erlbaum.

Hadwin, A. F. & Webster, E. A. (2013). Calibration in goal setting: Examining the nature of judgments of confidence. *Learning and Instruction, 24*(12), 37-47. doi: 10.1016/j.learninstruc.2012.10.001

Hattie, J. (2013). Calibration and confidence: Where to next? *Learning & Instruction, 24*, 62-66. doi: 10.1016/j.learninstruc.2012.05.009

Jacobse, A. E. & Harskamp, E. G. (2012). Towards efficient measurement of metacognition in mathematical problem-solving. *Metacognition and Learning, 7*(2), 133-149.

Kellog, R. T. (1987). Writing performance: Effects of cognitive strategies. *Written Communication, 4,* 269-298.

Kramarski, B. & Gutman, M. (2006). How can self-regulated learning be supported in mathematical E-learning environments? *Journal of Computer Assisted Learning 22*, 24-33. doi: 10.1111/j.1365-2729.2006.00157.x

Labuhn, A., Zimmerman, B. & Hasselhorn, M. (2010). Enhancing students' self-regulation and mathematics performance: The influence of feedback and self-evaluative standards. *Metacognition and Learning, 5*(2), 173-194. doi: 10.1007/s11409-010-9056-2

Lipko, A. R., Dunlosky, J., Hartwig, M. K., Rawson, K. A., Swan, K. & Cook, D. (2009). Using standards to improve middle school students' accuracy at evaluating the quality of their recall. *Journal of Experimental Psychology: Applied, 15*(4), 307-318. doi: 10.1037/a0017599

Macbeth, G., Razumiejczyk, E. & Ledesma, R. D. (2011). Cliff's Delta Calculator: A non-parametric effect size program for two groups of observations. *Universitas Psychologica, 10*(2), 545-555.

Maki, R. H., Shields M., Wheller, A. E. & Zacchilli, T. L. (2005). Individual differences in absolute and relative meta-comprehension accuracy. *Journal of Educational Psychology, 97,* 723-731. doi: 1037/0022-0663.97.4.723

McCormick, C. B. (2003). Metacognition and learning. In W. B. Reynolds & G. E. Miller (Eds.), *Handbook of psychology: Educational psychology* (pp. 79-102). Hoboken, NJ: Wiley.

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. doi: 10.1177/0731948711421762

Moos, D. C. & Ringdal, A. (2012). Self-Regulated Learning in the classroom: A literature review on the teacher's role. *Education Research International, 2012*, Article ID 423284. doi: 10.1155/2012/423284

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. doi: 10.1016/j.ijer.2011.08.004

Newton, S. J., Horner, R. H., Algozzine, R. F., Todd, A. W. & Algozzine, K. M. (2009). Using a problem-solving model to enhance data-based decision making in schools. In W. Sailor, G. Dunlap, G. Sugai, & R. Horner (Eds.), *Handbook of positive behaviour support. Issues in clinical child psychology* (pp. 551-580). New York, NY: Springer.

Newton, S. J., Horner, R. H., Todd, A. W., Algozzine, R. F. & Algozzine, K. A. (2012). A pilot study of a problem-solving model for team decision making. *Education and Treatment of Children, 35*(1), 25-49. doi: 10.1353/etc.2012.0001

Nietfeld, J. L., Cao, L. & Osborne, J. W. (2005). Metacognitive monitoring accuracy and student performance in the postsecondary classroom. *The Journal of Experimental Education, 74*, 7-28.

Nietfeld, J. L., Cao, L. & Osborne, J. W. (2006). The effect of distributed monitoring exercises and feedback on performance, monitoring accuracy, and self-efficacy. *Metacognition and Learning, 1*, 159-179.

Olive, T., Kellog, R. T. & Piolat, A. (2001). The Triple Task technique for studying the process of writing. In T. Olive., & C. M. Levy (Eds.), *Contemporary tools and techniques for studying writing* (Vol. 10, pp. 89-104). Dordecht: Kluwer Academic Publishers.

Olive, T. & Piolat, A. (2002). Suppressing visual feedback in written composition: Effects on processing demands and coordination of the writing processes. *International Journal of Psychology, 37*(4), 209-218. doi: 10.1080/00207590244000089

Özsoy, G. (2012). Investigation of fifth grade students' mathematical calibration skills. *Educational Sciences: Theory & Practice, 12*(2), 1190-1194.

Panadero, E. & Alonso-Tapia, J. (2014). How do students self-regulate? Review of Zimmerman's cyclical model of self-regulated learning. *Annals of Psychology, 30*(2), 450-462. doi: 10.6018/analesps.30.2.167221

Parkinson, M. M., Dinsmore, D. L. & Alexander, P. A. (2010). *Calibrating calibration: Towards conceptual clarity and agreement in calculation.* Paper presented at the annual Meeting of the American Educational Research Association, Denver, CO.

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 and Reasoning*, *16*(3), 198-220. doi: 10.1080/13546783.2010.509052

Pereis, F., Dignath, C. & Schmitz, B. (2009). Is it possible to improve mathematical achievement by means of self-regulation strategies? Evaluation of an intervention in regular math classes. *European Journal of Psychology of Education, 24*(1), I7-29. doi: 10.1007/BF03173472

Piolat, A., Kellogg, R. T. & Farioli, F. (2001). The triple task technique for studying writing processes: On which task is attention focused? *Current Psychology Letters: Brain, Behavior and Cognition, 4,* 67-83.

Piolat, A., Olive, T. & Kellogg, R. T. (2005). Cognitive effort during note taking. *Applied Cognitive Psychology, 19,* 291-312. doi: 10.1002/acp.1086

Piolat, A., Roussey, J. Y., Olive, T. & Farioli, F. (1996). Charge mentale et mobilisation des processus rédactionnels: examen de la procédure de Kellogg [Mental load and time processing of writing: Test of Kellogg's procedure]. *Psychologie Française, 41-4,* 339-354.

Polya, G. (1954). *Mathematics and plausible reasoning*. Princeton, NJ: Princeton University Press.

Rinne, L. F. & Mazzocco, M. M. M. (2014). Knowing right from wrong in mental arithmetic judgments: Calibration of confidence predicts the development of accuracy. *PLoS ONE 9*(7), e98663. doi: 10.1371/journal.pone.0098663

Romano, J. L., Kromrey, J. D., Coraggio, J., Skowronek, J. & Devine, L. (2006). *The impact of method choice in the evaluation of national student surveys: What statistic is really best with ordinal level data?* Paper presented at the annual meeting of the Florida Educational Research Association, Jacksonville, FL.

Rosário, P., Mourão, R., Núñez, J. C., González-Pienda, J. A. & Solano, P. (2008). Storytelling as a promoter of self-regulated learning (SRL) throughout schooling. In A. Valle, J. C. Núñez, R. G. Cabanach, J. A. González-Pienda & S. Rodríguez (Eds.), *Handbook of instructional resources and their applications in the classroom* (pp. 107-122). New York, NY: Nova Science.

Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning*, *4*, 33-45. doi: 10.1007/s11409-008-9031-3

Schraw, G., Kuch, F. & Gutierrez, A. P. (2012). Measure for measure: Calibrating ten commonly used calibration scores. *Learning and Instruction*, *24*, 48-57. doi: 10.1016/j.learninstruc.2012.08.007

Schunk, D. (2008). Metacognition, self-regulation, and self-regulated learning: Research recommendations. *Educational Psychology Review*, *20*, 463-467.

Schunk, D. & Pajares, F. (2009). Self-efficacy theory. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 35-53). New York, NY: Routledge.

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. doi: 10.1016/j.ijer.2013.10.008

Stoeger, H. & Ziegler, A. (2008). Evaluation of a classroom based training to improve self-regulation in time management tasks during homework activities with fourth graders. *Metacognition and Learning, 3*(3), 207-230. doi: 10.1007/s11409-008-9027-z

Stolp, S. & Zabrucky, 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.

Throndsen, I. (2011). Self-regulated learning of basic arithmetic skills: A longitudinal study. *British Journal of Educational Psychology, 81*, 558-578. doi: 10.1348/2044-8279.002008

Tillema, M., Van den Bergh, H., Rijlaarsdam, G. & Sanders, T. (2011). Relating self reports of writing behaviour and online task execution using a temporal model. *Metacognition and Learning, 6*(3), 229-253. doi: 10.1007/s11409-011-9072-x

Veenman, M. V. J., Elshout, J. J. & Groen, M. G. M. (1993). Thinking aloud: Does it affect regulatory processes in learning? *Tijdschrift voor Onderwijsresearch, 18*, 322-330.

Veenman, M. V. J. (2011). Learning to self-monitor and self-regulate. In R. Mayer & P. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 197-218). New York, NY: Routledge.

Whimbey, A. & Lochhead, J. (1993). *Comprender y resolver problemas [Problem solving and comprehension]*. Madrid: Visor.

Williams, J. R. (2008). Revising the Declaration of Helsinki. *World Medical Journal, 54*, 120-125.

Winne, P.H. (2001). Self-regulated learning viewed from models of information processing. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement* (pp. 153-190). New York, NY: Erlbaum.

Winne, P. H. (2004). Students' calibration of knowledge and learning processes: Implications for designing powerful software learning environments. *International Journal of Educational Research, 41*(6), 466-488. doi: 10.1016/j.ijer.2005.08.012

Winne, P. H. & Muis, K. (2011). Statistical estimates of learners' judgments about knowledge in calibration of achievement. *Metacognition and Learning, 6*(2), 179-193. doi: 10.1007/s11409-011-9074-8

Zimmerman, B. (2000). Attaining self-regulation. A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13-39). San Diego, CA: Academic Press.

Zimmerman, B. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal, 45*(1), 166-183. doi: 10.3102/0002831207312909

Zimmerman, B. J., Moylan, A., Hudesman, J., White, N. & Flugman, B. (2011). Enhancing self-reflection and mathematics achievement of at-risk urban technical college students. *Psychological Test and Assessment Modelling, 53*(1), 106-140.

Zimmerman, B. & Schunk, D. H. (2011). *Handbook of self-regulation of learning and performance.* New York, NY: Routledge.

METACOGNITIVE KNOWLEDGE AND SKILLS IN STUDENTS WITH DEEP APPROACH TO LEARNING. EVIDENCE FROM MATHEMATICAL PROBLEM SOLVING

TRINIDAD GARCÍA, MARISOL CUELI, CELESTINO RODRÍGUEZ, JENNIFER KRAWEC* & PALOMA GONZÁLEZ-CASTRO

UNIVERSITY OF OVIEDO, *UNIVERSITY OF MIAMI

SUBMITTED FOR PUBLICATION

ABSTRACT. Student approaches to learning and metacognitive strategies are two important conditioning factors in solving mathematical problems. The evidence suggests that it is the deep approach to learning which leads to student success in such tasks. The present study focused on analyzing the differences in metacognitive knowledge and skills in a sample of 524 fifth and sixth grade students divided into three groups based on their different levels of use of a deep approach (241= low; 152= medium; and 131= high). Metacognitive knowledge was assessed using the Learning Strategies Knowledge Questionnaire, while evidence about metacognitive skills was gathered by means of process measures (Triple Task Procedure) during students' solving of two mathematical word problems. Statistically significant differences in metacognitive knowledge were found among groups while differences in metacognitive skills were only found in the second task, with a low effect size. Results are discussed in terms of past and future research.

Keywords: Deep approach to learning, Elementary school, metacognitive knowledge, metacognitive skills, Mathematics problem solving.

RESUMEN. El enfoque de aprendizaje y las estrategias metacognitivas son importantes condicionantes en la resolución de problemas matemáticos. La investigación ha puesto de relevancia que el enfoque profundo de aprendizaje dirige al estudiante al éxito en la ejecución de estas tareas. Este trabajo ha pretendido analizar las diferencias en el conocimiento y habilidades metacognitivas de 524 estudiantes de quinto y sexto de primaria clasificados en tres grupos en función del nivel de uso del enfoque profundo (241= bajo; 152= medio; 131= alto). El conocimiento metacognitivo fue evaluado con el cuestionario de conocimiento de estrategias de aprendizaje, y las habilidades metacognitivas con medidas del proceso (Triple Tarea) durante la resolución de dos problemas matemáticos. Los resultados mostraron diferencias estadísticamente significativas en el conocimiento metacognitivo y, en las habilidades metacognitivas en la segunda tarea con un bajo tamaño del efecto. Se discuten los resultados en función de la investigación pasada y futura.

Palabras clave: enfoque profundo, educación primaria, conocimiento metacognitivo, habilidades metacognitivas, resolución de problemas matemáticos.

1. BACKGROUND

Solving mathematical problems is known to be an important issue in formal educational setting. However, students often have difficulties in this area. This activity, which is first introduced in the early elementary years, may suppose the basis to develop further problem-solving skills. Much of the research has focused on the cognitive and metacognitive processes involved in solving mathematical word problems, highlighting the importance of self-regulated learning (SRL) as a determining factor of problem-solving proficiency (Cleary & Chen, 2009; Krawec, Huang, Montague, Kressler, & Melia, 2012; Montague, Enders, & Dietz, 2011). In this way, recent studies have highlighted the relationship between SRL and the learning approaches adopted by the student as a key aspect of his or her achievement, specifically in problems that require deep processing (García, Betts, González-Castro, González-Pienda, & Rodríguez, 2015; Phan, 2011; Ranellucci et al., 2013).

1.1. Approaches to Learning

According to Baeten, Kyndt, Struyven and Dochy (2010), an approach to learning embeds both the intention of the student when starting a task and the learning processes and strategies they use to carry out this task. Biggs (1987) distinguished three approaches to learning (deep, surface and achieving), which have been substantiated in a wide range of studies in several educational levels (Kizilgunes, Tekkaya, & Sungur, 2009). A *deep approach* involves an attempt to integrate new information with prior knowledge and enables students to organize new information, relate ideas, and monitor their understanding of the information, which would be translated into a better performance (McInerney, Cheng, Mok, & Lam, 2012). This kind of learning is motivated intrinsically and implies a semantic understanding of the information, an essential component in acquiring meaningful and long-term knowledge. A deep approach to learning is intimately linked to academic success (Murayama, Pekrun, Lichtenfeld, & vom Hofe, 2013). A s*urface approach* to learning involves rote memorization without deep elaboration. This approach is motivated extrinsically, aimed at fulfilling situational demands such as getting the assignments done or the courses passed. It commonly leads to acquire knowledge that fades quickly (McInerney et al., 2012; Murayama et al., 2013). Finally, an *achieving approach* refers to students making an effective use of space and time in order to maximize outcomes.

The students´ choice of an approach to learning and its effectiveness relies on many different factors such as age, personality traits, preferences for teaching methods, self-efficacy beliefs, or the characteristics of the task (Baeten et al., 2010; Chen & McNamee, 2011; Malmberg, Jarvenoja, & Jarvela, 2013). Thus, students´ approach to learning may vary considerably. However, authors such as Biggs (1993) noted that only the deep approach is “task focused”; that is, students using a deep approach pursue meaning and integration of knowledge, which leads to better results (Cano, García, Justicia, & García-Berbén, 2014). As such, studies in different educational levels have focused on encouraging students to develop a deep approach to learning (Gargallo, Almerich, Suárez, García, & Garfella, 2013; Murayama et al., 2013).

In this sense, there is compelling evidence that a deep understanding of content is assumed to emerge principally from strategic activities that prompt the learner to generate inferences, connecting what is being learned to other content as well as to what the learner already knows (McNamara & Magliano, 2009). Thus, the use of a deep approach to learning involves strategies such as reading deeply, thinking about what one has read or done, and making connections with prior knowledge or courses of action. All these processes require time and mental effort, and are related to students´ ability to monitor their own learning progress in order to be successful (Biggs, 1987). This has led many authors to study to what extent a deep approach is linked to metacognitive and self-regulatory mechanisms (Phan, 2011; Ranellucci et al., 2013).

1.2. Metacognitive Knowledge and Skills

Metacognition is characterized by Pintrich (2000) as a superordinate ability to direct and regulate cognitive, motivational, and problem-solving processes in order to achieve a specific goal (in Ifenthaler, 2012). It involves two main components: knowledge and skills (Lucangeli & Cabriele, 2006). *Metacognitive knowledge* refers to declarative knowledge about learning strategies, procedural knowledge about how to use these strategies, and conditional knowledge about when and why to use them. *Metacognitive skills* (or application of this knowledge) involve those aspects that facilitate the control and regulation of one’s cognitive system and learning process.

These skills refer to abilities such as planning, self-monitoring, and self-evaluation and would represent the authentic procedures and strategies used during task performance to monitor and control one’s cognition (Efklides, Kiorpelidou, & Kiosseoglou, 2006; Pennequin et al., 2010). Traditionally, both metacognitive components have been assessed through self-report techniques, mainly by means of questionnaires and interviews. However, many authors argue that, given its strategic and complex nature, evidence from metacognitive skills should be gathered by using methods based on real-time, or concurrent measures (Veenman, 2011). This approach, commonly referred to as *on-line* methods, includes procedures such as the think-aloud (Montague et al., 2011) or the Triple Task protocols (Kellog, 1987; Piolat, Kellogg, & Farioli, 2001).

The distinction between these two components lies in the fact that, even when students seem to show good knowledge about mathematical concepts or metacognitive processes, they often fail solving problems as a consequence of a lack of adequate skills to properly use this knowledge or to structure the solution process. The problem is that even when students know the different strategies, they do not necessarily are able to use them purposefully (Bransford, Brown, & Cocking, 2000). It has been argued in this sense that metacognitive knowledge might precede metacognitive skills (Pennequin et al., 2010). In this way, these authors pointed out that while metacognitive knowledge seems to start developing at the age of 6, the proper application of this knowledge (i.e., metacognitive skills) seems not to reach maturity until early adolescence (11-12 years). As such, the present study focused on this particular age range, using a wide sample of 5th and 6th grade students from Northern Spain.

