TESIS DOCTORAL

2021

El Desarrollo del Pensamiento Matemático en la Adolescencia: el Funcionamiento Ejecutivo en la Resolución de Problemas Aritméticos Verbales

Gonzalo Duque De Blas

PROGRAMA DE DOCTORADO EN PSICOLOGÍA DE LA SALUD

Universidad Nacional de Educación a Distancia (UNED)

Directores:

Dra. Isabel Gómez Veiga

Dr. Juan A. García Madruga

UNED Escuela Internacional de Doctorado EIDUNED

TESIS DOCTORAL

EL DESARROLLO DEL PENSAMIENTO MATEMÁTICO EN LA ADOLESCENCIA: EL FUNCIONAMIENTO EJECUTIVO EN LA RESOLUCIÓN DE PROBLEMAS ARITMÉTICOS VERBALES

GONZALO DUQUE DE BLAS

Programa de doctorado de Psicología de la Salud

Departamento de Psicología Evolutiva y la Educación. Facultad de Psicología

Directores

Dra. Isabel Gómez Veiga
Dr. Juan A. García Madruga

Madrid, noviembre de 2021

Aerodinámicamente, el cuerpo de una abeja no está hecho para volar. Lo bueno es que la abeja no lo sabe.

Anónimo

Dadas naves y velas apropiadas a los vientos celestes, habrá quien se atreva, incluso, con esa vastedad.

Carta de Kepler a Galileo - 1610

FINANCIACIÓN DE LA TESIS DOCTORAL

La presente tesis ha sido financiada gracias a las ayudas para la formación de doctores del Ministerio de Economía y Competitividad con referencia BES-2015-073057. Esta ayuda se contempló como parte de un Proyecto de Investigación financiado por dicho ministerio y titulado "Memoria Operativa y Educación: comprensión lectora, razonamiento y habilidades socioemocionales", con referencia EDU2014-56423-R. La consecución de la ayuda permitió formalizar un contrato predoctoral con la Universidad Nacional de Educación a Distancia (UNED) durante el periodo comprendido entre 2015 y 2019.

Además, el doctorando recibió una ayuda del fondo destinado a "Planes de Promoción de la Investigación I y II de la UNED 2020" para la publicación en revistas de acceso abierto que sufragó el coste de la publicación del estudio 1 contenido en esta tesis.

AGRADECIMIENTOS

Quisiera agradecer el esfuerzo, tiempo y paciencia dedicados por mis directores de tesis, Isabel Gómez Veiga y Juan Antonio García Madruga, para que esto pudiera llegar a término.

También, al Departamento de Psicología Evolutiva y la Educación de la Facultad de Psicología de la UNED, por abrirme sus puertas y acogerme como alumno y compañero.

A Barbara Carretti, Maria Chiara Passolunghi, Eva Izquierdo, Lucia Ronconi, Antonino Visalli, Cesare Cornoldi y el Dipartamento di Psicologia Generale (DPG) de la Scuola di Psicologia dell'Università degli Studi di Padova (Italia), quienes me recibieron con calor y afecto durante los meses de otoño de 2019.

A mi madre, Beatriz, el ser más noble, bello y verdadero de cuantos he conocido.

A mi padre, Miguel Ángel, por transmitirme el valor de la cultura como simiente de la libertad.

A mis hermanos, Miguel Ángel y Beatriz, hombros siempre dispuestos a recogerme.

A mis sobrinos, Ángela, Nicolás y Gabriel, con la esperanza de que algún día entiendan y me recuerden por qué no pude dedicarles el tiempo que merecían.

A mi familia, por ser el referente de mis logros.

A mi amigo José Ángel García, compañero en mil batallas y cobaya a deshoras.

A Laura Casco, Beatriz Herranz y M.ª Ángeles Cuevas, amigas de la Universidad y fedatarias de que alguna vez dejé poso allí.

A mis amigos Alberto Peña y Vicente López-Peláez, con quienes comparto afición por la más bella de las artes.

A lo que esté por llegar, pues de lo contrario nada habría tenido sentido.

A todos aquellos doctorandos que, alentados por la verdad, dignifican el oficio de buscar respuestas. Sea para ellos este canto de esperanza:

Si el don, cercenado;
por la empresa incierta.
Si las ganas, yertas;
de su afán postrado.

Si el honor, ajado;
por razón perversa.
Y amistad, aviesa;
de valor truncado.

Si el calor, negado;
de una mano artera.
Si la boca, seca;
de un amor cansado.

Si ni leptón fiado;
y el barquero acecha.
Si el deber ya espera;
sin destino al lado.
Oiga al yo trenado,
clamar esta aleya:

¡Mantén senda regia
y la fe a tu lado!

Nubes negras – 7 de septiembre de 2019

PREFACIO

Fue esa clase de azar que sonríe a quien *busca sin término* la que, poco antes de iniciar la redacción de mi tercer estudio de tesis doctoral a comienzos de 2019, trajo ante mí el Blog de Juan de Burgos Román, Catedrático ya retirado de la Escuela Superior de Ingenieros Aeronáuticos de la Universidad Politécnica de Madrid (UPM). En él compartía una buena colección de ejercicios matemáticos, la mayor parte fuera de mi alcance, y también algunas de sus experiencias docentes y reflexiones de distinto ámbito. Una de las entradas disponibles, titulada “Me enseñaron a pensar”, llamó poderosamente mi atención. Recogía un suceso, quizá ficticio, pero célebremente conocido bajo el título *Angels on the head of a pin*; un texto divulgado por Alexander Calandra, profesor de química de la Universidad de Chicago, reconocido entre otras cosas por sus innovadores métodos de enseñanza de las Matemáticas. La anécdota dice así:

***

Dicen que Sir Ernest Rutherford (1871-1937), presidente de la Real Sociedad Británica y Premio Nobel de Química en 1908, contaba la siguiente anécdota.

“Hace algún tiempo, recibí la llamada de un colega. Estaba a punto de poner un cero a un estudiante por la respuesta que había dado en un problema de física, pese a que este afirmaba con rotundidad que su respuesta era absolutamente acertada. Profesores y estudiantes acordaron pedir arbitraje de alguien imparcial y fui elegido yo. Leí la pregunta del examen:

> *Demuestre cómo es posible determinar la altura de un edificio con la ayuda de un barómetro.*

El estudiante había respondido lo siguiente:

Lleve el barómetro a la azotea del edificio y átele una cuerda muy larga. Descuélguelo hasta la base del edificio, marque y mida. La longitud de la cuerda es igual a la longitud del edificio.

Realmente, el estudiante había planteado un serio problema con la resolución del ejercicio, porque había respondido a la pregunta correcta y completamente. Por otro lado, si se le concedía la máxima puntuación, podría alterar el promedio de su año de estudios, obtener una nota más alta y así certificar su alto nivel en física; pero la respuesta no confirmaba que el estudiante tuviera ese nivel. Sugerí que se le diera al alumno otra oportunidad. Le concedí seis minutos para que me respondiera la misma pregunta, pero esta vez con la advertencia de que en la respuesta debía demostrar sus conocimientos de física.

Habían pasado cinco minutos y el estudiante no había escrito nada. Le pregunté si deseaba marcharse, pero me contestó que tenía muchas respuestas al problema. Su dificultad era elegir la mejor de todas. Me excusé por interrumpirle y le rogué que continuara. En el minuto que le quedaba escribió la siguiente respuesta:

Coja el barómetro y láncelo al suelo desde la azotea del edificio, calcule el tiempo de caída con un cronómetro. Después aplique la correspondiente fórmula: la altura es igual a 0.5 por la aceleración de la gravedad en la Tierra por el cuadrado del tiempo; y así obtendremos la altura del edificio.