1.3. The present study

In this context, the present study was aimed at analyzing the relationship between a deep approach to learning and metacognition in its two dimensions (metacognitive knowledge and skills), focused the latter on mathematical problem-solving situations. To accomplish that, the sample of students was divided into three groups based on their use of a deep approach to learning (UDAL), determined by means of the Processes Study Inventory-PSI (Núñez et al., 2011; Rosário et al., 2013). Metacognitive knowledge was assessed through the Learning Strategies Knowledge Questionnaire (LSKQ: Núñez et al. 2011; Rosário, Mourão, Núñez, González-Pienda, & Solano, 2006), while the application of this knowledge (or metacognitive skills) were assessed through on-line methods. Specifically, the Triple Task Procedure in Mathematics (TPTM; García & González-Pienda, 2012; García et al., 2015) was used to obtain an evidence of the metacognitive processes involved in solving two mathematical word problems.

Given the main aim of the present study, two specific goals were established: 1) to determine if students with different levels of use of a deep approach to learning exhibited different degrees of metacognitive knowledge, evaluated by means of self-report methods;

and 2) to examine if those students showed different metacognitive skills, assessed through on-line measures. Given students´ age range (10-12 years), it is presumable that they have acquired a basic metacognitive knowledge, but they are not necessarily able to properly use this knowledge in mathematics problem-solving situations.

2. METHOD

2.1. Participants

Five hundred and twenty-four students from 12 primary schools in northern Spain took part in this study. Their ages ranged from 10 to 13 years (M = 10.99, SD = 0.716). Of these students, 49.6% were female (n = 260). The sample comprised 220 students from the 5th grade (42%; male = 108, female = 112) and 304 students from the 6th grade of elementary school (58%; male = 156, female = 148). Sample selection was made through convenience or accessibility procedures. Students volunteered for the study and presented informed consent from their parents after they had received a complete description of the study. Children with a diagnosis of severe learning disabilities were excluded from the analyses. Students were divided into three groups based on the deepness of their approach to learning. There were no differences among groups in age [$F(2, 523) = 1.166$, $p = .314$].

Group 1 (low use of a deep approach to learning -UDAL) was composed of students 241 with a score below the 50th percentile on the Processes Study Inventory-PSI (Núñez et al., 2011; Rosário et al., 2013). Of them, 116 were female (48.1%) and 125 male (51.9%). Ages ranged from 10 to 13 years (M = 10.99; SD = 0.671). A total of 95 students (39%) attended 5th grade and 147 (61%) attended 6th grade.

Group 2 (medium UDAL) was composed of 152 students with a score that ranged between the 50th and the 75th percentile on this questionnaire. Of them, 79 (52%) were female and 73 (48%) male. Age ranged between 10 and 13 years (M = 11.05; SD = 0.783). Of them, 62 students (40.8%) attended 5th grade and 90 (59.2%) attended 6th grade.

Group 3 (high UDAL) was composed of 131 students with a score over the 75th percentile on the aforementioned questionnaire: 65 (49.6%) female and 66 (50.4%) male. Age ranged between 10 and 12 years (M = 10.92; SD = 0.713). A total of 64 students (48.9%) attended 5th grade and 67 (51.1%) attended 6th grade.

2.2. Measures

Deep approach to learning was assessed by means of the Processes Study Inventory-PSI (Núñez et al., 2011; Rosário et al., 2013). This variable was used to establish the different groups of comparison in the present study. The PSI scale measures both deep and surface approaches and is made up of 12 items (6 per component). Given the aim of the present study, only the first dimension of the scale, related to the deep approach to learning, was considered. Items are positively expressed and are scored on a 5-point Likert-type format, ranging from 1 (never) to 5 (always). Sample items include “I attempt to understand and express in my own words what the book says”. The internal consistency of the dimension used, established through Cronbach’s alpha, was .72 in this sample. Although not excellent, this coefficient is higher than those found with similar tools, such as the Learning Process Questionnaire (LPQ; Biggs, Kember, & Leung, 2001).

Metacognitive knowledge was assessed by means of the Learning Strategies Knowledge Questionnaire (LSKQ: Núñez et al., 2011; Rosário et al., 2006). This test consists of 10 questions with 3 response options, 2 false and 1 true. Maximum score in this scale is 10. The items refer to important general strategies students use in learning situations. This scale does not differentiate among different phases in the learning process, in contrast to the measure of metacognitive skills presented below. Thus, it only evaluates a dimension, based on general metacognitive knowledge. Cronbach’s alpha of the total scale was .89 in the present sample. The ten strategies evaluated are shown in Appendix 1, as well as the items of the deep approach to learning dimension.

Metacognitive skills were assessed through process measures, taken during problem solving. The Triple Task Procedure in Mathematics-TTPM (García & González-Pienda, García et al., 2015) was used for this purpose. It is an adaptation of the Triple Task technique (Piolat, Kellogg, & Farioli, 2001), initially used in the field of written composition. In Triple Task procedures, students are asked to perform three tasks simultaneously: a) a primary task under investigation (mathematics problem solving in this study), b) a secondary probe task (based on Response Time-RT), and c) a third task of categorization, in which participants are asked to label the process that was interrupted by the probe. RT´s are also collected during a base-line condition, in which students are asked to perform a probe task as a single task. TR´s in the base-line are compared with RT´s during Triple Task. The degree of interference in RT (IRT) caused by the primary task provides a measure of the cognitive effort designated to the task.

Triple Task protocols are based on directed introspection, (i.e., students are provided with a system of categories or sub-processes according to which they categorize their cognitive activity). In this sense, the main characteristic of TTPM is related to the designed category system, based on the Self-Regulated Learning (SRL) phases of planning, execution and evaluation (Núñez et al., 2011, Zimmerman, 2000) and Bransford and Stein's (1993) IDEAL model of problem solving. This model defines five steps to be taken in order to solve a problem successfully. A total of eight categories or sub-processes were proposed. Table 1 shows these sub-processes in relation to the phases and steps of the two models. In accordance with previous studies (Piolat et al., 2001, 2005; Rodríguez et al., 2012), an additional category called "other" was incorporated to include all thoughts unrelated to the mathematical task (i.e., day-dreaming). The categories "reading" and "writing", which may result ambiguous at first glance, refer to reading the statement of the problem and writing a response to the problem, respectively. Prior to the administering the TTPM protocol, students were trained in recognizing each of the categories.

TABLE 1

Category system. Based on the Self-Regulation Model (Rosário et al., 2008; Zimmerman, 2000) and the IDEAL Model (Bransford and Stein, 1993)

SRL Model	IDEAL Model	Process categories (I am ...)
Planning	Identification of the problem	Reading
	Definition and representation	Drawing or summarizing
		Recalling similar problems
	Exploration of possible strategies	Thinking about a solution
Execution	Action based on the strategy	Calculating
		Writing a response
Evaluation	Look at effects of solutions	Reviewing
		Correcting mistakes
"Other"		Doing something unrelated

The TTPM was administered during the performance of two mathematical word problems. These problems were taken from the book *"Problem solving and comprehension"* (Whimbey & Lochhead, 1993), published in Spanish. A description of the two problems is given below:

Problem 1*: "Beatriz lends €700 to Susana. But Susana borrows €1500 from Ester and €300 from Juana. In addition, Juana owes Ester €300 and Beatriz €700. One day they meet at Beatriz's home to settle their debts. Who went back home with €1800 more than she brought?"* Correct Response: Esther

Problem 2: *"Paula, Mari and Juana have a total of 16 dogs, 3 of which are poodles, 6 are hounds, and the rest of them are German shepherds and Pekinese dogs. Juana does not like poodles and Pekinese dogs, but she has 4 hounds and 2 German shepherds, giving a total of 6 dogs. Paula has a poodle and 2 more dogs, which are German shepherds. Mari has 3 Pekinese dogs and several dogs of other breeds. Which breeds, and how many dogs of each breed, does Mari have?"* Correct Response: Mari has two poodles and two Pekinese dogs.

This kind of problems is stated in everyday language and deal with fairly realistic situations. They are characterized as *context problems* (Harskamp & Suhre, 2006), and do not contain algebraic expressions or other mathematical means to guide the students. As they do not involve specific mathematical contents, students´ failure in their resolution is assumed to be due to other, more strategic, processes.

2.3. Procedure

The study was conducted in accordance with The Helsinki Declaration of the World Medical Association (Williams, 2008). The evaluation was collectively administered during two regular classes. The deep approach to learning and metacognitive knowledge (using self-report measures and a strategy recognition test, respectively) were assessed in the first session. Metacognitive skills, evaluated by means of process measures (TTPM), were taken in the second session, concurrent to the performance of the two mathematical problems.

The first phase in the application of the TTPM consisted of training students to become familiar with the system of categories and the evaluation procedure. The hypothetical case of a boy of their age (Alex) who tried to solve a mathematical problem, was used for this purpose. After training, students performed a category-recognition test consisting of 18 multiple-choice items (2 per category), with four response alternatives. Students had to indicate the category that best fitted each proposed activity. For instance, for the statement

"Alex realized he made a mistake, so he is erasing", the alternatives were Alex is: "thinking about a solution", "writing a response", "doing something unrelated to the task" or "correcting mistakes".

After training, students performed a base-line RT´s task, based on responding to a tone presented at different time intervals (10-15 seconds). Then, students were told that this same sound would appear while solving the problems, after which a box with the category system would appear on their computer screen. They had to choose the category that best fitted the activity that they were doing in each moment. Although a measure of RT´s was also taken during TTPM administration, the analysis of this variable was not incorporated as it exceeds the purpose of the present study. Beeps were presented in intervals of 40-45 seconds and were presented during task performance until students indicated to have finished each problem, by clicking on a button set up to this purpose. Time intervals were programmed so that regardless of students' response speed and the time at which they started or finished each task, and the intervals among beeps were constant among participants. As data were collected from individual students simultaneously, headphones were provided in order that other students were not disturbed.

Data collection was implemented through Moodle platform (https://moodle.org). A special module was created on the platform, which hosted the evaluation procedure and stored data. In order to accomplish that, a multidisciplinary team, including psychologists, teachers, and a computer engineer (responsible for all technical issues) collaborated during the study. Students accessed this platform through an individual username and password in order to guarantee their anonymity. Once data were obtained and stored, they were automatically transferred to an Excel file for subsequent processing. Students were given the mathematical problems on paper, which were also displayed on the computer screen. Students used the paper to perform the problems, and they were asked to write a final answer to each problem when they finished.

Students who showed difficulties understanding the category system (those with less than 90% correct responses in the categorization test) were excluded from the analyses. Process variables were based on frequency counts. Frequency of each sub-process or category was established by dividing the frequency of election of each category by the total number of elections across categories. These frequency counts were then transformed into percentages by multiplying the quotient by 100.

2.4. Data analysis

The data were analyzed in two steps. First, in order to analyze differences among groups with different use of a deep approach to learning (UDAL) in metacognitive knowledge, Univariate Analysis of the Variance (ANOVA) was conducted. Considering the criterion of Finney and Di Stefano (2006), for whom intervals of ±2 and ±7 are the allowable values of skewness and kurtosis respectively to conduct parametric analyses, this variable met normality conditions (see results section). For the interpretation of effect sizes, Cohen´s (1988) criterion was used, which establishes that the effect is small when $\eta p2$ = .01 (d = .20), medium when $\eta p2$ = .059 (d = .50), and high if $\eta p2$ = .138 (d = .80). Schefee´s post-hoc analyses were carried out to determine differences between pairs of groups. Second, and given that process variables (TTPM sub-processes) did not meet normality conditions (see Tables 2 and 3), nonparametric analyses were conducted. Specifically, Kruskal-Wallis statistic was used to analyze differences among groups. As the nonparametric nature of these statistics did not allow to carry out post-hoc analyses, differences between pairs of groups were analyzed using Mann-Whitney U-statistic for each comparison. Cliff's delta was taken as a measure of effect size of the differences between pairs of groups (Macbeth, Razumiejczyk, & Ledesma, 2011). It provides a measure of dominance, or the degree of overlapping between two distributions of scores. The value of this statistic ranges from –1 (if scores in Group 2 are larger than scores in Group 1) to +1 (if scores in Group 2 are smaller than scores in Group 1), and takes on the value of zero if the two distributions are similar (i.e., absence of significant differences between groups in the measured variables).

SPSS v.19 was used to conduct statistical analyses, except for calculating effect size for process variables. Cliff's Delta Calculator (CDC; Macbeth et al., 2011) was used for this purpose. In order to interpret effect size, Cohen (1988) established a correspondence between Cohen's d and this statistic. In this sense, a δ value of ±.147 will have an effect size of d = .20, a δ value of ±. 33 will correspond to an effect size of d = .50, and a δ of ± .474 will have an effect size of d = .80 (Romano et al., 2006). Separate analyses were conducted for each task. A p-value ≤ .05 was established as criterion of statistical significance.

3. RESULTS

3.1. Differences among groups with different UDAL in Metacognitive Knowledge

The mean of the general group in the LSKQ scale (Núñez et al., 2011) indicated a moderate but not high metacognitive knowledge among the students (*M* = 5.527; *SD* = 1.870). Descriptive analyses also indicated that this variable met normality conditions (*Kurtosis* = - .234, *Skewness* = - .017).

ANOVA indicated the existence of statistically significant differences in metacognitive knowledge among groups with different use of a deep approach to learning (UDAL) [$F(2, 523) = 17.463$, $p \leq .001$] with a medium effect size ($\eta p^2 = .063$). In this sense, increasing levels of UDAL related to an increase in metacognitive knowledge (low UDAL: *M* = 5.037, *SD* = 1.919; medium UDAL: *M* = 5.793, *SD* = 1.714; high UDAL: *M* = 6.122, *SD* = 1.723). Scheffé's post-hoc analyses indicated that differences were significant between the groups with low and medium UDAL ($p \leq$. 001), and the groups with low and high UDAL ($p \leq$. 001).

3.2. Differences among groups with different UDAL in Metacognitive Skills

Descriptive statistics for process variables in each task are shown in Tables 2 and 3. Kurtosis and skewness indicated that these variables did not meet normality criterion (Table 2). Therefore, nonparametric analyses were conducted.

Regarding the metacognitive process showed by students in the general group, means indicate the presence of quite similar profiles in both problems (see Tables 2 and 3). In this sense, although students showed a tendency to read - the problem statement- more in the first than in the second task, profiles seemed to be coincident in the rest of sub-processes. Specifically, students reported spending most time doing calculations in comparison to the rest of sub-processes in both problems. As could be expected, the amount of time they spent on calculations was significantly higher than the time spent on writing an answer to the problem. Students also reported spending a large amount of time thinking about solutions, and less time using strategies such as recalling (recollecting previous knowledge) or drawing/summarizing (organizing the information) in the two problems. Finally, regarding evaluation processes, the use of reviewing and correcting mechanisms were the least frequently reported by students in both problems.

With relation to differences among groups with different use of a deep approach to learning (UDAL), Kruskal-Wallis statistic indicated a lack of significant differences among groups in Problem 1. Thus, results in this sense are explained in terms of tendencies.

Specifically, means in Table 2 indicated the following pattern: groups with medium and high UDAL tended to read, use information representation strategies, and recall similar problems more frequently than the group with low UDAL. On the contrary, this last group showed to spend more time thinking about solutions; the group with high UDAL tended to do calculations less, but to write a response and use revision mechanisms more frequently than the other groups; finally, no differences seemed to be observed in correcting mistakes.

TABLE 2

Descriptive statistics for process variables, and group differences in Metacognitive Skills. Problem 1

Process categories	Group 1 (*n*= 241)	Group 2 (*n*= 152)	Group 3 (*n*= 131)	General group		
	M (SD)	*M(SD)*	*M(SD)*	*M(SD)*	*Kurtosis*	*Skewness*
Reading	15.937 (14.460)	17.092 (15.555)	16.603 (15.823)	16.438 (15.109)	5.871	1.952
Drawing or summarizing	10.885 (17.408)	12.210 (17.996)	12.438 (16.556)	11.984 (17.352)	2.507	1.663
Recalling similar problems	2.829 (6.794)	3.065 (6.381)	3.328 (6.913)	3.022 (6.697)	8.160	2.693
Thinking about a solution	20.564 (18.424)	18.322 (16.619)	17.610 (15.273)	19.175 (17.181)	1.250	1.111
Calculations	27.240 (22.523)	28.848 (24.009)	26.877 (22.102)	27.616 (22.832)	-.314	.646
Writing	11.340 (11.999)	10.842 (11.204)	13.542 (13.023)	11.746 (12.066)	1.942	1.280
Reviewing	6.070 (9.412)	5.710 (8.695)	7.404 (9.466)	6.299 (9.230)	2.319	1.566
Correcting mistakes	3.688 (7.731)	3.875 (7.935)	3.771 (8.433)	3.763 (7.955)	8.431	2.658

Note. *M* = Mean; *SD* = Standard Deviation; Group 1 = low use of a deep approach to learning (UDAL), Group 2 = medium UDAL; Group 3 = high UDAL. Means refer to the mean frequency percentage reported by students in each category.

With regard to Problem 2, Kruskal-Wallis statistic showed statistically significant differences in reading ($\chi^2(2) = 8.215$, $p = .016$), writing ($\chi^2(2) = 7.170$, $p = .028$), and reviewing ($\chi^2(2) = 17.951$, $p \leq .001$). Means in Table 3 indicate that student with low UDAL spent more time reading than the rest of the groups, while the high UDAL group showed to write and review more frequently than those with low and medium UDAL. Mann-Whitney U-test conducted for pairs of groups indicated that there were significant differences in reading between the groups with low and medium UDAL ($U = 12533.500$, $p = .005$, *Cliff's* $\delta = .1189$), while differences in writing were found between the groups with medium and high UDAL ($U = 8160.500$, $p = .008$, *Cliff's* $\delta = -.1324$). Finally, significant differences in reviewing were found between the groups with low and high UDAL ($U = 12522.500$, $p \leq .001$, *Cliff's* $\delta = -.1389$) and between the groups with medium and high UDAL ($U = 8236.000$, $p = .003$, *Cliff's* $\delta = -.1324$). The negative sign of Cliff's δ statistic indicated that means were higher in the second group of comparison. Taking Cohen´s (1988) correspondence criterion into consideration, results from these analyses suggest a low effect size.

TABLE 3

Descriptive statistics for process variables, and group differences in Metacognitive Skills. Problem 2

Process categories	Group 1 (*n* = 241)	Group 2 (*n* = 152)	Group 3 (*n* = 131)	General group		
	M(SD)	*M(SD)*	*M(SD)*	*M(SD)*	*Kurtosis*	*Skewness*
Reading	14.771 (13.928)	11.914 (15.976)	12.816 (13.931)	13.454* (14.579)	8.040	2.232
Drawing or summarizing	13.215 (21.254)	12.782 (20.286)	14.389 (22.005)	13.383 (21.138)	3.043	1.868
Recalling similar problems	2.352 (6.143)	3.118 (8.081)	1.938 (4.918)	2.471 (6.510)	16.305	3.553
Thinking about a solution	18.527 (20.208)	19.322 (19.100)	17.954 (18.599)	18.614 (19.467)	2.110	1.439
Calculations	31.107 (25.056)	33.335 (26.090)	27.984 (24.668)	30.973 (25.298)	-0.847	0.438
Writing	13.933 (14.238)	11.776 (13.646)	15.771 (14.831)	13.767* (14.270)	2.435	1.375
Reviewing	3.062 (6.084)	3.703 (6.967)	6.984 (10.113)	4.229*** (7.683)	4.877	2.052
Correcting mistakes	3.203 (7.550)	4.203 (9.681)	2.236 (5.843)	3.251 (7.888)	17.255	3.498

Note. *M* = Mean; *SD* = Standard Deviation; Group 1 = low use of a deep approach to learning (UDAL), Group 2 = medium UDAL; Group 3 = high UDAL. Means refer to the mean frequency percentage reported by students in each category.

*p≤.05; **p≤.01; ***p≤.001

The process profile concerning the non-significant variables indicated the following tendency: the group with high UDAL tended to read, recall similar problems and think about solutions less, but use strategies of organization of the information (drawing or summarizing) more frequently than the remaining groups; the group with high UDAL also tended to perform calculations and correct mistakes to a lesser extent that their peers in the other groups.

Figure 1 shows the metacognitive process exhibited by the three groups of students in Problem 2.

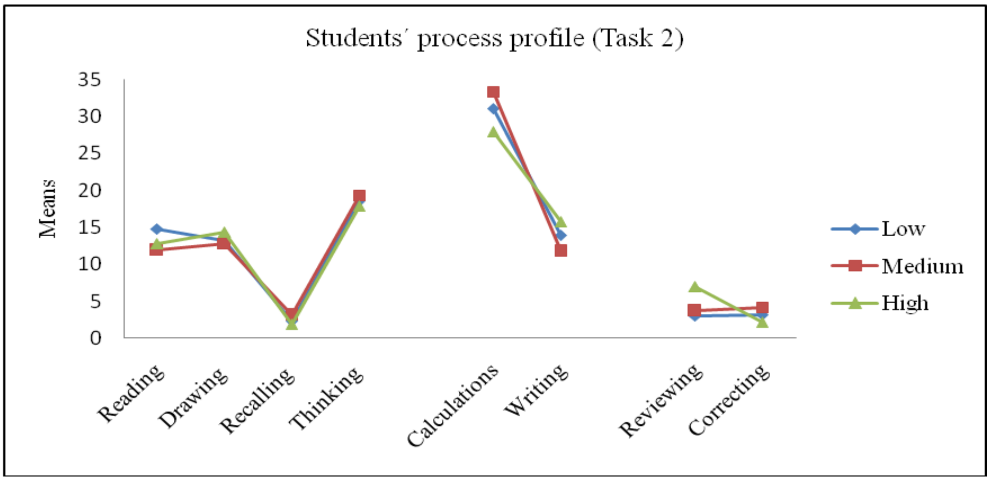


Figure 1. Process reported by students in the three groups in Problem 2.

Statistically significant differences among groups with different use of a deep approach to learning (UDAL) were found in the sub-processes of "reading", "doing calculations" and "reviewing".

4. Discussion and conclusions

The main purpose of the present study was to analyze the differences in metacognitive knowledge and skills showed by a sample of 524 upper-elementary students, divided into three groups as function of their different use of a deep approach to learning (low, medium, and high UDAL).

4.1. Differences among groups with different UDAL in Metacognitive Knowledge

Results indicated that students in the general group had a moderate but not high level of knowledge of the metacognitive strategies, evaluated by means of a strategy recognition test. However, statistically significant differences among groups with different UDAL were found in this variable. Specifically, a higher UDAL was related to increasing levels of metacognitive knowledge. These findings reflect those of previous studies that highlighted an association between metacognitive mechanisms and a deep approach to learning (Lazakidou & Retalis 2010; Malmberg et al., 2013; Rosário et al., 2013; Schmitz & Perel, 2012).

4.2. Differences among groups with different UDAL in Metacognitive Skills

With regard to self-regulation skills, assessed though the administration of the TTPM (Triple Task Procedure in Mathematics: García & González-Pienda, 2012; García et al., 2015) in two mathematical problem-solving tasks, the results did not show the existence of a clear pattern of differences. In this way, although some significant differences among groups were found in Problem 2, mainly concerning the sub-process of revision, it can be concluded that student seem to undertake the problems in the same way, regardless of the deepness of their approach to learning. The fact that only significant differences were found in the second problem could be related to the fact that it involves more relationships to establish. This could have made this problem more complex, enhancing the use of metacognitive skills, mainly those related to revision mechanisms. However, the absence of more problems, with different levels of difficulty, impedes to reach a clear conclusion in this sense.

There are two possible explanations for these findings:

A feasible explanation in this sense is related to the fact that, generally speaking, students in the present study showed moderate but not high levels of metacognitive knowledge. This result could be related to a lack of experience with or explicit use of metacognitive strategies in common learning situations, as authors such as Sengodan and Zanaton (2012) suggested. This lack of experience could then explain why students were not particularly successful identifying those strategies in the LSKQ. Considering metacognitive knowledge as the basis for the development of metacognitive skills (Pennequin et al., 2010; Weil et al., 2013), it is not surprising that the absence of high levels of knowledge of metacognitive strategies predicts a low application of them in real situations. However, previous studies also showed that students tend to demonstrate poor metacognitive skills when they are engaged in mathematical problem-solving situations. Many students forge ahead without considering alternative decisions, jumping immediately into calculations, giving impulsive responses, trying the same strategies even when they are wrong, or using trial and error as solving mechanisms (Cleary & Chen, 2009; Montague et al., 2011). In fact, students in the general group showed a very similar process in both problems, characterized by a lack of revision mechanisms, and a clear preference for using familiar processes such as doing calculations as a method to solve the problems, instead of using other kind of strategies (i.e., information representation), which have demonstrated to be useful for successfully problem solving in previous studies (Díez-Palomar, Menéndez, & Civil, 2011; García et al., 2015; Pantziara, Gagatsis, & Elia, 2009). Among other factors, this could be related to a lack of familiarity with these sorts of tasks. Previous studies focused on the process have highlighted that novice students tend to spend more time on familiar procedures (e.g., calculations) without making sure they follow a correct solution plan. On the other hand, experts are more capable of representing problem situations in many forms (e.g., a graph, a sketch or a table), they are more flexible in the strategies they use, and are more able to monitor their resolution process (Harskamp & Suhre 2006, 2007).

Another explanation for these results could be related to the distinction made between on-line and off-line measures, and the low degree of correspondence they show (Özcan, 2014; Veenman, 2011). It is necessary to point out that the use of deep approach to learning and metacognitive knowledge were assessed by means of questionnaires (i.e., off-line measures) in the present study, while metacognitive skills were evaluated though the administration of a measure of the process, concurrent to problem-solving performance (i.e., on-line measures). Thus, an absence of significant differences in metacognitive skills could be partially explained by the use of different types of measures in this case.

Implications

Beyond the relationship between the relationship between a deep approach to learning and the components of metacognition, is important to note that while students in the present study had a moderate (maybe lower than expected) metacognitive knowledge, they show poor metacognitive skills during problem-solving situations. This suggests, therefore, that they might have not developed adequate metacognitive skills yet, at least for the sort of tasks they were required to tackle in this study (i.e., mathematical problem solving). As Valle et al. (2009) pointed out, for students to be successful in learning, they need to be able to understand and know these strategies, but also put them into practice accurately. While the reasons behind this mismatch may arise from different sources, including students' age or maturity (Pennequin et al., 2010), a feasible explanation for the results in this study could be related to a lack of experience or familiarity with specific situations that potentially enhance the use of metacognitive strategies, such as the type of problems students used in the present study (i.e., mathematical word problems based on daily situations). This aspect has an important educational implication that is the need for teachers to provide students with learning situations and tasks that help them to develop these skills. As Harskamp and Suhre (2006) argue, teachers of mathematics often teach students to solve mathematical problems by copying standard solution methods or using invariant formulae. Thus, it is not surprising that many students find it difficult to solve new problems, especially those whose correct resolution does not depend on knowing specific mathematical contents or formulae application. These kinds of tasks represent an important requirement in further educational levels. Nevertheless, they are not promoted enough at Elementary School (Lazakidou & Retalis, 2010).

Limitations

Finally, we acknowledge the following limitations: first, the low effect size found in process measures calls for a certain degree of caution concerning the scope of our findings. As this low effect size seems to be related to the high variability observed in student problem-solving processes, obtaining more homogeneous groups may be imperative in future research; second, students were classified according to their learning approach to learning by administering a unique self-report scale. Collecting information from other sources, such as teachers´ reports, would be appropriate in future studies; third, only the deep approach to learning was taken into account to classify students. Additionally, some studies highlighted that students might not show only one approach and that the choice they make depends on the characteristics of the task or other personal or contextual features. Thus, different approaches and their possible interactions should be considered in further research (Chen & McNamee, 2011; Malmberg et al., 2013); finally, factors related to the application of on-line measures must also be noted. Specifically, the possibility that TTPM may result reactive for students, leading to changes in both process and performance. In this line, Branch (2000) found that an adequate degree of automation reporting these processes (i.e., previous training as included in the present study) could mitigate this problem. In addition, Kellog (1987, in Olive et al., 2001), using Triple Task procedures, found that although the use of this sort of measure may cause an increase in the time to complete the task, simply instructing participants to verbalize of categorize their thoughts during a task does not alter the sequence of the cognitive processes or performance. The use of low- to medium-rate time intervals between probes (e.g., 40-45 seconds, as in this study) has also demonstrated to help reduce this effect (García & Rodríguez, 2007). However, a greater degree of control over this possible effect should be included in future studies, maybe by retaining the measure of cognitive effort.

References

Baeten, M., Kyndt, E., Struyven, K. & Dochy, F. (2010). Using student-centered learning environments to stimulate deep approaches to learning: factors encouraging or discouraging their effectiveness. *Educational Research Review, 5*(3), 243-260, doi: 10.1016/j.edurev. 2010.06.001

Biggs, J. (1987). *Student Approaches to Learning and Studying Hawthorn*. Australia: Australian Council for Educational Research.

Biggs, J. (1993). What do inventories of students' learning processes really measure? A theoretical review and clarification. *British Journal of Educational Psychology, 63*(1), 3–19. doi: 10.1111/j.2044-8279.1993.tb01038.x

Biggs, J., Kember, D. & Leung, D.Y.P. (2001). The revised two-factor study process questionnaire: R-SPQ-2F. *British Journal of Educational Psychology*, *71*(1), 133-149. doi: 10.1348/000709901158433

Branch, J. L. (2000). Investigating the information-seeking processes of adolescents: The value of using think-alouds and think afters. *Library and Information Science Research, 22*(4), 371–392.

Bransford, J. D., Brown, A. L. & Cocking, R. R. (Eds.). (2000). How people learn: Brain, mind, experience, and school. Washington: National Academies Press.

Bransford, J. D., & Stein, B. S. (1993). *The ideal problem solver: A guide for improving thinking, learning and creativity (2nd Ed.).* New York, NY: W. H. Freeman.

Cano, F., García, A., Justicia, F. & García-Berbén, A. B. (2014). Enfoques de aprendizaje y comprensión lectora: el papel de las preguntas de los estudiantes y del conocimiento previo. *Revista de Psicodidáctica, 19*(2), 247-265. doi: 10.1387/RevPsicodidact.10186

Chen, L.G. & McNamee, G. (2011). Positive approaches to learning in the context of Preschool classrooms activities. *Early Childhood Education Journal, 39*(1), 71-78. doi: 10.1007.s10643-010-0441-x

Cleary, T.J. & Chen, P.P. (2009). Self-regulation, motivation, and math achievement in middle school: Variations across grade level and math context. *Journal of School Psychology*, *47*(5), 291-314. doi: 10.1016/j.jsp.2009.04.002

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)*. Hillsdale, NJ: Lawrence Erlbaum.

Díez-Palomar, J., Menéndez, J. M. & Civil, M. (2011). Learning mathematics with adult learners: drawing from parents´ perspective. *Revista Latinoamericana de Investigación en Matemática Educativa, 14*(1), 71-94.

Efklides, A., Kiorpelidou, K. & Kiosseoglou, G. (2006). Worked out examples in mathematics: Effects on performance and metacognitive experiences. In A. Desoete & M. Veenman (Eds.), *Metacognition in mathematics* (pp. 11–33). New York, NY: Nova Publishers.

Finney, S. J. & DiStefano, C. (2006). Non-normal and categorical data in structural equation modeling. In G.R. Hancock & R.O. Muller (Eds.), *Structural equation modeling: A second course* (pp. 269-314). Greenwich, CT: Information Age.

García, J. N. & Rodríguez, C. (2007). Influence of the recording interval and a graphic organizer on the writing process/product and on other psychological variables. *Psicothema, 19*(2), 198-205.

García, T. & González-Pienda, J. A. (2012). Evaluación del proceso de aprendizaje autorregulado en el área de las matemáticas mediante pizarras digitales. [Assessment of the self-regulated learning process in Mathematics through electronic blackboards]. In J. Dulac & C. Alconada (Coord.), *III Congreso Pizarra Digital. Publicación de comunicaciones* (pp.105-117). Madrid: Pluma y Arroba.

García, T., Betts, L., M., González-Castro, P., González-Pienda, J. A. & Rodríguez, C. (in press). On-line assessment of the process involved in Maths problem solving in fifth and sixth grade students: self-regulation and achievement. *Revista Latinoamericana de Investigación en Matemática Educativa.*

Gargallo, B., Almerich, G., Suárez, J. M., García, E. & Garfella, P. R. (2013). Learning styles and approaches to learning in excellent and average first-year university students. *European Journal of Psychology of Education, 28*(4), 1361-1379. doi: 10.1007/s10212-012-0170-1

Harskamp, E. & Suhre, C. (2006). Improving mathematical problem solving: A computerized approach. *Computers in Human Behavior, 22*(5), 801-815. doi: 10.1016/j.chb.2004.03.023

Harskamp, E. & Suhre, C. (2007). Schoenfeld´s problem solving theory in a student controlled learning environment. *Computers & Education, 49*(3), 822-839. doi: 10.1016/j.compedu.2005.11.024

Ifenthaler, D. (2012). Determining the effectiveness of prompts for self-regulated learning in problem-solving scenarios. *Educational Technology & Society, 15*(1), 38–52.

Kellog, R. T. (1987). Writing performance: Effects of cognitive strategies. *Written Communication, 4,* 269-298.

Kizilgunes, B., Tekkaya, C. & Sungur, S. (2009). Modeling the relations among students' epistemological beliefs, motivation, learning approach, and achievement. *The Journal of Educational Research, 102*(4), 243–255. doi: 10.3200/JOER.102.4.243-256

Krawec, J., Huang, J., Montague, M., Kressler, B. & Melia, A. (2012). The effects of cognitive strategy instruction on knowledge of math problems solving processes of middle school students with learning disabilities. *Learning Disability Quarterly, 36*(2), 80-92. doi: 10.1177/0731948712463368

Lazakidou, G. & Retalis, S. (2010). Using computer supported collaborative learning strategies for helping students acquire self-regulated problem-solving skills in mathematics. *Computers & Education, 54*(1), 3-13. doi: 10.1016/j.compedu.2009.02.020

Lucangeli, D. & Cabrele, S. (2006). The relationship of metacognitive knowledge, skills and beliefs in children with and without mathematical learning disabilities. In A. Desoete & M. V. Veenman (Eds.), *Metacognition in mathematics education* (pp. 103–133). New York, NY: Nova Science Publishers.

Macbeth, G., Razumiejczyk, E. & Ledesma, R. D. (2011). Cliff's Delta Calculator: A non-parametric effect size program for two groups of observations. *Universitas Psychologica, 10*(2), 545-555.

Malmberg, J., Jarvenoja, H. & Jarvela, S. (2013). Patterns in elementary school students' strategic actions in varying learning situations. *Instructional Science, 41*(5), 933-954. doi: 10.1007/s11251-012-9262-1

McInerney, D. M., Cheng, R. W., Mok, M. M. C. & Lam, A. K. H. (2012). Academic Self-Concept and Learning Strategies: Direction of Effect on Student Academic

Achievement. *Journal of Advanced Academics, 23*(3), 249-269. doi: 10.1177/1932202X12451020.

McNamara, D. S. & Magliano, J. P. (2009). Towards a comprehensive model of comprehension. In B. Ross (Ed), *The psychology of learning and motivation* (pp. 297-284). New York, NY: Elsevier Science.

Montague, M., Enders, C. & 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. doi: 10.1177/0731948711421762

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*(4), 1475-1490. doi: 10.1111/cdev.12036

Núñez, J. C., Cerezo, R., Bernardo, A., Rosário, P., Valle, A., Fernández, E. & Suárez, N. (2011). Implementation of training programs in self-regulated learning strategies in moodle format: Results of an experience in higher education. *Psicothema, 23*(2), 274-281.