En este punto le pregunté a mi colega si el estudiante se podía retirar. Le dio la nota más alta. Tras abandonar el despacho, me reencontré con el estudiante y le pedí que me contara sus otras respuestas a la pregunta.

Bueno, -respondió- hay muchas maneras. Por ejemplo, coges el barómetro en un día soleado y mides la altura del barómetro y la longitud de su sombra. Si medimos a continuación la longitud de la sombra del edificio y aplicamos una simple proporción, obtendremos también la altura del edificio.

¡Perfecto! - le dije. ¿Y de otra manera?

Sí -contestó-. Éste es un procedimiento muy básico para medir un edificio, pero también sirve. Coges el barómetro y te sitúas en las escaleras del edificio en la planta baja. Según subes las escaleras, vas marcando la altura del barómetro y cuentas el número de marcas hasta la azotea. Multiplicas al final la altura del barómetro por el número de marcas que has hecho y ya tienes la altura. Éste es un método muy directo.

Por supuesto, si lo que quiere es un procedimiento más sofisticado, puede atar el barómetro a una cuerda y moverlo como si fuera un péndulo. Si calculamos que, cuando el barómetro está a la altura de la azotea, la gravedad es cero, y si tenemos en cuenta la medida de la aceleración de la gravedad al descender el barómetro en trayectoria circular al pasar por la perpendicular del edificio, de la diferencia de estos valores, y aplicando una sencilla fórmula trigonométrica, podríamos calcular, sin duda, la altura del edificio.

En este mismo estilo de sistema, atas el barómetro a una cuerda y lo descuelgas desde la azotea a la calle. Usándolo como un péndulo puedes calcular la altura midiendo su periodo de precisión. En fin, -concluyó- existen otras muchas maneras. Probablemente, la mejor sea coger el barómetro y golpear con él la puerta de la casa del conserje. Cuando abra, decirle: “Señor conserje, aquí tengo un bonito barómetro. Si usted me dice la altura de este edificio, se lo regalo”.

En este momento de la conversación, le pregunté si no conocía la respuesta convencional al problema: la diferencia de presión marcada por un barómetro en dos lugares diferentes nos proporciona la diferencia de altura entre ambos lugares.

Él contestó que sí la conocía, pero que durante sus estudios sus profesores habían intentado enseñarle a pensar.

El estudiante se llamaba Niels Bohr (1885-1962), físico danés y premio Nobel de Física en 1922; conocido por ser el primero en proponer el modelo de átomo con protones y neutrones rodeados de electrones. Fue, fundamentalmente, un innovador de la teoría cuántica.”

***

Este texto constituye un magnífico y divertido alegato en favor del pensamiento divergente; aquel que nos permite afrontar un proceso de resolución de problemas a partir de soluciones creativas, innovadoras y, a veces, parsimoniosas. El ejercicio de la Ciencia impone la exploración de nuevas vías de conocimiento con las que afrontar los retos futuros, y ello requiere un tipo de pensamiento que, dominado a voluntad, sea capaz trascender del espacio encorsetado por las reglas lógicas y las costumbres.

Asimismo, subraya un efecto ocasionalmente adverso y, en todo caso, natural que se produce como consecuencia del propio proceso de aprendizaje; esto es, la adquisición y automatización de reglas heurísticas; atajos que, si bien nos permiten responder adecuadamente y con enorme eficiencia a un amplio corolario de situaciones, aplicadas sobre escenarios aparentemente similares pueden inducir una tendencia automática de respuesta y, en consecuencia, la toma de decisiones equivocadas. Estos dos estilos de pensamiento, uno automático y otro reflexivo, son -en síntesis- el núcleo central de este trabajo de investigación, cuyo contexto es el de la resolución de problemas aritméticos verbales.

ÍNDICE

LISTA DE SÍMBOLOS, ABREVIATURAS Y SIGLAS

OECD	Organización para la Cooperación y el Desarrollo Económicos
PISA	Programa para la Evaluación Internacional de los Alumnos
ESO	Educación Secundaria Obligatoria
STEM	Science, Technology, Engineering, and Mathematics
AWP	Arithmetic word problem
AWP-1-op	One operation arithmetic word problem
AWP-1-op	Two operations arithmetic word problem
AWP CON	Consistent arithmetic word problem
AWP INC	Inconsistent arithmetic word problem
AWP C-I	Consistent-then-Inconsistent arithmetic word problem
AWP I-C	Inconsistent-then-Consistent arithmetic word problem
WM	Working memory
Sup	Superficial response
FE	Funciones ejecutivas
EF	Executive functions
EC	Ejecutivo central
CE	Central Executive system

LISTA DE TABLAS

Introducción

Estudio 1

Estudio 2

Estudio 3

Anexo

LISTA DE FIGURAS

Introducción

Estudio 2

Estudio 3

RESUMEN

El desarrollo del pensamiento matemático en la adolescencia: el funcionamiento ejecutivo en la resolución de problemas aritméticos verbales

El objetivo de esta tesis fue profundizar en el estudio de los procesos cognitivos y metacognitivos que subyacen a la resolución de problemas aritméticos verbales (AWP) en la adolescencia. La resolución de problemas es una de las competencias curriculares básicas que deben adquirirse desde los primeros cursos de la escuela primaria, pues constituyen una simulación controlada de una situación de incertidumbre para la cual la persona a cargo de la resolución debe encontrar una respuesta aplicando todos sus recursos cognitivos y conocimientos. El nivel de competencia matemática mostrado por el alumnado español en las pruebas externas de conocimiento de la OCDE, el informe Pisa, ha permanecido por debajo de la media europea desde la fundación de este organismo. Entender las razones por las que persiste este bajo rendimiento resulta de interés general, ya que estos indicadores educativos permiten predecir el nivel de desarrollo científico, económico y laboral de una nación, y la prosperidad de sus habitantes. Dada la complejidad y diversidad de mecanismos cognitivos y metacognitivos implicados en la resolución de AWP, consideramos que esta tarea podría también resultar de utilidad para la predicción del rendimiento académico del estudiante de educación secundaria, incluso, en áreas no relacionadas necesariamente con las materias STEM.

Para ello, diseñamos tres estudios orientados contrastar diversas hipótesis. El primer estudio tenía por objetivo analizar la eficacia en la resolución de una tarea compuesta por 5 AWP modulados en dificultad en base a dos criterios concretos: número de operaciones y consistencia/inconsistencia de la estructura superficial de los problemas con la operación aritmética necesaria para la resolución. Tanto la eficacia, como el tipo de respuesta característica

a cada problema fueron analizados. Posteriormente, se analizó el grado de relación entre la resolución de problemas y tres habilidades cognitivas básicas: inteligencia fluida, la comprensión lectora y el razonamiento deductivo. Por último, comprobamos la capacidad explicativa de estos problemas sobre el rendimiento académico en dos materias muy diferentes: Matemáticas e Historia/Geografía. Para este estudio, participaron 135 estudiantes de 2º y 3º de educación secundaria (13.5 años). Los resultados confirmaron el éxito de la manipulación

experimental en la modulación de la dificultad de los problemas. La eficacia en la resolución de AWP estuvo correlacionada de forma moderada y positiva con la inteligencia y comprensión de textos, un aspecto también confirmado mediante análisis de regresión; y de forma baja con los procesos superficiales de la lectura y el razonamiento deductivo. Menos de la mitad de los AWP fueron resueltos correctamente, siendo los problemas inconsistentes los más difíciles. Los errores, definidos como la ejecución incorrecta de un plan -correcto o incorrecto- de resolución del problema, fueron el tipo de equivocación más frecuentemente cometida por encima de las respuestas superficiales, aquellas que suponen una ejecución correcta de un plan equivocado de solución. Asimismo, los AWP demostraron su capacidad explicativa del rendimiento académico en Matemáticas y en Historia/Geografía, al mismo o mayor nivel que la medida tradicional de inteligencia.