Özcan, C. Z. (2014). Assessment of metacognition in mathematics: which one of two methods is a better predictor of mathematics achievement? *International Online Journal of Educational Sciences*, *6*(1), 49-57. doi: 10.15345/iojes.2014.01.006

Pantziara, M., Gagatsis, A. & Elia, I. (2009). Using diagrams as tools for the solution of non-routine mathematical problems. *Educational Studies in Mathematics, 72,* 39-60.doi: 10.1007/s10649-009-9181-5

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*(3), 198-220. doi: 10.1080/13546783.2010.509052

Phan, H. P. (2011). Deep processing strategies and critical thinking: Developmental trajectories using latent growth analyses. *Journal of Educational Research, 104*(4), 283-294. doi: 10.1080/00220671003739382

Pintrich, P. R. (2000). The role of goal orientation in self–regulated learning. En M. Boekaerts, P.R. Pintrich & M. Zeidner (Eds.), *Handbook of self–regulation* (pp. 451–502). Academic Press.

Piolat, A., Kellogg, R. T. & Farioli, F. (2001). The triple task technique for studying writing processes: on which task is attention focused? *Current Psychology Letters: Brain, Behavior and Cognition*, *4,* 67–83.

Piolat, A., Olive, T. & Kellogg, R.T. (2005). Cognitive Effort during Note Taking. *Applied Cognitive Psychology*, *19*, 291–312. doi: 10.1002/acp.1086

Ranellucci, J., Muis, K. R., Duffy, M., Wang, X., Sampasivam, L. & Franco, G. (2013). To master or perform? Exploring relations between achievement goals and conceptual change learning. *British Journal of Educational Psychology, 83*(3), 431-451. doi: 10.1111/j.2044-8279.2012.02072.x

Romano, J. L., Kromrey, J. D., Coraggio, J., Skowronek, J. & Devine, L. (2006). *The impact of method choice in the evaluation of national student surveys: What statistic is really best with ordinal level data?* Paper presented at the annual meeting of the Florida Educational Research Association, Jacksonville, Florida.

Rosário, P., Mourão, R., Núñez, J. C., González-Pienda, J. A. & Solano, P. (2006). School-family: Is a reciprocal and positive relationship possible? *Papeles del Psicólogo, 27*(3), 174-182.

Rosário, P., Núñez, J. C., Ferrando, P., Paiva, O., Lourenço, A., Cerezo, R. & Valle, A. (2013). The relationship between approaches to teaching and approaches to studying: A two-level structural equation model for biology achievement in high school. *Metacognition and Learning, 8*(1), 44-77. doi: 10.1007/s11409-013-9095-6

Schmitz, B. & Perel, P. (2011). Self-monitoring of self-regulation during math homework behaviour using standardized diaries. *Metacognition Learning, 6*(3), 255–273. doi: 10.1007/s11409-011-9076-6

Sengodan, V. & Zanaton, Z. H. (2012). Students' learning styles and intrinsic motivation in learning mathematics. *Asian Social Science, 8*(16), 17-23. doi: 10.5539/ass.v8n16p17

Valle, A., Rodríguez, S., Cabanach, R. G., Núñez, J. C., González-Pienda, J. A. & Rosário, P. (2009). Diferencias en rendimiento académico según los niveles de las estrategias cognitivas y de las estrategias de autorregulación [Differences in academic performance by levels of cognitive strategies and self-regulation strategies]. *Summa Psicológica UST, 6*(2), 31-42.

Veenman, M. V. J. (2011). Learning to self-monitor and self-regulate. In R. Mayer & P. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 197-218). New York, NY: Routledge.

Weil, L. G., Flemingc, S. M., Dumontheil, I., Kilforda, E. J., Weil, R. S., Rees, G., … & Blakemorea. S. (2013). The development of metacognitive ability in adolescence. *Consciousness and Cognition, 22*, 264–271. doi: 10.1016/j.concog.2013.01.004

Whimbey, A., & Lochhead, J. (1993). *Comprender y resolver problemas [Problem solving and comprehension]*. Madrid: Visor.

Williams, J. R. (2008). Revising the Declaration of Helsinki. *World Medical Journal, 54*, 120-125.

Zimmerman, B. J. (2000). Attaining self-regulation. A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). San Diego, CA: Academic Press.

ANALYSIS OF THE FACTORS INVOLVED IN MAKING POST-PERFORMANCE JUDGMENTS IN MATHEMATICS PROBLEM SOLVING

TRINIDAD GARCÍA, EVELYN H. KROESBERGEN *, CELESTINO RODRÍGUEZ, PALOMA GONZÁLEZ-CASTRO & JULIO ANTONIO GONZÁLEZ-PIENDA

UNIVERSITY OF OVIEDO (SPAIN), * UTRECHT UNIVERSITY (THE NETHERLANDS)

SUBMITTED FOR PUBLICATION

ABSTRACT. 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. A total of 188 students took part in the study. They were divided into two groups as function of their judgments accuracy once completed the two tasks (accurate= 79; inaccurate= 109), and differences between these groups were examined. The discriminative value of these variables predicting group membership was analyzed, as well as the effect of age, gender and grade level. The results indicated that: 1) accurate students showed better levels of executive functioning, and more positive feeling, beliefs and motivation related to mathematics. They also spent more time on the tasks; 2) mathematics achievement, perceived usefulness of mathematics, and time spent on Task 1, significantly predicted group membership, classifying 71.3% of the sample correctly. These results support the relationship between academic achievement and calibration accuracy, suggest the need for consider a wide range of factors when performance judgments want to be explained.

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

RESUMEN. Este estudio analiza la influencia del funcionamiento ejecutivo, componentes afectivo-motivacionales relacionados con las matemáticas, rendimiento académico en la asignatura y las características de la tarea, en las habilidades de calibración de estudiantes de quinto y sexto curso tras la realización de dos problemas matemáticos. Un total de 188 estudiantes participaron en este estudio, 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 de los estudiantes al grupo. Los efectos de la edad, género y nivel educativo fueron analizados. Los resultados indicaron que: 1) 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; 2) 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%. 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.

1. BACKGROUND

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 et al., 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. Both would imply different processes. Specifically, predictions would reflect a sort of self-efficacy judgments, while postdictions would inform about monitoring processes during the task. This differentiation has been supported by previous research, showing that postdictions tend to be more accurate than predictions, presumably because of the additional feedback that working on the task provides (Bol, Hacker, O'Shea, & Allen, 2012; Hacker et al., 2008; Sheldrake, Mujtaba, & Reiss, 2014). In the current study, it is investigated which factors contribute to the post-performance calibration process.

Calibration informs about the status of one's knowledge or strategies at a cognitive level, and would thus be involved in the regulation of further effort and strategy use (Van Loon, Bruin, Van Gog, & Van Merriënboer, 2013). This relationship 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 were found to perform more successfully on problem-solving tasks, maybe 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 & Zabrucky, 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 threat learning (Dupeyrat, Escribe, Huet, & Régner, 2011;

Sheldrake et al. 2014). Specifically, Dupeyrat and colleagues found that under-confident students at grades 8 and 9 made less mathematics progress during the year and had lower performance approach goals than accurate and over-confident students. Having present 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 judgments accuracy. Specifically, high-achievers have shown to be more calibrated, and somewhat under-confident, in comparison to their low-achieving peers (Bol, Hacker, Walck, & Nunnery, 2011; Hacker et al., 2008; Özsoy, 2012). In this same direction, tasks characteristics (mainly difficulty) have been widely studied, leading to mixed results. Howie and Roebers (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).

An important issue in this sense is the fact that most research has been conducted with college or undergraduate students. In the context of elementary school and mathematical problem-solving, Boekaerts and Rozendaal (2010) found interesting results. They examined the effects of gender, type of mathematical problem (application versus computation), and time of measurement (predictions versus postdictions) on fifth grader´ 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, were less calibrated in this sort of problems and that, surprisingly, performance judgments were more accurate before task completion (i.e., predictions).

Calibration accuracy has also shown to be influenced by affective-motivational variables. Individual differences in self-efficacy beliefs, goal orientation, and susceptibility to social influences, and cultural issues, may account for differences in judgments 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 that 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 their study, students with accurate perceptions of performance in mathematics at year 10 reported higher intention to study the subject into years 12 and 13 (when mathematics is currently non-compulsory in England). They also showed 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 of making 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 are related to cognitive processes such as planning, goal setting, response inhibition, impulse control, attention, self-monitoring, and cognitive flexibility. These components have demonstrated to serve as powerful predictors of school readiness and achievement, in specific of mathematics (Bull & Lee, 2014; Kolkman, Hoijtink, Kroesbergen, & Leseman, 2013).To date however, literature about the relationship between calibration and EF is very scarce. 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 pay attention to irrelevant information, making it difficult to maintain a goal in working memory. Without a clear goal on mind, one would be less able to switch strategies when necessary, and even less, to accurately evaluate one's performance. A relationship between EF and calibration is also expectable from a developmental perspective. Roebers, 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 childhood.

The present study

This study analyzed the impact of three sets of variables on students' performance judgments 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 judgments 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 characteristics: perceived difficulty and total time spent on the tasks.

This study was aimed at asking the following questions:

1) Do the groups with different calibration levels differ in domain-general, domain-specific or task-related variables? Highly-calibrated students would show significantly better levels of executive functioning, higher motivation, and more positive feeling and attitudes towards mathematics than their inaccurate peers would. They would also show higher mathematics achievement. Finally, not only perceived difficulty, but also the time spent on the task would have an impact on calibration accuracy. Longer times would be related to more accurate judgments.
2) To which extent do these variables predict group membership? Some of these variables would predict group membership, being mathematics achievement one of the most important predictors.

2. METHOD

2.1. Participants

A total 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 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 that 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.

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

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.

2.2. Measures

2.2.1. 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 (α= .94) and moderate for its factors (.67 to .74).

2.2.2. 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= s*trongly 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).

2.2.3. Task 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.

2.2.4. 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.

2.3. 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. Students and parents had previously completed the questionnaires (IAM and EFS-F, respectively). Mathematical problems were given on paper. After solving each problem, students evaluated the degree to which they had perceived the task as easy or difficult, and made a performance judgment. Actual performance was established based on student written answers. Once evidence from post-performance judgments and actual performance was obtained, students were assigned to the groups as described above.

The problems were taken from the book "Problem solving and comprehension" (Whimbey & Lochhead, 1993) translated into Spanish. Based on daily and practical situations, they did not involve any specific curricular contents or algebraic expressions. Teachers evaluated their adequacy prior to the evaluation.

2.4. 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 η^2 = .010 (*d* = .20), a medium association as η^2 = .059 (*d* = .50), and a large association as η^2 = .138 (*d* = .80). Discriminant analyses were carried out to determine the significance of these variables predicting group membership, following a stepwise method. As this analysis requires equal 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.

3. RESULTS

3.1. Group differences in the studied variables

3.1.1. Domain-general variables

Statistically significant differences between groups in EF were found (*Wilks' λ* = .901, $F(8, 145) = 5.251$, $p = .026$, η^2= .099). Only the covariate gender was statistically significant (*Wilks' λ* = .913, $F(8, 154) = 2.094$, $p = .039$, η^2 = .087). ANCOVA´s 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.

3.1.2. Domain-specific variables

Statistically significant differences between groups were found in the affective-motivational components (*Wilks' λ* = .872, $F(5, 178) = 5.242$, $p< .001$, η^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.

3.1.3. 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.

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 characteristics								
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).

3.2. 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 1significantly predicted group membership. Although significant, the inter-group variability explained by the discriminant function ("Eigen value") 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 provide 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.

TABLE 3

Results of discriminant analysis, using stepwise method

	Standardize coefficients	Structure coefficients	Function coefficients	Wilks' Lambda	*F*	*f.d*	*p*
1. Academic grades	.716	.767	.692	.880	25.195	2,355	< .001
2. Perceived Usefulness	-.561	-.655	-.128	.837	17.899	4,708	< .001
3.Time spent on Task1	.402	.463	.003	.812	14.136	6,706	< .001
Constant			-1.884				
Eigenvalue	.232						
Wilks' Lambda	.812						
(χ2; *p*)	(χ2 = 38.246; p < .001)						
Canonical correlation	.434						

Note. f.d = freedom degrees

4. Discussion and conclusions

This study analyzed the impact of executive functions, affective-motivational variables related to mathematics and academic achievement in the subject, and task-related variables on students' post-performance judgments accuracy in two mathematical problems. The initial hypotheses were confirmed:

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 student from making impulsive and erroneous judgments. Nevertheless, 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 later functions and calibration. In this sense, ¨cool¨ functions would allow a rational processing of the information, being intimately linked to learning processes, while "hot" functions, such as emotional control or cognitive flexibility, would be more related to emotional information processing, and social adaptation. García, Rodríguez, González-Castro, Álvarez-García, & González-Pienda (in press) reported similar results using the EFS 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) with a similar sample, 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 in learning mathematical, or the value that they attribute to this subject (Adelson & McCoach, 2011). These components showed their impact on student 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., 2012; Hacker et al., 2008).

Third, a non-consistent pattern of results emerged as function of perceived difficulty. However, accurate students significantly spent more time 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 (an erroneous) judgments. However, this effect was not found for Task 2, suggests a greater impact of this variable when task is novel.

Finally, the predictive value of some of the studied variables in determining group membership was highlighted, being mathematics achievement the most important predictor. Students with higher levels of achievement, that 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; Thorell, Veleiro, Siu, & Mohammad, 2013).

Limitations

These findings advocate for examining a wide range of factors when differences in calibration accuracy want 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 judgments accuracy was also highlighted. However, the fact that only two mathematical tasks were used, and the dichotomous nature of the variables, may suppose an important constraint. Increasing the number of tasks would allow to obtain more continuous measures, as well as to explore 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. It should also be noted that the studied variables correctly classified 71% of the sample, showing greater predictive value concerning the inaccurate group. 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. Introducing an experimental design could provide interesting insight into the causal relationships. Thus, amore comprehensive study of the components involved in making accurate performance judgments is necessary. This would set the basis for more adapted and effective interventions.

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. doi: 10.1177/0748175611418522

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. doi: 10.1016/j.learninstruc.2009.03.002

Bol, L., Hacker, D. J., Walck, C. C. & Nunnery, J. A. (2012). The effects of individual or group guidelines on the calibration accuracy and achievement of high school biology students. *Contemporary Educational Psychology*, *37*, 280-287. doi: 10.1016/j.cedpsych.2012.02.004

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. doi: 10.1111/cdep.12059

Cleary, T. J. & Chen, P. (2009). Self-regulation, motivation, and math achievement in middle school: Variations across grade level and math context. *Journal of School Psychology*, *47*, 291-314. doi: 10.1016/j.jsp.2009.04.002

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York, NY: Academic Press.

Duperyrat, C., Escribe, 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. doi: 10.1016./j.ijer.2011.08.005

Finn, B. & Metcalfe, J. (2014). Over-confidence in children's multi-trial judgments of learning. *Learning and Instruction, 32,* 1-9. doi: 10.1016/j.learninstruc.2014.01.001

García, T., Álvarez-García, D., Cueli, M., González-Castro, P. & Álvarez, L. (2013). Psychometric properties of the Executive Functions Scale for Families (EFS-F). *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. doi: 10.7334/psicothema2013.149

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.*

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. & da Silva, E. H. (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.

Howie, 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. doi: 10.1016/j.lindif.2013.01.004

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. doi: 10.1245/073i9487M421762

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. doi: 10.1016/j.ijer.2011.08.004

Özsoy, G. (2012). Investigation of fifth grade students' mathematical calibration skills. *Educational Sciences: Theory & Practice, 12*(2), 1190-1194.

Rinne, L. F. & Mazzocco, M. M. (2014). Knowing right from wrong in mental arithmetic judgments: Calibration of confidence predicts the development of accuracy. *PLoSONE 9*(7), e98663. doi: 10.1371/journal.pone.0098663

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. doi: 10.1007/s11409-012-9089-9

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. doi: 10.1016/j.ijer.2013.10.008

Stolp, S. & Zabrucky, 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), 630-8. doi: 10.1080/09297049.2012.727792

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. doi: 10.1016/j.learninstruc.2012.08.005

Whimbey, A. & Lochhead, J. (1993). *Comprender y resolver problemas [Problem solving and comprehension]*. Madrid: Visor.

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. doi: 10.1111/j.1750-8606.2012.00246.x

Principales resultados y discusión

Resolver problemas matemáticos es una de las actividades más frecuentes en la edad escolar. Sin embargo, no por su frecuencia deja de ser una actividad compleja, cuya correcta ejecución depende de múltiples variables. En este sentido, una de las variables que ha demostrado influir en el rendimiento en estas tareas es el grado y efectividad del control que los estudiantes ejercen sobre sus procesos cognitivos, afectivos y motivacionales durante la resolución de problemas matemáticos. De este modo, se ha demostrado que los estudiantes que son exitosos en estas tareas muestran generalmente unas mejores habilidades metacognitivas. No obstante, la literatura también evidencia que los estudiantes en diferentes etapas educativas muestran dificultades para gestionar su proceso de resolución de problemas, la cual no solamente conduce a obtener un mal resultado en la tarea, sino también a la formación de juicios poco ajustados acerca de este resultado.

Estos juicios estarían íntimamente ligados a las habilidades de *calibración*, las cuales han mostrado tener un impacto importante en el rendimiento en matemáticas y resolución de problemas. No obstante, los estudios previos coinciden en señalar la existencia de unas pobres habilidades de calibración entre los estudiantes, lo cual afecta entre otros aspectos, al rendimiento, esfuerzo, persistencia e interés en la tarea o en la propia asignatura. Esto hace necesario incidir en el estudio de los factores que puedan determinar la precisión de estos juicios. En este sentido, se diferencian dos tipos de juicios (predicciones -o juicios realizados previamente a la ejecución de la tarea-, y postdicciones -o juicios realizados a posteriori-). Estos últimos serían susceptibles de ser analizados desde la perspectiva del análisis del proceso de resolución en problemas matemáticos, lo cual podría proporcionar información de utilidad acerca de cómo ciertos patrones de resolución conducen a la formación de juicios más o menos precisos. No obstante, y partiendo de la caracterización de la calibración como un mecanismo metacognitivo complejo, se hace necesario indagar en un rango más amplio de factores, atendiendo a diferentes niveles de análisis.

Dicho lo anterior, esta Tesis Doctoral ha tenido como objetivo responder a tres cuestiones principales:

1. ¿Cómo se puede evaluar el proceso y qué puede decir éste acerca de las estrategias metacognitivas implicadas en la resolución de problemas matemáticos?
2. ¿Puede el proceso informar acerca de los mecanismos implicados en la realización de juicios tras la ejecución de este tipo de tareas?
3. ¿Hay otras variables afectivas, motivacionales, cognitivas o conductuales que puedan explicar las diferencias en estos juicios?

Para cumplir con este objetivo, esta Tesis Doctoral se concretó en la realización de tres publicaciones y tres trabajos complementarios. Para la realización de estos estudios se contó con una muestra de 524 estudiantes de quinto y sexto curso (10-13 años), pertenecientes a 12 centros educativos en el Principado de Asturias (España), los cuales realizaron dos problemas matemáticos verbales basados en situaciones de la vida real. A continuación se presentan los resultados obtenidos en estos trabajos en relación a cada una de las preguntas planteadas, así como una discusión sobre los mismos.

Primera Cuestión

¿Cómo se puede evaluar el proceso y qué puede decir éste acerca de las estrategias metacognitivas implicadas en la resolución de problemas matemáticos?

***Artículo 1:* On-line assessment of the process involved in Maths problem-solving in fifth and sixth grade students: self-regulation and achievement.**

En este estudio se presenta y aplica el procedimiento de evaluación del proceso diseñado. Este procedimiento –Triple Task Procedure in Mathematics- se basó en la adaptación de la técnica de la Triple Tarea (Kellog, 1987; Piolat et al. 2001), tradicionalmente aplicada en composición escrita. Para ello se partió de la perspectiva del Aprendizaje Autorregulado (Zimmerman, 2000, 2008) y del modelo IDEAL de resolución de problemas, propuesto por Bransford y Stein (1993). De la combinación de estos dos modelos surgió el sistema de categorías o sub-procesos que supone la base de esta herramienta.

Los resultados obtenidos en este primer estudio, llevado a cabo con una muestra de 510 estudiantes de quinto y sexto curso de Educación Primaria, mostraron la existencia de una gran variabilidad en el proceso llevado a cabo por los estudiantes, lo que se tradujo en unas desviaciones típicas elevadas. Esta variabilidad se encontró fundamentalmente en los sub-procesos menos empleados por los estudiantes, entre ellos la revisión y corrección de errores. La escasa frecuencia de uso de estos sub-procesos es consistente con estudios previos que señalan como los estudiantes en estas etapas educativas muestran escasos mecanismos de evaluación (Cleary y Chen, 2009; Kajamies et al., 2010; Kramaski y Gutman, 2006). Estos estudiantes emplearon, no obstante, una cantidad considerable del tiempo destinado a la resolución de problemas a la fase de planificación, en comparación con las fases de ejecución y evaluación. Si bien este dato parecería contradecir los resultados mostrados en estudios previos a cerca de la impulsividad mostrada por los estudiantes y la ausencia de mecanismos de planificación, un análisis en profundidad de los sub-procesos que forman parte de esta fase mostró la presencia de unas estrategias de planificación ineficaces. En concreto, los estudiantes emplearon la mayor parte del tiempo dentro de esta fase a leer el enunciado y pensar formas de resolver el problema, en lugar de al uso de estrategias potencialmente más eficaces como la organización o representación de la información. Este patrón de resultados indicaría el empleo de unas estrategias de resolución pasivas por parte de los estudiantes, o incluso problemas en la compresión de la tarea. También, dentro de esta fase de planificación, el sub-proceso relativo a recordar problemas similares fue el menos empleado por los estudiantes, lo que sugeriría una falta de familiaridad con este tipo de tareas. Finalmente, los resultados revelaron también un empleo frecuente de procedimientos quizás más familiares para los estudiantes, concretamente la realización de cálculos.

En cuanto al análisis de las diferencias en el proceso entre los estudiantes con diferente rendimiento en las áreas (relación proceso-producto), los resultados indicaron la existencia de diferencias estadísticamente significativas entre los grupos en el uso de estrategias de organización y representación de la información en el primer problema matemático. En el caso del segundo de los problemas, estas diferencias se encontraron en todos los sub-procesos implicados en la planificación. En este sentido, los estudiantes que fueron exitosos resolviendo las tareas emplearon con mayor frecuencia estrategias de representación y organización de la información, mientras que destinaron significativamente menos tiempo a recordar problemas similares, leer y pensar en formas de solucionar el problema. Estos resultados sugieren por tanto, el empleo de unas estrategias de resolución más activas y eficaces por parte de los estudiantes con éxito en las tareas, siendo la fase de planificación y las estrategias de representación de la información claves en este sentido. El empleo de este tipo de estrategias se corresponderían con el proceso de *"traducción"*, cuya relevancia en resolución de problemas matemáticos ha sido ampliamente demostrada (Abdullah, Zakaria, y Halim, 2012; Csíkos, Szitányi, y Keleme, 2012; Díez-Palomar, Menéndez, y Civil, 2011; Pantziara, Gagatsis, y Elia, 2009; Stylianou, 2011). Finalmente, el hecho de que las diferencias en el proceso fueran más marcadas en el segundo problemas, el cual implicaba una mayor número de relaciones y variables a manejar, y podría haber resultado más complejo para los estudiantes, sugerirían un papel más importante de las estrategias de planificación cuando las demandas de la tarea se incrementan. No obstante, el hecho de emplear únicamente dos problemas matemáticos no permite obtener una conclusión clara acerca de esta cuestión.

Segunda Cuestión

¿Puede el proceso informar acerca de los mecanismos implicados en la realización de juicios tras la ejecución de este tipo de tareas?

***Trabajo Complementario 1:* Elementary students´ metacognitive process and post-performance calibration in mathematical problem solving.**

Para este estudio se contó con la muestra completa de 524 estudiantes. Previo al análisis del proceso, en este estudio se estableció un índice de calibración y se analizó la estabilidad de los juicios y la ejecución real de los estudiantes en los dos problemas matemáticos. En primer lugar, y en la línea de previos estudios llevados a cabo en estos niveles educativos y superiores en resolución de problemas (Bol et al. 2005, Bol, Riggs, Hacker, y Nunnery, 2010; Hacker et al 2008b; Özsoy 2012), los resultados revelaron la existencia de unas pobres habilidades de calibración post-ejecución en la muestra, con la gran mayoría de los estudiantes mostrando un exceso de confianza en sus juicios. En torno al 60% de los estudiantes en el primer problema y 45% en el segundo juzgaron su ejecución de una forma positiva cuando el resultado fue negativo. Los estudiantes mostraron también un elevado grado de estabilidad tanto en sus juicios como en su ejecución real, aunque ligeramente mayor en el caso de los primeros. Este resultado apoyaría inicialmente los resultados obtenidos en trabajos previos como el de Bol et al. (2012), Dinsmore y Parkinson

(2013) o Stolp y Zabrucky (2009) entre otros. Estos sugieren además que sería la estabilidad de estos juicios –y de sus sesgos– uno de los aspectos que explicaría la conocida resistencia de la calibración a la mejora. No obstante, el hecho de emplear únicamente dos problemas impone limitaciones al alcance de estos resultados.

En cuanto al análisis del proceso metacognitivo llevado a cabo por los estudiantes que mostraron diferente precisión en sus juicios (Precisos vs. Imprecisos), los resultados indicaron la existencia de diferencias estadísticamente significativas entre los grupos en varios sub-procesos. Concretamente, los estudiantes que fueron precisos a la hora de juzgar su ejecución informaron acerca de un uso significativamente más frecuente de estrategias de organización y representación de la información que el grupo de estudiantes imprecisos. Este segundo grupo mostró en su lugar haber empleado una cantidad de tiempo significativamente mayor escribiendo una respuesta y revisando su ejecución que el primer grupo. Estos estudiantes también emplearon más tiempo corrigiendo errores, aunque estas diferencias no fueron estadísticamente significativas. Este patrón de resultados indicaría, al igual que ocurría en el primer estudio en cuanto al éxito en resolución de problemas, el uso de unas estrategias de planificación más eficaces por parte de los estudiantes precisos en sus juicios. Este resultado sugeriría la presencia de un mayor control metacognitivo en este grupo de estudiantes, al menos en cuanto a esta fase se refiere. Esto es consiste con lo expuesto por autores como Bol et al. (2012) o Dunlosky y Rawson (2012) en cuanto a la existencia de una importante relación entre unas buenas habilidades de autorregulación del aprendizaje y la calibración. Por su parte, las diferencias encontradas en los sub-procesos de escritura, revisión –y, aunque no estadísticamente significativas, en corrección– indicarían un empleo de estrategias de *"ensayo y error"* por parte de los estudiantes poco calibrados, las cuales han demostrado ser relativamente frecuentes en estudiantes en estas edades (Callander, 2011; Clements, Sarama, y Liu, 2008; Dijk, Van Oers, y Terwel, 2003).

Finalmente, este estudio ha mostrado la relevancia del rendimiento académico en la emergencia de diferentes niveles de calibración, así como de diferencias en el propio proceso metacognitivo durante la realización de los problemas matemáticos. Concretamente, y en la línea de estudios previos, un mayor rendimiento en matemáticas dio lugar a un incremento en la precisión de los juicios post-ejecución. Esta variable explicó también las diferencias previas en el empleo de estrategias de representación y organización de la información, encontradas entre los estudiantes con diferente precisión en sus juicios. Estos resultados confirman la importancia de tener en cuenta el papel del rendimiento académico cuando se estudia la calibración. Esta variable ha demostrado ser uno de los predictores más potentes en la precisión de este tipo de juicios, generando incluso diferencias en la eficacia de los programas de intervención en calibración (Hacker et al., 2000; 2008b; Nietfeld et al., 2006). La importancia de esta variable se ha visto confirmada así mismo en el último estudio que forma parte de esta Tesis Doctoral, cuyos resultados se discutirán más adelante.

Tercera Cuestión

¿Hay otras variables afectivas, motivacionales, cognitivas o conductuales que puedan explicar las diferencias en estos juicios?

En este punto, y una vez la relación entre el proceso de resolución de problemas y los juicios post-ejecución, los restantes cuatro trabajos de esta Tesis Doctoral tuvieron por objetivo realizar un análisis comprensivo de variables externas al proceso que pudieran explicar las diferencias en calibración.

Dos de estos estudios se centraron en el análisis de las funciones ejecutivas y un enfoque profundo de aprendizaje como posibles candidatos a explicar estas diferencias, partiendo del estudio de su relación con los componentes metacognitivos de conocimiento y habilidades (o aplicación de este conocimiento en situaciones de aprendizaje). Con carácter previo al estudio de la relación entre funciones ejecutivas y metacognición, se realizó un análisis de la utilidad de la escala BRIEF (Gioia et al., 2000) en la evaluación de las funciones ejecutivas, tendiendo como referencia una muestra clínica. En el cuarto y último de los estudios de esta sección se analizó el valor predictivo o discriminante de las variables que mostraron ser significativas en los dos estudios anteriores, junto con otros componentes más ampliamente estudiados (concretamente, el rendimiento académico en matemáticas, ciertos componentes afectivo-motivacionales y las características de la tarea), sobre la precisión de los juicios post-ejecución.

Artículo 2: **Psychometric characteristics of the BRIEF scale for the assessment of executive functions in Spanish clinical population.**

Los resultados obtenidos en este estudio indicaron la presencia de unas aceptables propiedades psicométricas en la escala, así como una estructura factorial similar a la mostrada por la escala original en Inglés. No obstante, una de las limitaciones de este instrumento reside en su longitud (un total de 86 ítems), lo cual hace difícil su aplicación, fundamentalmente en aquellos casos en que las familias deben cumplimentar un elevado número de cuestionarios, entrevistas, etc. En este sentido, disponer de una escala más breve facilitaría la labor de clínicos e investigadores; al menos como una medida de screening, a partir de la cual obtener información preliminar que oriente una evaluación más exhaustiva. Con este objetivo, se elaboraron las escalas de *Evaluación del Funcionamiento Ejecutivo* para familias y profesorado (García et al., 2013, 2014b), que suponen la base para dos de los estudios que se discuten a continuación.

***Artículo 3*: Metacognición y funcionamiento ejecutivo en Educación Primaria.**

Este estudio abordó el análisis de la relación entre funcionamiento ejecutivo y metacognición, partiendo del componente de conocimiento metacognitivo y su relación con las habilidades metacognitivas. Este trabajo se llevó a cabo con una muestra reducida de 114 estudiantes. Puesto que el funcionamiento ejecutivo se evaluó en base a las informaciones proporcionadas familias y profesorado, se contó incluyeron en este estudio solamente aquellos estudiantes que disponían de los cuestionarios cumplimentados por ambos informantes. Los estudiantes fueron divididos en dos grupos en función de su conocimiento metacognitivo (Alto vs. Bajo) y se analizaron las diferencias entre los grupos en habilidades metacognitivas (o aplicación de este conocimiento en las fases de Planificación, Ejecución y Evaluación del aprendizaje, evaluado mediante auto-informe) y de funcionamiento ejecutivo. Los resultados indicaron en primer lugar como un elevado conocimiento metacognitivo se relacionó con unas mejores habilidades metacognitivas, siendo estadísticamente significativas estas diferencias en las fases de Planificación y Ejecución del aprendizaje. Estos resultados siguen la línea de estudios previos como los llevados a cabo por Blöte et al. (2004), Pennequin et al. (2010) o Weil et al. (2013), los cuales evidencian la relevancia de un buen conocimiento metacognitivo como base para el desarrollo de las habilidades metacognitivas.

Un aspecto interesante en este sentido, es que si bien en esta ocasión se emplearon medidas de auto-informe para evaluar el comportamiento de los estudiantes en las tres fases del aprendizaje analizadas, los estudiantes informaron sobre un mayor uso de estrategias metacognitivas en la fase de Ejecución en comparación con las fases restantes. Este resultado podría estar relacionado con el hecho de que los estudiantes en estas etapas educativas (últimos años de Educación Primaria) tienden a dar una mayor prominencia a la ejecución de la tarea, frente a la elaboración de un plan previo y la evaluación del progreso y de los resultados obtenidos (Cleary y Chen, 2009; Kramarski y Gutman, 2006; Montague, Enders, y Dietz, 2011). Por tanto, es en cierto modo esperable que los estudiantes muestren un mejor control metacognitivo sobre los procesos de ejecución, en parte porque les resultan más familiares.

En cuanto a la relación entre funciones ejecutivas y metacognición, tanto familias como profesorado informaron acerca de unas habilidades significativamente mejores en atención sostenida, focalización de la atención, memoria funcional, organización y fundamentalmente planificación en el grupo de estudiantes con un alto conocimiento metacognitivo. No se encontraron diferencias estadísticamente significativas en flexibilidad cognitiva, inhibición de respuestas o control emocional. Este patrón de diferencias parece abogar a favor de la distinción propuesta por varios autores entre las denominadas funciones cálidas (hot) y frías (cool) (Brock, Rimm-Kaufman, y Nathanson, 2009; Zelazo y Carlson, 2012). Las funciones cálidas estarían implicadas en el tratamiento de la información emocional, incluyendo componentes como el control de impulsos y de las propias emociones, la interpretación de señales corporales, la toma de decisiones y el reconocimiento de la perspectiva del otro, y estarían más relacionadas con el desempeño social del individuo en su entorno; mientras que las funciones denominadas frías permitirían un tratamiento más racional de la información, y estarían relacionadas con el razonamiento y el procesamiento de información abstracta, como por ejemplo la memoria de trabajo, la planificación, conceptualización y categorización entre otras. Estas funciones estarían más relacionadas, por tanto, con aspectos del aprendizaje, y serían las que habrían generado diferencias estadísticamente significativas entre los grupos en el presente estudio. Este patrón de resultados y los elevados tamaños del efecto obtenidos llevaron a incorporar las funciones ejecutivas como posibles predictores de la calibración en el último estudio de esta Tesis Doctoral.

Trabajo Complementario 2: **Metacognitive knowledge and skills in students with Deep approach to learning. Evidence from mathematical problem solving.**

El cuarto estudio analizó las diferencias en conocimiento metacognitivo y habilidades metacognitivas –estas últimas evaluadas mediante la medida del proceso elaborada– entre tres grupos de estudiantes con diferente uso de un enfoque profundo de aprendizaje (Bajo-Medio-Alto). Se contó para ellos con la muestra completa de estudiantes.

Los resultados indicaron la existencia de diferencias estadísticamente significativas entre los grupos en el conocimiento metacognitivo, mostrando los estudiantes con un elevado uso del enfoque profundo un mejor conocimiento, lo cual es coherente desde con estudios previos que muestran la existencia de una asociación entre ambos componentes (Lazakidou y Retalis 2010; Malmberg, Jarvenoja, y Jarvela, 2013; McNamara, 2011; Rosário et al., 2013; Schmitz y Perel, 2012). No obstante, del análisis de las diferencias en habilidades metacognitivas no se desprendió un patrón de resultados consistente, encontrándose diferencias aisladas en el segundo problema, principalmente en el sub-proceso de revisión. Concretamente, un alto empleo de un enfoque profundo se relacionó con un mayor empleo de estas estrategias.

Esta ausencia de diferencias puede explicarse desde el punto de vista del conocimiento metacognitivo como la base para desarrollo de las habilidades metacognitivas. En este sentido, la muestra de estudiantes analizada mostró un bajo conocimiento metacognitivo en general, lo cual podría explicar la ausencia de unas adecuadas habilidades metacognitivas y la emergencia de diferencias en este sentido. No obstante, es necesario tener en cuenta en este punto las diferencias en el tipo de medidas empleadas en la evaluación del enfoque profundo de aprendizaje y el conocimiento metacognitivo, y la empleada para la evaluación de las habilidades metacognitivas y el grado de correspondencia entre las mismas. De este modo, son numerosos los estudios que coinciden en señalar la existencia de una baja convergencia entre las denominadas medias "off-line" –medidas tomadas antes o después de la ejecución de la tarea, principalmente cuestionarios– y medidas "on-line" –tomadas de forma concurrente con la tarea, como en este caso fueron evaluadas las habilidades metacognitivas– (Cromley y Azevedo, 2006; Desoete, 2008; Jacobse y Harskamp, 2012; Özcan, 2014; Saraç y Karakelle, 2012). Desde este punto de vista, es posible que la ausencia de diferencias importantes en las habilidades metacognitivas se haya debido en parte a la baja correspondencia entre diferentes tipos de medidas.