En el segundo experimento, realizado en colaboración con la Dra. Maria Chiara Passolunghi (Universidad de Trieste, Italia) y la Dra. Barbara Carretti (Universidad de Padua, Italia), tuvo por objetivo estudiar la relación entre la eficacia en la resolución de AWP y diversas habilidades cognitivas (inteligencia fluida, comprensión lectora, inhibición y actualización) en una muestra de 182 participantes italianos de 5º de enseñanza primaria (10.5 años). Se utilizaron dos versiones de 6 AWP con el fin de comprobar si existían diferencias entre el uso del término verbal "más" o el término verbal "menos" en sus diferentes modalidades (consistente/inconsistente). Asimismo, se comprobó si existían diferencias en la eficacia de la resolución debida al orden en el que estaba insertada la inconsistencia en los problemas de dos operaciones. Otras medidas cognitivas, como la inteligencia fluida, la comprensión lectora, la actualización de contenidos en la memoria operativa y la inhibición de respuestas prepotentes fueron tomadas para comprobar su relación con los problemas matemáticos. Los resultados obtenidos mostraron una relación moderada entre la inteligencia fluida y la comprensión lectora con los AWP en enseñanza primaria. No existieron diferencias significativas entre las versiones de los problemas por efecto del uso del adverbio de cantidad ("más" o "menos"). Asimismo, se comprobó que los problemas más difíciles de resolver fueron los inconsistentes. Estas diferencias en eficacia estuvieron relacionadas con las habilidades de actualización e inhibición de los participantes. El orden de presentación de la inconsistencia afectó a la eficacia de la resolución de los problemas. Cuando la inconsistencia estuvo localizada en el primer término u operación, la eficacia de los participantes se reducía. Este efecto del orden pudo relacionarse con la eficiencia de los procesos de actualización de los participantes, de modo que cuando ésta era controlada, las diferencias encontradas relativas al orden de presentación de la inconsistencia se neutralizaban. No obstante, este resultado debe ser considerado con cierta cautela debido a que el orden de presentación de los problemas no fue contrabalanceado.

El tercer estudio tuvo por objetivo estudiar la relación existente entre dos tipos de problemas aritméticos: los AWP y los problemas de reflexión cognitiva (CRT). Ambos tipos de problemas muestran ciertas semejanzas, especialmente con los AWP inconsistentes, en relación a que ambos son capaces de generar una tendencia automática de respuesta superficial que es necesaria inhibir para hallar la solución. También, la eficacia en la resolución de ambos tipos de problemas se ha asociado a un rendimiento académico más elevado en matemáticas, por lo que su uso combinado podría constituir una medida más predictiva del logro. Para ello, AWP y CRT fueron comparados en función de cuatro criterios distribuidos en dos experimentos distintos: (1) la eficacia en la resolución y el tipo de respuesta que generan; (2) la experiencia de dificultad percibida por los participantes; (3) la relación que tienen con las medidas de funcionamiento ejecutivo, un aspecto de particular interés en esta tesis y; (4) su capacidad explicativa sobre el rendimiento académico.

El primer experimento tuvo por objetivo analizar las diferencias en la eficacia de resolución en estos dos problemas y estudiar el tipo de respuesta (intuitiva vs reflexiva) que se produce con más frecuencia. Además, se comprobó cómo se relacionaban con las medidas de inteligencia fluida y comprensión lectora. Para este estudio, participaron 81 estudiantes españoles de 4º curso de Educación Secundaria Obligatoria (ESO). Los resultados obtenidos indicaron que la eficacia de resolución en AWP fue notablemente mayor que la obtenida en CRT, con marcado efecto suelo. También difirieron en el tipo de respuesta más frecuente que producen. En el caso del AWP, las respuestas más frecuentes fueron los aciertos y los errores, mientras que en el CRT fue mayoritaria la respuesta superficial. Ambos problemas correlacionaron con la inteligencia y la comprensión de textos, aunque las asociaciones con CRT fueron más débiles debido al escaso número de aciertos.

El experimento 2 tuvo por objetivo confirmar los resultados obtenidos en el experimento 1 relativos a la eficacia en los dos tipos problemas analizando, también, si dicha eficacia podía estar relacionada con la percepción de la dificultad de los participantes. Asimismo, se comprobaron las relaciones existentes entre la resolución de los dos tipos de problemas y diversas medidas cognitivas generales (inteligencia fluida y comprensión) y ejecutivas (actualización, inhibición y planificación). Por último, se comprobó la capacidad explicativa de ambos problemas sobre dos medidas distintas de rendimiento académico: una medida general a final del curso y una medida específica externa de conocimientos matemáticos para secundaria (COMPEMAT). Participaron 82 estudiantes españoles de 3º y 4º curso de Educación Secundaria Obligatoria (ESO). En relación a la eficacia obtenida en los problemas, los resultados fueron similares a los del experimento 1. En ambos tipos de problemas se constató un marcado sesgo positivo de los participantes hacia la dificultad que les suponía la resolución de los problemas ("bastante fáciles"). No existió correlación entre la eficacia de la resolución y la percepción de la dificultad. Respecto a los dos criterios en que la dificultad de los AWP fue modulada, se obtuvieron resultados diversos. La diferencia en la precisión de respuesta entre los problemas consistentes e inconsistentes fue significativa. Esta diferencia tuvo que ver más con la inteligencia y la inhibición de los participantes que con el resto de medidas. Respecto a las diferencias entre problemas de una o dos operaciones, fueron significativas sólo en los problemas consistentes. Tomados en conjunto los problemas, los análisis de senderos indicaron que la resolución de AWP es una actividad compleja en la que intervienen de manera más significativa los procesos de inhibición, planificación y la inteligencia fluida; mientras que la resolución de CRT parece estar ligada a la inteligencia fluida y la inhibición de respuestas, aunque no pudo encontrarse significación estadística debido al bajo rendimiento obtenido en esta tarea. Por último, los AWP fueron capaces de contribuir de manera significativa a la explicación de las dos medidas de rendimiento académico, aunque de manera diferente. Con la medida de conocimientos matemáticos (COMPEMAT) obtuvo significación estadística, aunque el factor más explicativo fue la inteligencia. Sin embargo, fue el factor más explicativo en la media global del alumno a final del año académico. CRT sólo pudo demostrar marginalmente su capacidad para explicar el rendimiento académico en ambas medidas.

Concluyendo, este estudio ha permitido constatar la diversidad de procesos cognitivos y también algunos metacognitivos implicados en la resolución de AWP en la adolescencia, así como determinar en qué grado contribuyen cada una de las medidas cognitivas tomadas (inteligencia, comprensión, actualización, inhibición y planificación) a la explicación de las diferencias en la eficacia de resolución que se encuentran entre los distintos tipos de AWP. Como hemos demostrado a lo largo de este trabajo, los AWP ofrecen potencial suficiente para explicar el rendimiento académico de los alumnos en, al menos, el mismo grado que otras medidas tradicionales como la inteligencia fluida y en áreas de conocimiento diversas, como la Historia/Geografía, o la media académica general a final del curso. Asimismo, a pesar de compartir ciertas características, deben considerarse los problemas CRT como un tipo especial de problema aritmético muy diferente del AWP debido a su gran dificultad y capacidad para generar respuestas superficiales incluso entre adultos y expertos. Este hecho impide la posibilidad de que estos problemas pudieran ser utilizados para la predicción de rendimiento académico en la adolescencia. Futuros estudios deberían ir dirigidos a explorar la eficacia y transferencia de los programas de intervención para la mejora de la resolución de AWP basados en el entrenamiento de procesos ejecutivos en educación primaria y secundaria.