Una de las conclusiones de este estudio, y que no está relacionada directamente con los resultados obtenidos, ha sido la necesidad de considerar no solamente uno sino varios enfoques de aprendizaje dentro de un mismo individuo, los cuales pueden varias de acuerdo con factores tanto de la tarea, de la persona o del contexto (Aharony, 2006; Baeten et al., 2010; Chen y McNamee, 2011; Malmberg et al., 2013). El patrón de resultados obtenido, junto con la especificidad de la variable a predecir (es decir, juicios post-ejecución en resolución de problemas matemáticos) y el argumento anteriormente expuesto llevaron a desestimar el enfoque profundo de aprendizaje como posible variable predictora en la última parte de esta Tesis Doctoral.

***Trabajo Complementario 3:* Analysis of the factors involved in making post-performance judgments in mathematical problem solving.**

El último trabajo de esta Tesis Doctoral se realizó con una muestra de 188 estudiantes, asignados a dos grupos en función de la precisión de sus juicios post-ejecución en las dos tareas (Precisos vs. Imprecisos). Previo al análisis de su valor predictivo o discriminante, se analizaron las diferencias entre estos dos grupos de estudiantes en las siguientes variables: funciones ejecutivas, variables afectivo motivacionales relacionadas con las matemáticas y rendimiento en la asignatura, y características de la tarea –dificultad percibida y tiempo total empleado en resolver cada problema–. Los resultados mostraron la presencia de diferencias estadísticamente significativas entre los grupos en la mayoría de las variables analizadas:

En primer lugar, los estudiantes precisos en sus juicios mostraron ser significativamente menos impulsivos que sus compañeros en el otro grupo, mientras que mostraron unas mejores habilidades en planificación, memoria funcional, atención sostenida y focalización de la atención, evaluados estos componentes en base a las informaciones proporcionadas por las familias en este caso. Este patrón de diferencias recuerda a la distinción mencionada previamente en el tercer estudio de esta Tesis Doctoral entre las llamadas funciones ejecutivas cálidas (hot) y frías (cool) (Zelazo y Carlson, 2012), y la relación de estas últimas con la metacognición. No obstante, un dato interesante en este estudio es el hecho de que la inhibición de respuestas –que se incluiría dentro de las denominadas funciones cálidas– generó en este caso diferencias estadísticamente significativas entre los estudiantes con diferente precisión en sus juicios. En este sentido, un mejor nivel de control inhibitorio habría favorecido a los estudiantes, evitando respuestas impulsivas y permitiéndoles ser más reflexivos a la hora de elaborar sus juicios. Estos resultados son consistentes con la caracterización de las funciones ejecutivas como mecanismos de control de la conducta y del aprendizaje (García, González-Castro, Areces, Cueli, y Rodríguez, 2014b; Meltzer, 2013; Lyons y Zelano, 2011).

En segundo lugar, los estudiantes precisos en sus juicios presentaron asimismo unos sentimientos, creencias, motivaciones y actitudes significativamente más positivas en relación a las matemáticas que el otro grupo. Estos resultados son consistentes con los obtenidos por Narciss, Koerndle y Dresel (2011) y Sheldrake et al. (2014) entre otros en cuanto a la relación entre unas buenas habilidades de calibración y altos niveles de satisfacción, interés y percepción de auto-eficacia en el aprendizaje. Los estudiantes en este primer grupo mostraron también un rendimiento significativamente mejor en matemáticas, en la línea de los resultados obtenidos en el segundo estudio de esta Tesis Doctoral y de los trabajos previos mencionados con anterioridad.

En tercer lugar, la dificultad percibida de la tarea no generó diferencias estadísticamente significativas entre los grupos, mientras que el tiempo total empleado en los problemas demostró ser una variable importante, fundamentalmente en el caso del primer problema matemático. En este sentido, los estudiantes mejor calibrados emplearon más tiempo en la realización de la tarea que los estudiantes en el otro grupo, lo cual es coherente con las diferencias previamente mencionadas en cuando a una menor impulsividad en este primer grupo. En este sentido, un mejor control de la impulsividad habría ayudado a los estudiantes de este grupo, especialmente la primera vez que éstos se enfrentaron a la tarea. La relevancia del tiempo empleado en la tarea apoyaría en cierta medida los estudios previos que muestran como el propio hecho de verse envuelto en el proceso de realización de la tarea proporcionaría información útil para la formación de juicios a posteriori (Boekaerts y Rozendaal, 2010). Desde esta perspectiva sería comprensible que un mayor tiempo de exposición a la tarea se relacionara con unos juicios más precisos.

En cuanto al valor predictivo de estas variables sobre los juicios post-ejecución, el análisis discriminante llevado a cabo reveló el papel significativo de las variables rendimiento académico en matemáticas, utilidad percibida o valor de la asignatura, y el tiempo de resolución en el primer problema como las variables que predijeron significativamente la pertenencia de los estudiantes a su grupo. Estas variables permitieron asignar correctamente al 71.3% de la muestra. Un aspecto relevante en este sentido es el hecho de que ninguno de los componentes de las funciones ejecutivas que resultaron significativos en el análisis de diferencias se encontrase entre estas variables predictivas. Este resultado podría deberse a una elevada asociación entre estos componentes y otras de las variables que si demostraron ejercer un efecto significativo en esta segunda parte del análisis, presumiblemente el rendimiento en matemáticas. De hecho, la existencia de una elevada asociación entre las funciones ejecutivas y el rendimiento en esta área ha sido ampliamente demostrada en estudios previos (Bull y Lee, 2014; Thorell, Veleiro, Siu, y Mohammad, 2013). Finalmente, si bien estos resultados revelaron la conveniencia de analizar un amplio rango de variables cuando las diferencias en calibración quieren ser explicadas, es necesario considerar que las variables analizadas en este estudio permitieron clasificar correctamente al 71.3% de la muestra, siendo su valor predictivo mayor en el grupo de estudiantes calibrados. Esto indicaría que otras variables adicionales podrían explicar las diferencias en calibración, las cuales deben ser propiamente analizadas en futuros estudios.

Implicaciones

De los resultados obtenidos en esta Tesis Doctoral se derivan una serie de implicaciones en dos sentidos: en cuanto a la utilidad del procedimiento de evaluación diseñado para la evaluación del proceso, y en relación a la calibración post-ejecución y sus determinantes.

En cuanto a la medida del proceso

Los resultados obtenidos muestran la utilidad de la medida del proceso diseñada y puesta a prueba –*Triple Task Procedure in Mathematics (TTPM)*– en la evaluación de los procesos metacognitivos implicados en la resolución de problemas matemáticos. Una de las características principales de esta medida es que se basa en la introspección dirigida, lo cual la diferencia de otros métodos más conocidos como los protocolos Think-aloud. No obstante, y a pesar de las limitaciones que el empleo de esta técnica haya podido generar (las cuales se expondrán más adelante), el empleo de la introspección dirigida ha supuesto en cierto modo una ventaja, al proporcionar a los investigadores la posibilidad de acotar el número y tipología de categorías o sub-procesos a estudiar. Buena parte de la utilidad del propio método de evaluación diseñado reside en el sistema de categorías elaborado, basado en las fases del Aprendizaje Autorregulado (Zimmerman, 2000, 2008) y el modelo IDEAL de Bransford y Stein (1993).

Este sistema de categorías ha permitido en este caso subdividir un proceso tan complejo como la resolución de problemas matemáticos en sub-procesos o actividades aisladas y susceptibles de ser analizadas. Esto ha permitido obtener una evidencia inicial acerca de ciertas fortalezas y debilidades de los estudiantes en la realización de este tipo de tareas desde la perspectiva del control que estos ejercen sobre sus procesos cognitivos. Estos procesos o patrones de resolución de problemas se han relacionado con aspectos tan importantes como el resultado final –o producto– de la tarea y la precisión de los juicios realizados tras la misma. Disponer de este tipo de información es relevante tanto desde el punto de vista de la enseñanza como del aprendizaje de los estudiantes:

Desde el punto de vista de la enseñanza, la forma en la que el profesorado conceptualiza y organiza el conocimiento dentro de una materia o plantea una actividad concreta, así como el conocimiento que éste tiene acerca de los procesos seguidos por los estudiantes han demostrado ser importantes predictores tanto del rendimiento como de las habilidades de calibración en resolución de problemas y otros ámbitos. Este aspecto ha sido relevante en estudios llevados a cabo con muestras normativas de estudiantes pero especialmente en el caso de aquellos que incorporan muestras con dificultades del aprendizaje (Hattie, 2013; Hattie y Gan, 2011; Héfer, 2009; Maggioni y Parkinson, 2008). En esta línea, uno de los resultados obtenidos en la presente Tesis Doctoral ha sido la existencia de una elevada variabilidad entre los procesos metacognitivos mostrados por los estudiantes durante la resolución de los problemas matemáticos. Este resultado evidencia la necesidad de tener presentes las características y peculiaridades del alumnado, aproximándose por tanto a una perspectiva cada vez más centrada en el estudiante. Autores como Hattie (2013), y Maggioni y Parkinson (2008) sugieren el empleo de una serie de estrategias por parte de los docentes en este sentido, tales como anticipar a los estudiantes qué van a aprender o a qué tipo de actividad se van a enfrentar, proporcionar información clara sobre los criterios de éxito en la tarea, confirmar lo que los estudiantes ya saben o qué conocimientos o habilidades deben adquirir, explorar sus creencias, afectos y motivaciones, o proporcionar feed-back como una forma de reducir la fisura entre lo que un estudiante cree que sabe o entiende y lo que realmente sabe o entiende. Estrategias como estas permitirían al profesorado proporcionar un mejor "andamiaje" a los estudiantes en sus procesos de aprendizaje.

Desde la perspectiva del aprendizaje de los estudiantes, que éstos conozcan y sobretodo reflexionen sobre su propio proceso de aprendizaje es el primer paso para que aprendan a gestionarlo de una forma más efectiva, regulando sus procesos cognitivos, afectivos y motivacionales. La presencia de unas estrategias de planificación inefectivas, así como la ausencia de mecanismos de evaluación del progreso durante e incluso después de la realización de la tarea han sido precisamente dos de las principales dificultades mostradas por los estudiantes en los diferentes estudios realizados dentro de esta Tesis Doctoral. Los datos procedentes del análisis del perfil metacognitivo de los estudiantes y su baja calibración son muestra de ello y dan soporte a la afirmación realizada por Ifenthaler (2012) en cuanto a que el vínculo entre el conocimiento sobre la propia actividad de resolución de problemas y la regulación de la misma reside en buena parte en la capacidad de los estudiantes para reflexionar sobre dicha actividad.

En cuanto a la calibración post-ejecución y sus determinantes

Si bien el propio proceso metacognitivo mostrado por los estudiantes durante la resolución de problemas matemáticos ha mostrado cierto valor explicativo sobre los juicios post-ejecución, uno de los resultados principales de este trabajo de Tesis Doctoral es la necesidad de tener en cuenta un número amplio de variables cuando las diferencias en calibración quieren ser explicadas. En el caso concreto de la resolución de problemas matemáticos, estas variables explicativas responderían a lo que podrían denominarse diferentes niveles. En esta Tesis Doctoral se han establecido al menos tres niveles de análisis: un nivel más general en el que se situarían variables de tipo conductual como las funciones ejecutivas, un nivel intermedio en el que ciertas variables afectivo motivacionales relacionadas con las matemáticas y el rendimiento en la asignatura tendrían un papel relevante, y finalmente, un nivel más específico, centrado en el análisis de ciertas características de la tarea. Sin olvidar el posible papel explicativo de variables como la edad, el género o el curso.

Obviamente, existen infinidad de posibilidades de análisis y el rango de factores a analizar es inmenso. Los estudios realizados en este Tesis Doctoral supondrían una primera aproximación en este sentido y sugerirían la necesidad de seguir profundizando en el estudio de los factores determinantes de la calibración. Analizar estos factores es importante teniendo en cuenta que tanto los trabajos previos como los resultados obtenidos en esta Tesis Doctoral muestran que los estudiantes tienden a realizar juicios poco precisos sobre su ejecución. Estos juicios tienden además a ser estables, lo que ha demostrado suponer un reto para la intervención (Hacker et al., 2000, 2008a; Nietfeld et al., 2006). Si bien es común observar un exceso de confianza en estos juicios, como ha ocurrido en los estudios aquí llevados a cabo, la literatura sobre el tema sugiere que tanto el exceso como la falta de confianza pueden suponer condiciones negativas para el aprendizaje, afecto y motivaciones de los estudiantes (Bouffard et al., 2011; Gonida y Leondari 2011; Gramzow et al. 2003). Por lo tanto, no sería tanto el tipo de juicios sino su precisión el aspecto sobre el que sería necesario incidir en último término. El problema en torno en qué variables incidir en cada caso lo resolvería precisamente un análisis comprensivo de los factores determinantes de estos juicios.

Limitaciones y líneas futuras

Una vez discutidos los principales resultados obtenidos en esta Tesis Doctoral y su relevancia en el campo de estudio en que se circunscribe, se incluye una serie de limitaciones y líneas futuras de trabajo:

En primer lugar, uno de los aspectos a considerar es la existencia de unos bajos tamaños del efecto en algunos de los estudios llevados a cabo, lo cual sugiere la conveniencia de ser cautelosos al establecer el alcance de los resultados. Esto ocurre especialmente en el caso de los estudios basados en el análisis del proceso. La gran variabilidad en el proceso mostrado por los estudiantes, lo cual se tradujo en unas elevadas desviaciones típicas en las variables del proceso, podrían explicar en parte este resultado. Obtener grupos más homogéneos, quizás clasificando a los estudiantes en base a su perfil de resolución en ciertos sub-procesos o bien acotando el análisis de proceso a un número más reducido de momentos y no a la tarea completa. Los estudiantes mostraron de hecho una gran variabilidad en el tiempo total empleado en resolver los problemas matemáticos.

En segundo lugar, el empleo de únicamente dos problemas matemáticos en estos estudios debe ser tenido en cuenta. Esta decisión se debió principalmente a las propias características del método de evaluación del proceso empleado –*Triple Task Procedure in Mathematics (TTPM)*– el cual requiere de una fase previa de entrenamiento en el reconocimiento del sistema de categorías a analizar, y a las propias restricciones temporales que llevar a cabo todo el proceso de evaluación durante una única sesión supuso. Disponer de un mayor número de problemas habría sido de utilidad para analizar en más profundidad el proceso llevado a cabo por los estudiantes durante la resolución de problemas matemáticos. Si bien a la vista de los resultados parece que el segundo de los problemas habría sido capaz de mostrar un patrón de diferencias más preciso, incluir un mayor número de problemas, con diferentes grados de dificultad e incluso tipologías se considera un aspecto esencial para futuros estudios. En este línea, el trabajo realizado por Boekaerts y Rozendaal (2010) con estudiantes de quinto curso mostró como éstos obtuvieron un peor resultado en problemas basados en la vida real que en aquellos que requerían principalmente habilidades de computación, tendiendo a sobre-estimar su ejecución en mayor medida en los primeros.

Tercero, es también necesario considerar la naturaleza dicotómica de las variables rendimiento (Éxito vs. Fracaso) y juicios post-ejecución (Precisos vs. Imprecisos) empleadas en los estudios realizados. Esto se debió principalmente a las características de los problemas matemáticos empleados, que requerían un único resultado, a lo sumo dos en el segundo de los problemas. Este hecho, junto con la intención de proporcionar a los estudiantes un sistema de estimación de su propia ejecución más sencillo, llevó al empleo de medidas dicotómicas en la calibración post-ejecución. Si bien las medidas dicotómicas en calibración son frecuentemente empleadas, es necesario no olvidar que disponer de una medida más continua de esta variable aumentaría considerablemente las posibilidades de tratamiento de los datos.

En cuarto lugar, las características del protocolo de evaluación del proceso diseñado deben ser tenidas en consideración, concretamente la posibilidad de que la tarea de categorización, consistente en la presentación del sistema de categorías cada 40-45 pidiendo a los estudiantes señalar el sub-proceso que estaban realizando en cada momento, haya interrumpido su flujo de pensamiento, con efectos sobre el propio proceso y el resultado final en la tarea. Este posible efecto de reactividad se ha intentado reducir mediante un entrenamiento previo en el reconocimiento de las categorías empleadas, haciendo esta tarea lo más automática posible. No obstante, este aspecto debe ser más estrictamente controlado en futuros estudios ya sea empleando métodos de evaluación del procesos potencialmente menos intrusivos como medidas Think-aloud, o bien mediante el análisis de la medida del “esfuerzo cognitivo” en el procedimiento de la Triple Tarea (es decir, la diferencia en los tiempos de respuesta entre la línea base -tarea simple de tiempos de respuesta- y los tomados mientras los estudiantes resuelven los problemas matemáticos).

Finalmente, y no por ello menos importante, es necesario no olvidar que el proceso de Autorregulación del Aprendizaje se define como un proceso abierto y cíclico, en el cual hay lugar para diferentes relaciones entre procesos y fases, siendo el concepto de “recursividad” esencial. Definida como un movimiento de ida y vuelta o "estructuración retrospectiva y proyectiva" (Perl, 1994), ha sido ampliamente estudiada en composición escrita (Makalela, 2004; Olive et al., 2002; Paklans, 2008; Yood, 2005) pero no así en resolución de problemas. Si bien este aspecto no ha sido analizado en ninguno de los estudios expuestos en este trabajo, supone una de las principales líneas de continuación de esta Tesis Doctoral.

Síntesis de resultados

A continuación se presentan, de una forma resumida, los principales resultados obtenidos en los diferentes estudios llevados a cabo dentro de la presente Tesis Doctoral:

- El análisis del proceso muestra la presencia de unas estrategias de planificación ineficaces, así como una ausencia de mecanismos de evaluación, en la muestra general de estudiantes.
- Los estudiantes en la muestra general presentan una clara preferencia hacia sub-procesos familiares, como la realización de cálculos matemáticos, como medio para resolver los problemas.
- Este patrón de resolución de problemas se refleja en un bajo rendimiento en las tareas, así como en la formulación de juicios poco precisos acerca de la propia ejecución.
- Un buen rendimiento en las tareas de resolución de problemas se relaciona con unas estrategias de planificación más eficaces, siendo las estrategias de representación u organización de la información especialmente útiles.
- Los estudiantes en la muestra analizada muestran una baja calibración post-ejecución en tareas de resolución de problemas matemáticos, mostrando un exceso de confianza en sus juicios.
- Los estudiantes mejor calibrados usan unas estrategias de planificación más eficaces durante la realización de problemas matemáticos. Éstos muestran un uso más frecuente de estrategias de representación y organización de la información, mientras que los estudiantes no calibrados tienden a usar mecanismos de ensayo y error.
- Un mejor rendimiento académico en matemáticas se relaciona con unas mejores habilidades de calibración, y esta variable puede explicar las diferencias previas en el uso de estrategias de representación y organización de la información.
- Un mayor conocimiento metacognitivo (evaluado mediante un test de reconocimiento de estrategias) se asocia con unas mejores habilidades metacognitivas (evaluadas mediante auto-informe), y de funcionamiento ejecutivo (evaluadas a través de las informaciones proporcionadas por familias y profesorado).

- Un alto conocimiento metacognitivo se relaciona con mejores niveles en organización, planificación, procesos atencionales y memorísticos; componentes que formarían parte de las denominadas funciones ejecutivas "frías".
- El empleo de un enfoque profundo de aprendizaje (evaluado mediante auto-informe) se relaciona con un mejor conocimiento metacognitivo (evaluado mediante un test de reconocimiento de estrategias). No se encuentra, no obstante, una clara relación entre este enfoque y las habilidades metacognitivas, evaluadas mediante la administración de la medida del proceso diseñada en tareas de resolución de problemas. Estos resultados plantean la cuestión en torno a la distinción entre diferentes tipos de medidas (off-line vs. on-line) y su grado de correspondencia.
- La adopción de un determinado enfoque de aprendizaje varía dependiendo de las circunstancias personales, contextuales y de la tarea. Se deben por tanto analizar no un único sino varios enfoques de aprendizaje en la misma persona, con el fin de conocer su influencia en cualquier proceso cognitivo.
- Los estudiantes precisos en sus juicios muestran unas mejores creencias, actitudes y motivaciones hacia las matemáticas, siendo el valor o utilidad percibida de la asignatura una de las variables que significativamente predicen la calibración post-ejecución.
- Unas mejores habilidades en las denominadas funciones ejecutivas "frías" y un mayor control inhibitorio se asocian con mejores niveles de calibración post-ejecución.
- Un mayor tiempo empleado en resolver los problemas matemáticos se asocia con la formulación de juicios más precisos tras la ejecución de la tarea. Este resultado es coherente con la presencia de un mayor control de la impulsividad en los estudiantes precisos en sus juicios.
- El rendimiento académico en matemáticas, el valor o utilidad percibida de la asignatura, y el tiempo total empleado en resolver el primero de los problemas matemáticos, demuestran tener un importante valor predictivo sobre los juicios post-ejecución en la presente muestra de estudiantes.
- Las variables edad, género y curso no ejercen un efecto significativo sobre la calibración post-ejecución de los estudiantes evaluados.
- Los componentes de las funciones ejecutivas no predicen la calibración post-ejecución en la presente muestra, debido posiblemente a la relación entre estas variables y el rendimiento académico en matemáticas.

Conclusiones

Una vez presentados y discutidos los resultados obtenidos en los diferentes estudios que forman parte de esta Tesis Doctoral, se pueden formular las siguientes conclusiones:

1) *El procedimiento de evaluación del proceso diseñado y puesto a prueba -Triple Task Procedure in Mathematics- resulta de utilidad en el estudio del proceso metacognitivo implicado en la resolución de problemas matemáticos.*

   Este método proporciona información acerca del perfil del proceso mostrado por los estudiantes, posibilitando el establecimiento de una relación entre proceso y producto (o resultado de la ejecución), así como entre proceso y calibración post-ejecución.

   La aplicación de este método de evaluación ayuda a identificar ciertas fortalezas y debilidades en los patrones de resolución de problemas de los estudiantes, lo cual es especialmente relevante desde el punto de vista de los procesos de Enseñanza/Aprendizaje. Tanto docentes como el propio alumnado podrían beneficiarse de este tipo de información para introducir mejoras en este ámbito, favoreciendo entre los estudiantes la adopción de grados crecientes de control sobre su propio proceso de aprendizaje.

2) *Es necesario tener en cuenta las características individuales del alumnado con el fin de promover mejoras en sus habilidades de resolución de problemas matemáticos.*

   La gran variabilidad en los procesos mostrados por los estudiantes durante este tipo de tareas sugiere la necesidad de tener presentes las características individuales observadas en dicho proceso. Las diferencias en el rendimiento previo en matemáticas suponen en este mismo sentido una importante variable a considerar. Diferentes niveles de rendimiento pueden relacionarse con patrones diferenciales de resolución de problemas, e incluso con diferencias en la precisión de los juicios realizados por los estudiantes sobre su ejecución. La intervención para la mejora de las habilidades de resolución de problemas y la calibración ha demostrado depender en buena medida de esta variable en numerosos trabajos previos. Es importante en este sentido prestar una especial atención a los estudiantes con bajos niveles de rendimiento, teniendo en cuenta desde las características de sus procesos, hasta sus motivos y preferencias hacia el empleo de determinados procedimientos durante la realización de la tarea.

3) *La calibración es un proceso metacognitivo complejo que depende de múltiples factores, los cuales deben ser analizados desde un punto de vista comprensivo.*

   Es necesario estudiar un amplio rango de variables con el fin de obtener una visión lo más comprensiva posible sobre los factores que influyen en la formación y precisión de este tipo de juicios. Estos factores deben responder asimismo a diferentes niveles de análisis, desde los aspectos más generales como las habilidades cognitivas o de control del comportamiento, hasta los más específicos como las características de la tarea. Es importante también tener en cuenta como diferentes variables, incluso pertenecientes a diferentes niveles, interactúan entre si. Un mejor conocimiento de estos factores y sus relaciones permitirá diseñar estrategias de intervención potencialmente más eficaces.

4) *Finalmente, ciertas características del procedimiento de evaluación del proceso diseñado y del diseño de la investigación deben ser tenidas en cuenta en la interpretación de los resultados obtenidos, así como en futuros estudios en esta línea.*

   Es posible que el tipo de medida de evaluación del proceso empleada, tomada de forma concurrente a la realización de la tarea y basada en la introspección dirigida, pueda ejercer un efecto en el propio proceso de los estudiantes (es decir, resulte reactiva). Del mismo modo, la posible recursividad del proceso de resolución de problemas debe ser analizada en futuros estudios. Finalmente, el número, tipología y dificultad de los problemas matemáticos empleados, así como la naturaleza dicotómica de algunas de las variables estudiadas, son deben ser tenidas en cuenta como posibles limitaciones en estos trabajos.

Conclusions

Having presented and discussed the results of the different studies conducted within this PhD-Thesis, the following prime conclusions from the research can be summarised:

1) *The method of assessing the process that was designed and tested (the Triple Task Procedure in Mathematics) was found to be useful in the study of the metacognitive processes involved in solving mathematical problems.*

   This method provides evidence about the profile of the process shown by students, enabling the establishment of a relationship between process and product (or task-result), as well as between process and post-performance calibration.

   The application of this assessment method helps to identify certain strengths and weaknesses in student problem-solving activities that are especially relevant from the perspective of Teaching-Learning processes. Both teachers and students themselves could benefit from this type of information by utilising it to attain improvements in this area, which in turn, would also allow students to adopt increasing degrees of control over their own learning processes.

2) *It is also essential to take into consideration a student's individual characteristics in order to fully promote improvements in their ability to solve mathematical problems.*

   The great variability shown by students while solving mathematical problems suggests that there is a need to pay attention to the individual characteristics observed during the process. Differences in mathematics achievement are also an important variable to consider in this context. Different achievement levels may be related to differential patterns of problem solving, and even to differences in the accuracy of performance judgments made by students. Previous research has shown that an important part of the effectiveness of interventions aimed at improving problem-solving skills and calibration depends largely upon this variable. Moreover, it is important to pay special attention to students with low levels of academic achievement in this area by addressing different variables in the analysis, from the characteristics of student problem-solving processes to their motives as well as their preferences for the use of certain procedures while performing the task.

3) *Calibration is a complex process that relies on multiple factors, and thus must be examined from a comprehensive analysis perspective.*

   It is necessary to study a wide range of variables in order to obtain a complete picture of the factors that influence the formation and accuracy of judgments. These factors must also correspond to different levels of analysis, from general aspects such as cognitive abilities or behavioral control, to more specific such as the characteristics of the task. It is also important to note how different variables (even those restricted to different levels) interact with one another. A better understanding of these factors and their interrelationships will allow us to design intervention strategies that are potentially more effective.

4) *Finally, certain aspects of the assessment procedure and the research design must be taken into consideration before the results obtained in the different studies included in this Dissertation can be fully interpreted and applied to future research on this issue.*

   It is possible that the type of assessment measure used, taken concurrently with the completion of the task and based on introspection directed, have an effect on the process itself of (i.e., resulting reactive). Similarly, the possible recursion in problem-solving processes must also be analyzed in future studies. Finally, the number, type and difficulty of the mathematical problems used in these studies, as well as the dichotomous nature of some of the variables, need to be taken into account as potential limitations.

Referencias bibliográficas

Abdullah, N., Zakaria, E. y Halim, L. (2012). The effect of a thinking strategy Approach through visual representation on achievement and conceptual understanding in solving mathematical word problems. *Asian Social Science, 8*(16), 16-34. doi: 10.5539/ass.v8n16p30

Adelson, J. L. y McCoach, D. B. (2011). Development and psychometric properties of the Math and Me Survey: Measuring third through sixth graders´ attitudes toward Mathematics. *Measurement & Evaluation in Counseling & Development, 44*(4), 225-247. doi: 10.1177/0748175611418522

Aharony, N. (2006). The use of deep and surface learning strategies among students learning English as a foreign language in an Internet environment. *British Journal of Educational Psychology, 76*(4), 851-866. doi: 10.1348/000709905X79158

Alexander, P. A. (2013). Calibration: What is it and why it matters? An introduction to the special issue on calibrating calibration. *Learning & Instruction, 24*, 1-3, doi: 10.1016/j.learninstruc.2012.10.003

Anderson, V. M., Jacobs, R. y Anderson, P. (2008). *Executive functions and the frontal lobes: A lifespan perspective.* New York, NY: Taylor y Francis.

Azevedo R. y Cromley, J.G. (2004) Does training of self- regulated learning facilitate student's learning with hypermedia? *Journal of Educational Psychology 96*, 523-535.

Baeten, M., Kyndt, E., Struyven, K. y Dochy, F. (2010). Using student-centred learning environments to stimulate deep approaches to learning: Factors encouraging or discouraging their effectiveness. *Educational Research Review, 5*(3), 243-260. doi: 10.1016/j.edurev. 2010.06.001

Blanco, L. J., Guerrero, E. y Caballero, A. (2013). Cognition and affect in Mathematics problem solving with prospective teachers. *The Mathematics Enthusiast, 10*(1-2), 335-364.

Blöte, A. W., Van Otterloo, S. G., van Stevenson, C. E. y Veenman, M. V. J. (2004). Discovery and maintenance of the many-to-one counting strategy in 4-year-olds: A microgenetic study. *British Journal of Developmental Psychology, 22,* 83–102.

Boekaerts, M. y 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. doi: 10.1016/j.learninstruc.2009.03.002

Bol, L., Hacker D. J., O'Shea P. y Allen D. (2005). The influence of overt practice, achievement level, and explanatory style on calibration accuracy and performance. *Journal of Experimental Education*, *73*, 269-290. doi: 10.3200/JEXE.73.4.269-290

Bol, L., Hacker, D. J., Walck, C. C. y Nunnery, J. A. (2012). The effects of individual or group guidelines on the calibration accuracy and achievement of high school biology students. *Contemporary Educational Psychology*, *37*, 280-287. doi: 10.1016/j.cedpsych.2012.02.004

Bol, L., Riggs R., Hacker D. J. y Nunnery J. (2010). The calibration accuracy of middle school students in math classes. *Journal of Research in Education*, *21*, 81-96.

Bouffard, T., Vezeau, C., Roy, M. y Lengelé, A. (2011). Stability of biases in self-evaluation and relations to well-being among elementary school children. *International Journal of Educational Research*, *50,* 221-229. doi: 10.1016/j.ijer.2011.08.003

Bransford, J. D. y Stein, B. S. (1993). *The ideal problem solver: A guide for improving thinking, learning and creativity* (2nd Ed.). New York, NY: W. H. Freeman.

Brock, L. L., Rimm-Kaufman, S. E. y Nathanson, L. (2009). The contributions of 'hot' and 'cool' executive function to children's academic achievement and learning-related behaviors, and engagement in kindergarten. *Early Childhood Research Quarterly*, *24*(3), 337-349.

Bull, E. y Lee, K. (2014). Executive functioning and mathematics achievement. *Child Development Perspectives, 8*(1), 36-41. doi: 10.1111/cdep.12059

Butler D. L. y Cartier S.C. (2005). *Multiple complementary methods for understanding self-regulated learning as situated in context.* Accepted for presentation at the April 2005 Annual Meetings of the American Educational Research Association, Montreal, QC.

Callander, S. (2011). Searching and learning by trial and error. *American Economic Review, 101*(6), 2277-2308. doi: 10.1257/aer.101.6.2277

Castro, E. (2008). Resolución de problemas. Ideas, tendencias e influencias en España. Castro, E. (2008). Resolución de problemas: ideas, tendencias e influencias en España. En R. Luengo, B. Gómez, M. Camacho y L. Blanco (Eds.), *Investigación en educación matemática XII* (pp. 113-140). Badajoz: Sociedad Española de Investigación en Educación Matemática, SEIEM.

Chen, L. G. y McNamee, G. (2011). Positive approaches to learning in the context of Preschool classroom activities. *Early Childhood Education Journal*, *39*(1), 71-78. doi: 10.1007.s10643-010-0441-x

Cleary, T. J. (2009). Monitoring trends and accuracy of self-efficacy beliefs during interventions: Advantages and potential applications to school-based settings. *Psychology in the Schools*, *46*, 154-171. doi: 10.1002/pits.20360

Cleary, T. J. y Chen, P. (2009). Self-regulation, motivation, and math achievement in middle school: Variations across grade level and math context. *Journal of School Psychology, 47,* 291-314. doi: 10.1016/j.jsp.2009.04.002

Clements, D. H., Sarama, J. y Liu, X. (2008). Development of a measure of early mathematics achievement using the Rasch model: The Research-based Early Maths Assessment. *Educational Psychology, 28*(4), 457-482. doi: 10.1080/01443410701777272

Cromley, J. G. y Azevedo, R. (2006). Self-report of reading comprehension strategies: What are we measuring? *Metacognition and Learning, 1,* 229-247.

Csíkos, C., Szitányi, J. y Keleme, R. (2012). The effects of using drawings in developing young children's mathematical word problem solving: A design experiment with third-grade Hungarian students. *Educational Studies in Mathematics*, *81,* 47-65. doi: 10.1007/s10649-011-9360-z

Das, R. y Das, G. C. (2013). Math anxiety: The poor problem-solving factor in school mathematics. *International Journal of Scientific and Research Publications, 3*(4), 1-5.

Desoete, A. y Roeyers, H. (2006). Metacognitive macro-evaluations in mathematical problem solving. *Learning and Instruction, 16,* 12-25.

Desoete, A. (2008). Multi-method assessment of metacognitive skills in elementary school children: how you test is what you get. *Metacognition and Learning, 3*(3), 189‒206.

Díez-Palomar, J., Menéndez, J. M. y Civil, M. (2011). Learning mathematics with adult learners: drawing from parents´ perspective. *Revista Latinoamericana de Investigación en Matemática Educativa, 14*(1), 71-94.

Dijk, I. M. A. W., Van Oers, B. y Terwel, J. (2003). Providing or designing? Constructing models in primary maths education. *Learning and Instruction, 13*, 53-72. doi: 10.1016/S0959-4752(01)00037-8

Dinsmore, D. L. y Parkinson, M. M. (2013). What are confidence judgments made of? Students' explanations for their confidence ratings and what that means for calibration. *Learning and Instruction, 24,* 4-14. doi: 10.1016/j.learninstruc.2012.06.001

Dunlosky, J. y Rawson, K. A. (2012). Over-confidence produces underachievement: Inaccurate self-evaluations undermine students' learning and retention. *Learning and Instruction, 22,* 271-280. doi: 10.1016/j.learninstruc.2011.08.003

Dupeyrat, C., Escribe, C., Huet, N. y Régner, I. (2011). Positive biases in self-assessment of mathematics competence, achievement goals, and mathematics performance. *International Journal of Educational Research, 50*(4), 241-250.

Efklides, A. y Misailidi, P. (2010). *Trends and prospects in metacognition research*. New York, NY: Springer.

Farré, A. y Narbona, J. (1997). *Escala de Défi cit de Atención e Hiperactividad (E.D.A.H.)* Madrid: TEA Ediciones.

Gálvez, G., Cosmelli, D., Cubillos, L., Leger, A. M., Mena, A., Tanter, E., … y Soto-Andrade, J. (2011). Estrategias cognitivas para el cálculo mental. *Revista Latinoamericana de Investigación en Matemática Educativa, 14*(1), 9-40.

García, T., Álvarez-García, D., Cueli, M., González-Castro, P. y Álvarez, L. (2013). Propiedades psicométricas de la Escala de Funcionamiento Ejecutivo para Familias. *En revisión.*

García, T., Álvarez-García, D., González-Castro, P., Álvarez, L. y Segurola, L. A. (2014a). Propiedades psicométricas de la Escala de Funcionamiento Ejecutivo para Profesorado. *Revista de Psicología y Educación, 9*(1), 49-69.