SUMMARY

The development of mathematical thinking in adolescence: executive functioning in verbal arithmetic problem solving

The aim of this thesis was to deepen the study of the cognitive and metacognitive processes that underlie the resolution of arithmetic word problems (AWP) in adolescence. Problem solving is one of the basic curricular skills that should be acquired from the first years of primary school, since they constitute a controlled simulation of a situation of uncertainty for which the solver must find a solution by applying all his or her cognitive resources and knowledge. The level of mathematical competence shown by Spanish students in the OECD's external tests of knowledge, the Pisa report, has remained below the European average since the foundation of this organisation. Understanding the reasons for this persistent underperformance is of general interest, since these educational indicators predict the nation's level of scientific, economic and labour market development, and the prosperity of its people. Given the complexity and diversity of cognitive and metacognitive mechanisms involved in AWP solving, we consider that this task could also be useful for the prediction of secondary school students' academic performance, even in areas not necessarily related to STEM subjects.

To this end, we designed three studies aimed at testing different hypotheses. The first study aimed to analyse the efficiency in solving a task consisting of 5 AWPs that were modulated in difficulty according to two specific criteria: number of operations and consistency/inconsistency of the superficial structure of the problems with the necessary arithmetic operation for the resolution. Both the efficiency and the type of response characteristic of each problem were analysed. Subsequently, we analysed the extent of relationship between problem solving and three

basic cognitive skills: fluid intelligence, reading comprehension and deductive reasoning. Finally, we tested the explanatory capacity of these problems on academic performance in two very different subjects: Mathematics and History/Geography. For this study, 135 students in the 2nd and 3rd grades of secondary education (13.5 years) participated. The results confirmed the success of the experimental manipulation in modulating the difficulty of the problems. Efficacy in solving AWPs was positive and moderately correlated with higher intelligence and reading comprehension, also confirmed by regression analysis; and low correlated with surface reading processes and deductive reasoning. Less than half of the AWPs were solved correctly, with inconsistent problems being the most difficult. Errors, defined as the incorrect execution of a problem-solving plan, were the most frequently type of error committed over superficial responses, those involving correct execution of a wrong solution plan. The AWP also demonstrated its ability to explain academic performance in Mathematics and History/Geography at the same or higher level than the traditional measure of intelligence, respectively.

In the second experiment, conducted in collaboration with Dra. Maria Chiara Passolunghi (University of Trieste, Italy) and Dra. Barbara Carretti (University of Padua, Italy), the aim was to study the relationship between AWP solving efficiency and several cognitive skills (fluid intelligence, reading comprehension, inhibition and updating) in a sample of 182 Italian participants in 5th grade (10.5 years old). Two versions of 6 AWPs were used to test whether there were differences between the use of the verbal term "more" or the verbal term "less" in its different modalities (consistent/inconsistent). We also tested whether there were differences in solving efficiency due to the order in which the inconsistency was inserted in the two-operation problems. Other cognitive measures, such as fluid intelligence, reading comprehension, updating and inhibition of prepotent responses were taken to test their relationship with the solving accuracy of arithmetical problems. The results obtained showed a moderate relationship between fluid intelligence and reading comprehension with AWPs in primary school. There were no significant differences between the versions of the problems due to the effect of the use of the adverb of quantity ("more" or "less"). In addition, it was found that the most difficult problems to solve were the inconsistent ones. These differences in efficiency were related to the participants' updating and inhibition skills. The order of presentation of the inconsistency affected problem-solving efficacy. When the inconsistency was located in the first term or operation, participants' efficacy was reduced. This order effect could be related to the efficiency of the participants' updating processes, so that when this was controlled, the differences found regarding the order of presentation of the inconsistency were neutralised. However, this result should be viewed with some caution because the order of presentation of the problems was not balanced.

The third study aimed to study the relationship between two types of arithmetic problems: AWP and cognitive reflection problems (CRT). Both types of problems show certain similarities, especially with inconsistent AWP, in that both are capable of generating an automatic tendency of superficial response that needs to be inhibited in order to find the solution. Also, both problems have been associated with higher academic achievement, so their combined use could be a more predictive measure of achievement. To this end, AWP and CRT were compared on the basis of four criteria distributed in two different experiments: (1) the efficiency in solving and the type of response they generate, (2) the experience of difficulty perceived by the participants, (3) the relationship they have with cognitive measures and (4) their explanatory capacity on academic performance.

The first experiment aimed to analyse the differences in solving efficiency in these two problems and to study the type of response (intuitive vs. reflexive) that occurs more frequently. In addition, it was tested how they were related to measures of fluid intelligence and reading comprehension. For this study, eighty-one students in the 4th grade of Secondary Education participated. The results obtained indicated that the resolution efficiency in AWP was notably higher than that obtained in CRT, with a marked ‘floor effect’. They also differed in the type of response most frequently produced. In the case of AWP, the most frequent response was 'correct' and 'errors' responses, while 'superficial' was the majority response for CRT. Both problems correlated with intelligence and text comprehension, although the associations with CRT were weaker due to the low number of hits.

The aim of experiment 2 was to confirm the results obtained in experiment 1 regarding efficiency in the two types of problems, also analysing whether this efficiency could be related to the participants' perception of difficulty. We also tested the relationships between the resolution of the two types of problems and several general cognitive measures (fluid intelligence and comprehension) and executive measures (updating, inhibition and planning). Finally, the explanatory power of both problems was tested on two different measures of academic performance: a general end-of-year measure and a specific external measure of mathematical literacy for secondary school (COMPEMAT). Eighty-two students in the 3rd and 4th years of Secondary Education participated. In relation to the efficiency obtained in the problems, the results were similar to those of experiment 1. For both types of problems, there was a marked positive bias from participants towards the difficulty of solving the problems, which was considered "quite easy". There was no correlation between problem-solving efficiency and perceived difficulty. For the two criteria in which the difficulty of the AWPs was modulated, different results were obtained. The difference in response accuracy between consistent and inconsistent problems was significant. This difference had more to do with the intelligence and inhibition of the participants than with the other measures. Regarding the differences between one- and two-operation problems, they were significant only in the consistent problems. Taken together across problems, path analyses indicated that AWP solving is a complex activity in which inhibition, planning and fluid intelligence processes are most significantly involved; whereas CRT solving appears to be linked to fluid intelligence and response inhibition, although statistical significance could not be found due to the low performance on this task. Finally, AWPs were able to contribute significantly to the explanation of the two measures of academic performance, although in different ways. With the measure of mathematical knowledge (COMPEMAT) it obtained statistical significance, although the most explanatory factor was intelligence. However, it was the most explanatory factor in the student's overall average at the end of the academic year. As for CRT, it could only marginally demonstrate its ability to explain academic performance on both measures.

In conclusion, this study has allowed us to ascertain the diversity of cognitive and also some metacognitive processes involved in the resolution of AWPs in adolescence, as well as to determine how much each of the cognitive measures taken (intelligence, comprehension, updating, inhibition and planning) contribute to the explanation of the differences in resolution efficacy found among the different types of AWPs. As we have shown throughout this thesis, AWPs offer sufficient potential to explain students' academic performance to at least the same degree as other traditional measures such as fluid intelligence and in diverse subject areas, such as History/Geography, or the overall academic average at the end of the year. Also, despite sharing certain characteristics, CRT problems should be considered as a special type of arithmetic problem very different from AWP because of their great difficulty and ability to generate superficial responses even among adults and experts. This fact precludes the possibility that these problems could be used for the prediction of academic performance in adolescence. Future studies should be aimed at exploring the efficacy and transferability of intervention programmes for the improvement of AWP resolution based on executive process training in primary and secondary education.

1. INTRODUCCIÓN

Las Matemáticas son una herramienta imprescindible en cualquier ámbito de la vida. Es el lenguaje de la ciencia y el modo mediante el cual creamos representaciones de la realidad con el fin de analizarlas, computarlas, elaborar predicciones o determinar el mejor modo de producir cambios y transformaciones. Uno de los factores que mejor predice el desarrollo de un país es la educación de sus habitantes, pues de su nivel dependerá que surjan nuevos productos, patentes o que se realicen contribuciones relevantes al conocimiento científico que mejoren las condiciones de vida de la población. Existen muchos estudios que relacionan la competencia matemática con una mayor empleabilidad y nivel de vida (ver Caspi et al., 1998; Marjoribanks, 2005; Watt et al., 2012), por lo que es necesario garantizar un nivel adecuado de competencia a lo largo de toda la vida académica del estudiante. Este es uno de los principios fundacionales del informe PISA: valorar la educación que adquieren las nuevas generaciones con el fin de pronosticar su desarrollo y empleabilidad posterior.

El último informe PISA publicado hasta la fecha presente (OECD, 2019) indicó que los estudiantes españoles de 4º de Educación Secundaria Obligatoria (ESO) lograron alcanzar una puntuación media de 481 puntos en competencia matemática, lo que nos sitúa por debajo de la media de la UE y de los países miembros de la OCDE. Desde 2006 hasta 2015, las diferencias entre el alumnado español y la media de la OCDE en competencia matemática han ido decreciendo, aunque no como consecuencia de lograr un mejor desempeño y unas puntuaciones más elevadas, sino por el hecho de que -sostenidamente- la media de la OCDE ha ido decreciendo. Actualmente, el nivel de conocimientos matemáticos de los alumnos españoles es similar a los obtenidos en 2009, por lo que el incremento neto a lo largo de la última década es prácticamente nulo.

Paralelamente, el número de estudiantes que acceden en las enseñanzas universitarias del ámbito de las ciencias o STEM (Sciences, Technology, Engineering and Mathematics) es alrededor de 5 puntos porcentuales inferior a la medida europea, según el último informe de la

Conferencia Rectora de Universidades Españolas (Hernández-Armenteros et al., 2018). A pesar de la creciente oferta de puestos de trabajo en las empresas tecnológicas y las mejores condiciones laborales que se ofrecen en términos de salario, en comparación con otras áreas, el número de estudiantes matriculados en estas enseñanzas que reciben cada año las Universidades es inferior al de los países de nuestro entorno, especialmente entre las mujeres. Las razones de estos resultados son de naturaleza variada, por lo que cualquier esfuerzo que se realice en la dirección de comprender cómo mejorar los resultados en la educación en general, y las Matemáticas en particular, resulta de interés público.

Esta tesis doctoral tiene por objeto estudiar los procesos cognitivos y metacognitivos implicados en la resolución de problemas aritméticos verbales (AWP) en la adolescencia, una competencia curricular básica que todo estudiante debe adquirir desde la Educación Primaria (EP) y que, por su naturaleza, supone la adquisición de unas habilidades generalizables a todos los ámbitos de la vida.

Todo proceso formal de resolución de problemas implica la búsqueda racional de los cambios necesarios para transitar desde un estado inicial problemático hasta un objetivo o meta final mediante la aplicación de los conocimientos, habilidades y estrategias que dispone la persona a cargo de su resolución. Es un proceso complejo y no exento de problemas que demanda una enorme diversidad de habilidades cognitivas generales y específicas, tales como: a) *las habilidades de razonamiento*, por medio de las cuales deducimos las relaciones que se establecen entre los distintos elementos del problema; b) *la comprensión lectora*, gracias a la cual logramos entender el enunciado del problema y logramos generar un modelo situacional del mismo basado en los aspectos relevantes para la resolución; c) *la planificación* de la secuencia de resolución, mediante la cual se establecen las metas y submetas de la tarea, junto con los procedimientos necesarios para cada una de ellas; d) *la memoria operativa*, gracias a la cual se logra mantener y actualizar en la mente de la persona los distintos resultados parciales, así como el objetivo global que lo motiva; e) *la inhibición* de respuestas superficiales o tendencias heurísticas de respuesta que, aplicadas sobre situaciones con apariencia similar, conducirían a la toma de decisiones equivocadas; f) *la verificación de resultados* que se realiza sobre el producto resultante con el objetivo de detectar incongruencias mediante, por ejemplo, la búsqueda de contraejemplos, y; g) *el control metacognitivo*, necesario para dirigir estratégicamente la acción y los recursos que se disponen, y sobreponerse a las reacciones emocionales que pueda generar la tarea, todos ellos causantes en algunos casos de una subestimación de la dificultad de la tarea y un bajo rendimiento.

Las características intelectuales y formativas de la persona que resuelve el problema constituirían, por tanto, la caja de herramientas necesaria para llevar a cabo todo el proceso. Sin embargo, estas habilidades sufren cambios y transformaciones como consecuencia del desarrollo normotípico del individuo, la progresiva maduración de sus estructuras cerebrales y el incremento de la eficiencia de los procesos mentales subyacentes a la tarea. Tales habilidades gozarían de un desarrollo significativo con la llegada de la adolescencia. En esta etapa se produce la conquista del pensamiento hipotético deductivo y la abstracción como mecanismo de taxonomización de la realidad (Inhelder & Piaget, 1955; Piaget & Inhelder, 1973) que permite a la persona desvincularse del objeto material e incrementar su conocimiento sobre el funcionamiento de las cosas, así como la adquisición de las estrategias necesarias para alcanzar las metas propuestas y, el incremento del conocimiento acerca de sí mismo. Además, se producen otros cambios que afectan al funcionamiento ejecutivo que, en conjunto, logran una mejora de la capacidad de resolución de problemas, la aritmética y la geometría (Eccles et al., 2003).

La adolescencia es un periodo de cambios, también, en relación a los intereses que muestran los alumnos en sus asignaturas y el despertar vocacional (Hoff et al., 2018). Estos intereses, aunque cambiantes a lo largo de esta etapa, pueden verse determinados por diversos factores entre los que destacan las actitudes hacia las Matemáticas, las creencias del estudiante acerca de dicha materia de estudio, su motivación y autorregulación emocional, las experiencias que va adquiriendo como consecuencia de su estudio y los resultados académicos que logra en los centros educativo, entre otros (ver una revisión más exhaustiva en Hannula et al., 2016), de modo que es más probable que desarrollen vocaciones en áreas en las que su percepción de autoeficacia y la facilidad con la que predicen un resultado positivo son más elevadas (Nauta et al., 2002). Por esta razón es importante determinar los factores que puedan contribuir a un rendimiento ineficiente en las áreas científicas y la educación en general con el fin de evitar que talentos potenciales acaben siendo desperdiciados por causas ciertamente controlables.

En este sentido, y relacionado con el rendimiento académico, otro de los objetivos principales que tiene esta tesis es comprobar en qué medida la eficacia en la resolución de problemas aritméticos verbales podría utilizarse como medida para la predicción de un rendimiento académico, no sólo en asignaturas relacionadas con esa misma área de conocimiento, sino también como una medida del rendimiento global del estudiante de Educación Secundaria. La resolución de problemas aritméticos no sólo es un ejercicio que requiere un conocimiento procedimental de las competencias aritméticas básicas para el cálculo de las operaciones. Representan, también, un ejercicio de comprensión llevado al campo de la aritmética, así como un ejercicio de reflexión gracias al cual se determinan las relaciones entre los elementos del problema y se deducen las operaciones que, posteriormente, se ejecutan. Estas habilidades no se emplean exclusivamente en el área de las Matemáticas y las ciencias en general. También son comunes a otro tipo de asignaturas, como Historia, en las que existe un creciente interés por desarrollar el pensamiento crítico del alumno de modo que sea capaz de extraer conclusiones racionales a partir de eventos complejos tras evaluar y relativizar los distintos puntos de vista desde los que pueden percibirse y se desarrollan los acontecimientos. Por esta razón, consideramos que los problemas aritméticos verbales contarían con potencial suficiente para ayudar a pronosticar el rendimiento académico global de los estudiantes.