García, T., González-Castro, P., Areces, D., Cueli, M. y Rodríguez, C. (2014b). Executive functions in children and adolescents: the types of assessment measures used and implications for their validity in clinical and educational contexts. *Papeles del Psicólogo, 35*(3), 215-223.

Geary, D. C. (2004). Mathematics and learning disabilities. *Journal of Learning Disabilities, 37,* 4-15.

Gick, M. L. (1986). Problem-solving strategies. *Educational Psychologist, 21*, 99-120.

Gioia, G.A., Isquith, P.K., Guy, S. y Kenworthy, L. (2000). *BRIEF: Behavior Rating Inventory of Executive Function professional manual.* Lutz, FL: Psychological Assessment Resources.

Gonida, E. N. y Leondari, A. (2011). Patterns of motivation among adolescents with biased and accurate self-efficacy beliefs. *International Journal of Educational Research, 50*(4), 209-211. doi: 10.1016/j.ijer.2011.08.002.

Gramzow, R. H., Elliot, A. J., Asher, E. y McGregor, H. A. (2003). Self-evaluation bias and academic performance: Some ways and some reasons why. *Journal of Research in Personality*, *37*, 41-61.

Jarero, M., Aparicio, E. y Sosa, L. (2013). Pruebas escritas como estrategia de evaluación de aprendizajes matemáticos. Un estudio de caso a nivel superior. *Revista Latinoamericana de Investigación en Matemática Educativa, 16* (2), 213-243.doi: 10.12802/relime.13.1623

Hacker D. J., Bol L. y Bahbahadi, K. (2008b). Explaining calibration accuracy in classroom contexts: The effects of incentives, reflection, and explanatory style. *Metacognition Learning, 3,* 101-121. doi: 10.1007/s11409-008-9021-5

Hacker, D. J., Bol, L., Horgan, D. D. y Rakow, E. A. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology*, *92,* 160-170.

Hacker, D. J., Bol, L. y Keener, M. C. (2008a). Metacognition in education: A focus on calibration. En J. Dunlosky y R. Bjork (Eds.), *Handbook of memory and metacognition* (pp. 429-455). Mahwah, NJ: Erlbaum.

Hadwin, A. F. y Webster, E. A. (2013). Calibration in goal setting: Examining the nature of judgments of confidence. *Learning and Instruction, 24*(12), 37-47. doi: 10.1016/j.learninstruc.2012.10.001

Hattie, J. (2013). Calibration and confidence: Where to next? *Learning & Instruction, 24,* 62-66. doi: 10.1016/j.learninstruc.2012.05.009

Hattie, J. y Gan, M. (2011). Instruction based on feedback. En R. E. Mayer y P. Alexander (Eds.). *Handbook of research on learning and instruction* (pp. 249-271). New York, NY: Routledge.

Harskamp, E. y Suhre, C. (2006). Improving mathematical problem solving: A computerized approach. *Computers in Human Behavior*, *22*(5), 801-815. doi: 10.1016/j.chb.2004.03.023

Héfer, B. (2009). Three essential components of college teaching: Achievement calibration, self-efficacy, and self-regulation. *College Student Journal, 43*(2), 562-570.

Ifenthaler, D. (2012). Determining the effectiveness of prompts for self-regulated learning in problem-solving scenarios. *Educational Technology & Society, 15*(1), 38-52.

Jacobse, A. E. y Harskamp, E. G. (2012). Towards efficient measurement of metacognition in mathematical problem-solving. *Metacognition and Learning, 7*(2), 133-149.

Kajamies, A. Vauras, M. y Kinnunen, R. (2010). Instructing low-achievers in mathematical word problem solving. *Scandinavian Journal of Educational Research, 54*(4), 335-355. doi: 10.1080/00313831.2010.493341

Kellog, R. T. (1987). Writing performance: Effects of cognitive strategies. *Written Communication, 4*, 269-298.

Kramarski, B. y Gutman, M. (2006). How can self-regulated learning be supported in mathematical E-learning environments? *Journal of Computer Assisted Learning 22,* 24-33. doi: 10.1111/j.1365-2729.2006.00157.x

Lazakidou, G. y Retalis, S. (2010). Using computer supported collaborative learning strategies for helping students acquire self-regulated problem-solving skills in mathematics. *Computers & Education 54,* 3-13. doi: 10.1016/j.compedu.2009.02.020

Lyons, K. E. y Zelazo, P. D. (2011). Monitoring, metacognition, and executive function: elucidating the role of self-reflection in the development of self-regulation. En J. Benson (Ed.), *Advances in Child Development and Behavior* (pp.379-412). Burlington: Academic Press.

Makalela, L. (2004). Differential error types in second-language students´written and spoken texts: Implications for Instruction in writing. *Written Communication, 21,* 368. doi: 10.1177/0741088304270026

Maggioni, L. y Parkinson, M. M. (2008). The role of teacher epistemic cognition, epistemic beliefs, and calibration in instruction. *Educational Psychology Review, 20*(4), 445-462. doi: 10.1007/s10648-008-9081-8

Malmberg, J., Jarvenoja, H. y Jarvela, S. (2013). Patterns in elementary school students' strategic actions in varying learning situations. *Instructional Science, 41*(5), 933-954. doi: 10.1007/s11251-012-9262-1

McCormick, C. B. (2003). Metacognition and learning. En W. B. Reynolds y G. E. Miller (Eds.), *Handbook of psychology: Educational psychology* (pp. 79-102). Hoboken, NJ: Wiley.

McNamara, D. S. (2011). Measuring deep, reflective comprehension and learning strategies: challenges and successes. *Metacognition and Learning, 6*(2), 195-203. doi: 10.1007/s11409-011-9082-8

Meltzer, L. (2013). Executive function and metacognition in students with learning disabilities: New Approaches to Assessment and Intervention. *International Journal for Research in Learning Disabilities, 1*(2), 31-63.

Montague, M. (2008). Self-Regulation and mathematics instruction. *Learning Disabilities Research & Practice, 22(*1), 75-83.

Montague, M., Enders, G. y 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. doi: 10.1177/0731948711421762

Nietfeld, J. L., Cao, L. y Osborne, J. W. (2005). Metacognitive monitoring accuracy and student performance in the postsecondary classroom. *The Journal of Experimental Education, 74,* 7-28.

Nietfeld, J. L., Cao, L. y Osborne, J. W. (2006). The effect of distributed monitoring exercises and feedback on performance, monitoring accuracy, and self-efficacy. *Metacognition and Learning, 1,* 159-179.

Núñez, J. C., Cerezo, R., Bernardo, A., Rosário, P., Valle, A., Fernández, E. y Suárez, N. (2011). Implementation of training programs in self-regulated learning strategies in moodle format: Results of an experience in higher education. *Psicothema, 23*(2), 274-281.

Núñez, J. C., Solano, P., González-Pienda, J. A. y Rosário, P. (2006). El aprendizaje autorregulado como medio y meta de la educación. *Papeles del Psicólogo, 27*(3), 139-146.

OECD. (2003). *The PISA 2003 Assessment Framework: Mathematics, Reading, Science and Problem Solving Knowledge and Skills*. PISA: OECD Publishing.

OECD. (2014). *PISA 2012 Results: Creative Problem Solving: Students' Skills in Tackling Real-Life Problems (Volume V)*. PISA: OECD Publishing.

Olive, T., Kellog, R. T. y Piolat, A. (2001). The Triple Task technique for studying the process of writing. En T. Olive. y C. M. Levy (Eds.), *Contemporary tools and techniques for studying writing* (pp. 89-104). Dordecht: Kluwer Academic Publishers.

Olive, T. y Piolat, A. (2002). Suppressing visual feedback in written composition: Effects on processing demands and coordination of the writing processes. *International Journal of Psychology, 37*(4), 209-218. doi: 10.1080/00207590244000089

Özcan, C. Z. (2014). Assessment of metacognition in mathematics: which one of two methods is a better predictor of mathematics achievement? *International Online Journal of Educational Sciences, 6*(1), 49-57. doi: 10.15345/iojes.2014.01.006

Özsoy, G. (2012). Investigation of fifth grade students' mathematical calibration skills. *Educational Sciences: Theory & Practice, 12*(2), 1190-1194.

Pantziara, M., Gagatsis, A. y Elia, I. (2009). Using diagrams as tools for the solution of non-routine mathematical problems. *Educational Studies in Mathematics, 72,* 39-60. doi: 10.1007/s10649-009-9181-5

Pennequin, V., Sorel, O., Nanty, I. y 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 and Reasoning, 16*(3), 198-220. doi: 10.1080/13546783.2010.509052

Pereis, F., Dignath, C. y Schmitz, B. (2009). Is it possible to improve mathematical achievement by means of self-regulation strategies? Evaluation of an intervention in regular math classes. *European Journal of Psychology of Education, 24*(1), 17-29. doi: 10.1007/BF03173472

Perl, S. (1994). *Landmark Essays on Writing Process.* Davis, CA: Hermagoras.

Phan, H. P. (2011). Deep processing strategies and critical thinking: Developmental trajectories using latent growth analyses. *Journal of Educational Research, 104*(4), 283-294. doi: 10.1080/00220671003739382

Plakans, L. (2008). Comparing composing processes in writing-only and reading-to-write test tasks. *Assessing Writing*, *12*(2), 111-129. doi: 10.1016/j.asw.2008.07.001

Ramirez, G., Gunderson, E. A., Levine, S. C. y Beilock, S. L. (2013). Math anxiety, working memory, and math achievement in early elementary school. *Journal of Cognition and Development, 14*(2), 187-202. doi: 10.1080/15248372.2012.664593

Ranellucci, J., Muis, K.R., Duffy, M., Wang, X., Sampasivam, L. y Franco, G. (2013). To master or perform? Exploring relations between achievement goals and conceptual change learning. *British Journal of Educational Psychology, 83*(3), 431-451. doi: 10.1111/j.2044-8279.2012.02072.x

Rinne, L. F. y Mazzocco, M. M. M. (2014). Knowing right from wrong in mental arithmetic judgments: Calibration of confidence predicts the development of accuracy. *PLoS ONE 9*(7), e98663. doi: 10.1371/journal.pone.0098663

Rosário, P. (2004). *Estudar o Estudar: As (Des)venturas do Testas*. Porto: Porto Editora.

Rosário, P., González-Pienda, J. A., Pinto, R., Ferreira, P., Lourenço, A. y Paiva, O. (2010). Efficacy of the program "Testas's (mis)adventures" to promote the deep approach to learning. *Psicothema*, *22*, 828-834.

Rosário, P., Mourão, R., Núñez, J. C., González-Pienda, J. A. y Solano, P. (2006). School-family: Is a reciprocal and positive relationship possible? *Papeles del Psicólogo, 27*(3), 174-182.

Rosário, P., Núñez, J. C., Ferrando, P., Paiva, O., Lourenço, A., Cerezo, R. y Valle, A. (2013). The relationship between approaches to teaching and approaches to studying: A two-level structural equation model for biology achievement in high school. *Metacognition and Learning, 8*(1), 44-77. doi: 10.1007/s11409-013-9095-6

Sakiz, G., Pape, S. J. y Woolfolk, A. (2012). Does perceived teacher affective support matter for middle school students in mathematics classrooms? *Journal of School Psychology, 50,* 235-255. doi: 10.1016/j.jsp.2011.10.005

Santos, M. (2007). *La Resolución de Problemas matemáticos. Fundamentos cognitivos*. México: Trillas.

Saraç, S. y Karakelle, S. (2012). On-line and off-line assessment of metacognition. *International Electronic Journal of Elementary Education, 4(2), 301-315*

Schunk, D. y Pajares, F. (2009). Self-efficacy theory. En K. R. Wentzel y A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 35-53). New York, NY: Routledge.

Sheldrake, R., Mujtaba, T. y 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. doi: 10.1016/j.ijer.2013.10.008

Schmitz, B. y Perel, P. (2011). Self-monitoring of self-regulation during math homework behaviour using standardized diaries. *Metacognition and Learning, 6*, 255-273. doi: 10.1007/s11409-011-9076-6

Stacey, K. (2005). The place of problem solving in contemporary mathematics curriculum documents. *Journal of Mathematical Behaviour, 24*, 341–350.

Stein, M. K., Remillard, J. y Smith, M. S. (2007). How curriculum influences student learning. En F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning: a project of the National Council of Teachers of Mathematics* (pp. 319-369). Charlotte, NC: Information Age Publishing.

Stolp, S., y Zabrucky, 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.

Stylianou, D. A. (2011). An examination of middle school students' representation practices in mathematical problem solving through the lens of expert work: towards an organizing scheme. *Educational Studies in Mathematics, 76*, 265-280. doi: 10.1007/s10649-010-9273-2

Thorell, L. B., Veleiro, A., Siu, A. F. y Mohammadi, H. (2013). Examining the relation between ratings of executive functioning and academic achievement: findings from a cross-cultural study. *Child Neuropsychology, 19*(6), 630-8. doi: 10.1080/09297049.2012.727792

Veenman, M. V. J. (2011). Learning to self-monitor and self-regulate. En R. Mayer y P. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 197-218). New York, NY: Routledge.

Verschaffel, L., Greer, B. y De Corte, E. (2000). *Making sense of word problems*. Lisse: Swets & Zeitlinger.

Waters, S. K., Cross, D. S. y Runions, K. (2009). Social and ecological structures supporting adolescent connectedness to school: A theoretical model. *Journal of School Health, 79*(11), 516-524.

Weil, L. G., Flemingc, S. M., Dumontheil, I., Kilforda, E. J., Weil, R. S., Rees, G., … y Blakemorea, S. (2013). The development of metacognitive ability in adolescence. *Consciousness and Cognition, 22*, 264–271. doi: 10.1016/j.concog.2013.01.004

Whimbey, A. y Lochhead, J. (1993). *Comprender y resolver problemas.* Madrid: Visor.

Yood, J. (2005). Present-Process: The Composition of Change. *Journal of Basic Writing, 24*(2), 4-25.

Zelazo, P. D. y Carlson, S. M. (2012). Hot and cool executive function in childhood and adolescence: Development and Plasticity. *Child Development Perspectives, 6(*4), 354-360. doi: 10.1111/j.1750-8606.2012.00246.x

Zhu, Z. (2007). Gender differences in mathematical problem solving patterns: a review of literature. *International Education Journal*, *8*(2), 187-203.

Zimmerman, B. (2000). Attaining self-regulation. A social cognitive perspective. En M. Boekaerts, P. R. Pintrich y M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13-39). San Diego, CA: Academic Press.

Zimmerman, B. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal*, *45*(1), 166-183. doi: 10.3102/0002831207312909

Zimmerman, B. J., Moylan, A., Hudesman, J., White, N. y Flugman, B. (2011). Enhancing self-reflection and mathematics achievement of at-risk urban technical college students. *Psychological Test and Assessment Modelling*, *53*(1), 106-140.

Anexo Escala EFE-Profesorado

A continuación se presentan afirmaciones que describen comportamientos en la infancia y adolescencia. Nos gustaría saber con qué frecuencia ha presentado **su alumno/a** estas conductas **en los últimos 6 meses**.

La escala contempla cinco alternativas de respuesta, en función de la frecuencia con que se dan estas conductas. Utilice los siguientes criterios para responder:

1 = Nunca; **2** = Pocas veces;
3 = Algunas veces; **4** = Muchas veces; **5** = Siempre

Por favor, conteste todos los enunciados. Responda rodeando con un círculo su respuesta. Si comete un error al responder, tache la respuesta errónea, y marque con un círculo la nueva respuesta.

1. Le cuesta seguir las normas (en el centro, con las actividades del aula o en el juego)	1	2	3	4	5
2. Es incapaz de cambiar su conducta ante diferentes situaciones	1	2	3	4	5
3. Le cuesta organizarse sin supervisión externa	1	2	3	4	5
4. Cuando se le da más de una instrucción, recuerda solo una de ellas	1	2	3	4	5
5.Parece sentirse inquieto	1	2	3	4	5
6. Comete errores por descuido en las actividades o en los exámenes	1	2	3	4	5
7. Le cuesta aceptar otros puntos de vista (en los trabajos en grupo, en el juego, con el profesorado, etc.)	1	2	3	4	5
8. Le cuesta preparar la mochila para salir de clase	1	2	3	4	5
9. Interrumpe en clase o se entromete en las actividades de sus compañeros/as	1	2	3	4	5
10. Le cuesta terminar lo que empieza	1	2	3	4	5
11. Tiene problemas para recordar cosas, incluso durante pocos minutos	1	2	3	4	5
12. Le cuesta estimar el tiempo que necesita para realizar actividades o exámenes	1	2	3	4	5
13. Su trabajo es poco cuidadoso	1	2	3	4	5
14. Reacciona de forma más exagerada ante pequeñas cosas que el resto de sus compañeros/as	1	2	3	4	5
15. Se distrae con sus propios pensamientos	1	2	3	4	5
16. Suele actuar sin pensar	1	2	3	4	5
17. Se enfada fácilmente	1	2	3	4	5
18. Necesita ayuda para seguir realizando una tarea	1	2	3	4	5
19. Presta poca atención a los detalles	1	2	3	4	5
20. Comienza tareas o trabajos en el último minuto	1	2	3	4	5
21. Le cuesta adaptarse a nuevas situaciones	1	2	3	4	5
22. Es incapaz de revisar los posibles errores cuando realiza una actividad o un examen	1	2	3	4	5
23. Cuando se le manda a por una cosa. se la suele olvidar	1	2	3	4	5
24. Le cuesta terminar lo que empieza	1	2	3	4	5
22. Es muy impulsivo/a en la forma de hablar o de comportarse	1	2	3	4	5
25. Los estallidos de ira o llanto son intensos, pero terminan rápidamente	1	2	3	4	5
26. Actúa de modo más incontrolado que los demás (en cambios de clase, recreos, etc.)	1	2	3	4	5
27. Le cuesta escuchar cuando se le habla	1	2	3	4	5