En los siguientes epígrafes se incluye una breve revisión teórica de los aspectos más relevantes relativos a la resolución de problemas aritméticos. En primer lugar, se realizará un repaso sobre la representación numérica y la adquisición y desarrollo del pensamiento matemático. El tipo de problemas aritméticos, las etapas de resolución, los modelos teóricos propuestos, así como las estrategias que se emplean serán tratados posteriormente. Más tarde, hablaremos de las habilidades generales implicadas en la resolución de problemas aritméticos, diferenciando entre dos tipos de habilidades cognitivas generales, la comprensión lectora y el razonamiento deductivo, y las habilidades metacognitivas. Por último, estudiaremos en profundidad la memoria operativa y los procesos ejecutivos, piezas clave del sistema cognitivo que están estrechamente ligadas a la resolución de problemas.

1.1. La representación numérica y el pensamiento matemático

El pensamiento matemático es una habilidad que implica manejar códigos lingüísticos y símbolos numéricos con los que se opera para conseguir una respuesta -generalmente- numérica (Sternberg y Ben-Zeev, 1996). No es, de ningún modo, una facultad limitada al mero cálculo de una serie de magnitudes, sino un ejercicio de *pensamiento productivo* de amplio espectro que incluiría, entre otros, el estudio de patrones, la resolución de problemas o el planteamiento de conjeturas (Nickerson, 2010).

El desarrollo del pensamiento matemático está ligado necesariamente a la capacidad de representar mentalmente el conocimiento y el desarrollo de la aritmética. Una representación constituiría un modelo análogo al fenómeno percibido que se incorpora y construye en la mente del sujeto a partir de una experiencia idiosincrática con la realidad externa. Esta representación es producto de un código responsable de su codificación y almacenamiento que habilitaría la posibilidad de que su contenido fuera manipulado en ausencia de la realidad material, permitiendo así la predicción de sucesos con los que poder regular la conducta. El tipo de representaciones y el formato en que se codifican las representaciones es un asunto extenso y origen de posturas teóricas diversas. Por esa razón, nos referiremos únicamente a las representaciones de tipo numérico.

1.1.1. Piaget y el desarrollo del pensamiento matemático

Para Piaget (1952), el pensamiento matemático es un proceso constructivo que se desarrollaría como consecuencia del desarrollo de otras estructuras lógicas más generales. El número no sería una entidad material que se aprehende, sino el resultado de una abstracción o síntesis lógica que se construye en la mente del niño, y que deriva de dos procesos lógicos concretos: a) *la clasificación jerárquica de clases*, una función que permitiría a la persona agrupar diferentes elementos en torno a un criterio de semejanza (Ej. Dos plátanos y tres manzanas forman un conjunto de cinco piezas de fruta) y; b) *la seriación*, que permite agrupar un número discreto de elementos en base a algún criterio que, también, los puede diferenciar. A partir de este proceso podría establecerse, por ejemplo, una correspondencia transitiva y unidireccional entre dos elementos (Ej. Marta es más alta que Antonio).

En el proceso de construcción del número, los niños aprenderían que la realidad material puede ser entendida y estructurada gracias a que puede ser contada y ordenada, lo que les permite adquirir de manera progresiva un sentido de conservación del número y su cardinalidad. La comprensión del número para este autor no podría iniciarse antes de los 7-8 años, en el estadio de las operaciones concretas, debido a varias razones. En primer lugar, porque su conducta permanece todavía anclada a la información perceptiva que dispone del objeto y cuya presencia es necesaria para que el pensamiento pueda operar sobre él. También, porque el número de características mediante las que se clasifican los objetos es muy limitado. La realidad física se entiende como una entidad estática que perdería su identidad a medida que se introducen cambios en los objetos y se transforma su apariencia o el modo en que se presentan. La capacidad para operar mentalmente con una representación del objeto permitiría al niño poder reconstruirlo y revertir los cambios que puedan ser introducidos en los mismos, lo que permitiría conservarlos como entidades independientes de sus atributos perceptivos. Para ello, será necesario que adquiera también ciertas nociones de tipo espacio-temporal a partir de las cuales sea posible establecer un estado inicial y final que anteceden o resultan de la secuencia de cambio que transforma el objeto (García-Madruga & Vila, 2013).

Para Piaget una de las diferencias entre el pensamiento adolescente y el pensamiento infantil tendría que ver con la capacidad de razonar a partir de hipótesis sin vínculo material, lo cual permite considerar las diferentes alternativas de respuesta como explicaciones plausibles de un mismo fenómeno. Así, entre los 8 y los 11 años los niños desarrollarían una capacidad de abstracción que les permitiría relacionar conjuntos de elementos heterogéneos que, en un estadio formal, podrían combinarse con el fin de establecer relaciones y causalidades, dando así inicio al pensamiento hipotético-deductivo característico de este periodo.

1.1.2. Representación numérica, subitación y conteo

Sin embargo, los estudios llevados a cabo mediante paradigma de habituación han permitido determinar que los niños a partir de los 5-6 meses ya poseen una representación del número que les permite distinguir entre conjuntos de pequeñas cantidades de objetos (ver Starkey & Cooper, 1980; Wynn, 1995). Esta habilidad no-simbólica, automática e innata de reconocer a primera vista entre un número discreto de elementos (entre 2 y 5 elementos, según las condiciones de la tarea), denominada *subitación* (Ginsburg, 1978; Kaufman et al., 1949), no es un mero proceso de reconocimiento de patrones (ver Starkey et al., 1990), ni tampoco un proceso restringido al ámbito de la percepción visual, sino que también abarca otros procesos perceptivos que, por ejemplo, dotarían al individuo de la capacidad para discriminar entre un número limitado de sonidos a desde los primeros años de vida (Starkey et al., 1983).

Por otro lado, tampoco esta habilidad sería una cualidad unívocamente humana, sino que también podría estar presente en otras especies animales (Agrillo, 2015; Beran et al., 2015). De hecho, algunos autores consideran que esta capacidad de conteo a simple vista es una característica seleccionada a lo largo de la evolución de las especies debido a la ventaja competitiva que otorga al animal en las tareas relacionadas con el aprovisionamiento de la comida, la maximización de energía dedicada a su búsqueda o la planificación de la caza (De Cruz, 2006).

La *subitación* sería, pues, una habilidad producida por la expresión conjunta de dos sistemas o módulos numéricos de procesamiento diferenciados (Feigenson et al., 2004). Por un lado, existiría un sistema encargado de diferenciar con precisión entre conjuntos de elementos pequeños, mientras que otro se encargaría de la función de discriminar, de modo aproximado entre colecciones grandes de elementos, siempre que exista una elevada diferencia proporcional entre ellos. Estas propuestas teóricas contarían, además, con el respaldo de estructuras cerebrales diferenciadas dependientes de cada sistema (ver Hyde & Spelke, 2011).

La representación numérica adquiere mayor grado de complejidad a medida que el niño desarrolla sus habilidades lingüísticas e incorpora las etiquetas (palabras o símbolos arábigos) que le permiten denominar, aunque de forma rudimentaria, un conjunto limitado de elementos (cardinalidad), y establecer un orden entre ellos (ordinalidad). A este respecto, algunos teóricos como Dehaene (1992; Dehaene & Cohen, 1995) señalarían la existencia de un triple código para el procesamiento numérico compuesto por:

1. *Una representación verbal* (grafémica o fonémica) para cada número, que se correspondería con las etiquetas verbales que se emplean (uno, dos, tres…) o su pronunciación.

2. *Una representación numérica*, de origen arábigo, que dispondría a los símbolos en una secuencia ordenada de dígitos (1, 2, 3…).

3. *Una representación analógica de la magnitud*: por medio de la cual, las magnitudes que representan los números se distribuirían a lo largo de una recta numérica mental. En esta línea numérica mental, los números quedarían ordenados de acuerdo a su magnitud, y las comparaciones entre ellos podrían llevarse a cabo estimando mentalmente su ubicación en la recta numérica (Laski & Siegler, 2007). Como señalan algunos autores, una deficiente capacidad para relacionar y situar los números en su representación abstracta a lo largo de una línea mental numérica podría estar detrás de ciertas dificultades que se pueden encontrar durante la adquisición de conocimientos matemáticos y operaciones de cálculo, lo que pronosticaría un bajo rendimiento en Matemáticas (Friso-van den Bos et al., 2015).

A los 2-3 años, la mayoría de los niños son capaces de enunciar los números y surgiría una cierta habilidad para contar verbalmente. Según Gelman y Gallistel (1978), esta habilidad incipiente aparece por la presencia de una serie de principios lógicos que actuarían como una guía para el desarrollo del conteo y su generalización a partir de las experiencias particulares que tienen en los distintos contextos de su entorno (Cahoon et al., 2017; LeFevre et al., 2009). Estos principios del *conteo* son:

1. *Principio de correspondencia uno a uno*: durante el conteo, cada objeto se designaría con una única etiqueta verbal, sea o no correcta la etiqueta.

2. *Principio de orden estable*: mediante el cual se establece una secuencia coherente de conteo que es invariable entre los distintos conjuntos que se contabilizan. Primaría la consistencia o coherencia entre conjuntos, frente a que la secuencia fuera o no correcta.

3. *Principio de cardinalidad*: establecería que el valor de la última etiqueta verbal representa la cantidad total del conjunto.

4. *Principio de abstracción*: establece que los principios anteriores son aplicables a cualquier tipo de conjunto, sean o no homogéneos sus elementos.

5. *Principio de irrelevancia*: el orden por el que se inicia el conteo es irrelevante para su designación cardinal.

1.1.3. Esquemas protocuantitativos y estrategias de conteo

Las experiencias que tienen los niños en su entorno posibilitan el aprendizaje de los distintos tipos de relaciones numéricas que pueden establecerse entre los elementos del conteo y que son relevantes para el posterior desarrollo de la aritmética. Estas relaciones lógicas conformarían los denominados *esquemas protocuantitativos* (Resnick, 1989, 1992) que permitirían establecer juicios sobre las cantidades sin necesidad de atender al hecho numérico concreto:

1. *Esquema de comparación*: mediante el cual se asignan etiquetas lingüísticas que permiten comparar magnitudes. Un ejemplo de ello podrían ser el uso de adverbios de cantidad como etiquetas de comparación: “más”, “menos”, “igual”.

2. *Esquema de incremento-decremento*: permitiría razonar sobre los cambios que se producen a partir del incremento o la reducción de las cantidades sin necesidad de una experiencia directa, previa o posterior con la realidad material.

3. *Esquema parte-todo*: permitiría reconocer que cualquier unidad puede descomponerse en piezas más elementales (partes), del mismo modo que la reagrupación de esas partes permitiría volver a constituir la unidad. Este sería el primer conocimiento adquirido sobre la propiedad aditiva de las cantidades.

Con la escolarización, se produce un incremento progresivo de conocimientos y estrategias en los niños que facilita el desarrollo y comprensión del número, el conteo y la aritmética. Las estrategias aritméticas pueden definirse como el conjunto de procedimientos utilizados para alcanzar un objetivo de nivel superior (Lemaire & Reder, 1999). Dados dos conjuntos de elementos que deben contarse, a cada tipo de situación de conteo le correspondería una estrategia de conteo distinta, encontrándose pues los siguientes tipos de estrategias que pudieren emplearse (Orrantia, 2006):

1. *Contar todo*: se aplicaría en situaciones aditivas de cambio y combinación.
    i. Se cuenta el número de objetos de uno de los conjuntos, o el conjunto inicial.
    ii. Se cuenta el número de objetos que se añaden.
    iii. Se cuenta el total de objetos para determinar el resultado
2. *Separar de: se aplicaría en situaciones sustractivas de cambio.*
    i. Se cuentan los objetos que representan el conjunto inicial.
    ii. Se quitan los objetos que especifica el conjunto de cambio.
    iii. Se cuentan los objetos que quedan para determinar el resultado.
3. *Igualación:* se aplica en situaciones aditivas en las que debe igualar la cantidad de ambos conjuntos siendo la diferencia desconocida.
    i. Se crean dos filas de objetos para representar ambos conjuntos
    ii. Se añaden objetos a la fila más pequeña hasta igualar a la fila más grande.
    iii. Se cuenta el número de objetos añadidos.
4. *Comparación: se aplica en situaciones sustractivas en las que no se conoce la diferencia entre los conjuntos:*
    i. Se crean dos filas de objetos para representar ambos conjuntos.

ii. Se cuenta el número de objetos no emparejados en la fila más grande.

Para llevar a cabo estas estrategias, es frecuente que los niños se apoyen en objetos o en sus propios dedos para representar el conteo, así como que verbalicen todo el proceso (Orrantia, 2006). Con el paso del tiempo, la maduración y la mejora en la eficiencia con la que se ejecutan los procesos, algunas de estas estrategias caerán en desuso en favor de otras más eficientes y sustentadas en la memoria (Geary & Hoard, 2003; Lemaire & Siegler, 1995). De ese modo, se desarrollará la capacidad de adecuar la estrategia más apropiada a cada problema a partir de las experiencias previas adquiridas hasta alcanzar el desarrollo de las competencias formales en la adolescencia.

Uno de estos cambios en las estrategias de conteo tiene que ver con el tránsito entre el conteo uno a uno de los elementos y el conteo realizado a partir del agrupamiento de los objetos en conjuntos discretos de elementos, una estrategia frecuentemente utilizada por los adultos (Aoki, 1977). Este cambio viene producido por el incremento en la eficiencia de conteo que proporciona el agrupamiento de elementos frente al conteo individualmente seriado de los elementos. A este respecto, Camos (2003) llevó a cabo un estudio que tenía por objetivo analizar el uso de las diferentes estrategias que se despliegan durante el conteo de puntos en función de la edad; en particular, desde la infancia a la edad adulta. Las estrategias de conteo observadas en este estudio fueron las siguientes:

1. *Conteo uno a uno*: consiste en identificar cada uno de los elementos de un conjunto con el cardinal que le corresponde, produciendo así una secuencia de números en la que el último de sus números naturales se identifica con el total del conjunto.
2. *Conteo en base a n*: siendo *n* un número concreto que puede oscilar entre 2 y 6, generalmente. Se identificaría con una estrategia en la que el participante llevaría a cabo, por ejemplo, un conteo de dos en dos elementos (Ej. 2, 4, 6, 8…), o de tres en tres (Ej. 3,

6, 9, 12…), etc. Este tipo de estrategia vendría también facilitada por la adquisición de conocimientos sobre *hechos aritméticos*, como son -por ejemplo- las tablas de multiplicar.

3. *Conteo por adición*: en esta estrategia, el número total de elementos se obtendría sumando sucesivamente grupos de puntos, ya sea diciendo explícitamente las operaciones en curso (Ej: tres más dos, cinco; más dos…) o dando sólo los resultados intermedios de las operaciones (Ej. tres, cinco, siete). La diferencia con la estrategia anterior de conteo en base a un número concreto es que en esta estrategia el número con el que se incrementa la suma de elementos puede variar durante el transcurso del conteo, mientras que en la estrategia anterior el incremento se fija previamente al inicio de la secuencia de conteo.
4. *Conteo por multiplicación*: consiste en agrupar los elementos de un conjunto en subgrupos y multiplicar el número de subgrupos por su dimensión o número de elementos que lo componen, obteniéndose así la cifra total del conjunto.

Estas estrategias básicas de conteo podrían, asimismo, combinarse entre sí durante la secuencia de conteo, formando una secuencia estratégica diversa y de mayor complejidad.

Los resultados obtenidos por esta autora indicaron que los participantes más jóvenes tendían a utilizar una estrategia de conteo de tipo uno a uno (88%), que se reducía al 24% alcanzada la edad adulta. A partir de la adolescencia, el resto de estrategias incrementaban su frecuencia de uso, siendo la estrategia de conteo en base a un número concreto (de dos en dos, de tres en tres, etc.) la estrategia más frecuentemente utilizada, incluso, entre los adultos. Las estrategias de adición (8%) y multiplicación (14%) serían estrategias más frecuentemente utilizadas por adolescentes y adultos.

Todos estos incrementos y decrementos en la frecuencia de utilización de las distintas estrategias de conteo se explicarían en base a la ganancia en la eficiencia en el conteo (menor tiempo invertido y menor número de errores) que conlleva el uso de unas estrategias frente a otras, así como por la adecuación del uso de la estrategia más apropiada para las características específicas de la tarea, una noción adquirida a través de la práctica y la experiencia acumulada. No obstante, existirían también otros cambios en las estrategias de conteo derivados de la maduración y el envejecimiento que podrían conformar perfiles distintos para adultos mayores y adultos jóvenes. Estos cambios afectarían al repertorio de estrategias que disponen, la distribución de las estrategias, su ejecución y su selección (ver revisión en Uittenhove & Lemaire, 2015; también, Lemaire, 2010).

Uno de los vehículos principales a través del cual tiene lugar la transmisión y/o el descubrimiento de las estrategias aritméticas son los problemas aritméticos. La resolución de problemas permite un entorno apropiado, cotidiano, dinámico y diverso para el que es necesario aplicar procedimientos aritméticos básicos con el fin de resolver la situación problemática planteada. El tipo de problemas, sus criterios de clasificación y sus estrategias de resolución serán objeto de estudio en el siguiente epígrafe.

1.2. Problemas aritméticos verbales: clasificación, modelos teóricos y estrategias de resolución

Cuando el pensamiento se dirige explícitamente hacia un propósito determinado, podemos decir que estamos frente a una situación de resolución de problemas (Johnson-Laird, 1988). Los problemas matemáticos constituyen una simulación controlada de una situación de incertidumbre para la que debe encontrarse una solución aplicando una secuencia concreta de cálculo. El acto de resolver un problema representa la búsqueda de las acciones que son apropiadas para lograr un propósito (Pólya, 1981). Esta respuesta es el resultado de un complejo proceso que se ve afectado por los conocimientos matemáticos que dispone el resolutor (Fyfe et al., 2012; Orrantia, 2003), sus habilidades cognitivas (Friso-van den Bos et al., 2013; Peng et al., 2019), las estrategias que posee (Gick, 1986; Mayer, 1992) y su control metacognitivo (Desoete et al., 2001; Fletcher &

Carruthers, 2012; Gómez-Chacón et al., 2014). Esto los convierte en un instrumento esencial no sólo en el ámbito educativo, sino también desde el punto de vista experimental, ya que ofrecen un contexto de alta validez ecológica en el cual es posible modular el tipo y grado de dificultad que entraña su resolución.

1.2.1. Clasificación de los problemas aritméticos verbales

Existe una gran diversidad de problemas aritméticos verbales (AWP). Una de las clasificaciones más aceptadas es la realizada por Riley y colaboradores (1984). Estos autores clasificarían los AWP en función del tipo de proceso que describe el enunciado del problema, estableciendo 4 tipos de categorías generales: cambio, combinación, comparación e igualación (ver tabla 1).

Los problemas de *cambio* se corresponderían a aquellos en los que se establece un estado inicial, un estado intermedio de transformación y un estado final. Existirían, por tanto, tres tipos de problemas de cambio diferentes, en función de si la pregunta o incógnita del problema alude al estado inicial, al estado final, o al proceso de cambio que se efectúa.

En relación a los problemas de *combinación*, podrían encontrarse de dos tipos en función de si la pregunta planteada en el problema alude al cálculo total combinado de los elementos del problema, o al cálculo del valor de una de las partes a partir de la combinación del valor de la parte restante y el total.

Por último, los problemas de *comparación* se clasificarían en función del tipo de relación -comparativa- que se establezca entre los elementos del problema. Así, el problema podría describir dos cantidades conocidas y requerir que se determine la diferencia entre ellas (problemas de diferencia desconocida), o determinar una cantidad desconocida a partir de la relación que se describa con el elemento de magnitud conocida.

Tabla 1: Clasificación de los de problemas aritméticos verbales (extraído de Riley et al., 1984, p.160)

PROBLEMAS DE ACCIÓN	PROBLEMAS ESTÁTICOS
CAMBIO *Resultado desconocido* 1. José tenía 3 canicas. Entonces Tomás le dio 5 canicas más. ¿Cuántas canicas tiene ahora José? 2. José tenía 8 canicas. Luego le dio 5 canicas a Tomás. ¿Cuántas canicas tiene ahora José? *Cambio desconocido* 3. José tenía 3 canicas. Luego Tomás le dio más canicas. Ahora José tiene 8 canicas. ¿Cuántas canicas le dio Tomás? 4. José tenía 8 canicas. Luego le dio dos canicas a Tomás. Ahora José tiene 3 canicas. ¿Cuántas canicas le dio a Tomás? *Inicio desconocido* 5. José tenía algunas canicas. Entonces Tomás le dio 5 canicas más. Ahora José tiene 8 canicas. ¿Cuántas canicas tenía José al principio? 6. José tenía sorne canicas. Luego le dio 5 canicas a Tomás. Ahora José tiene 3 canicas. ¿Cuántas canicas tenía José al principio?	COMBINACIÓN *Combinar valor desconocido* 1. José tiene 3 canicas. Tomás tiene 5 canicas. ¿Cuántas canicas tienen en total? *Subconjunto desconocido* 2. José y Tomás tienen 8 canicas en total. José tiene 3 canicas. ¿Cuántas canicas tiene Tomás? COMPARACIÓN *Diferencia desconocida* 1. José tiene 8 canicas. Tomás tiene 5 canicas. ¿Cuántas canicas tiene José más que Tomás? 2. José tiene 8 canicas. Tomás tiene 5 canicas. ¿Cuántas canicas tiene Tomás menos que José? *Cantidad comparada desconocida* 3. José tiene 3 canicas. Tomás tiene 5 canicas más que José. ¿Cuántas canicas tiene Tomás? 4. José tiene 8 canicas. Tomás tiene 5 canicas menos que José. ¿Cuántas canicas tiene Tomás?

Tabla 1 (continuación): Clasificación de los problemas aritméticos verbales (extraído de Riley et al., 1984, p.160)

PROBLEMAS DE ACCIÓN	PROBLEMAS ESTÁTICOS
IGUALACIÓN 1. José tiene 3 canicas. Tomás tiene 8 canicas. ¿Qué podría hacer José para tener tantas canicas como Tomás? (¿Cuántas canicas necesita José para tener tantas como Tomás?). 2. José tiene 8 canicas. Tomás tiene 3 canicas. ¿Qué podría hacer José para tener tantas canicas como Tomás?	COMPARACIÓN (continuación) *Referente desconocido* 5. José tiene 8 canicas. Él tiene 5 canicas más que Tomás. ¿Cuántas canicas tiene Tomás? 6. José tiene 3 canicas. Él tiene 5 canicas menos que Tomás. ¿Cuántas canicas tiene Tomás?

No obstante, los AWP también podrían categorizarse en base a otro tipo de criterios que aluden a los objetos que aparecen en el problema, las cantidades que relacionan a los elementos del problema, la pregunta que plantea el problema, las operaciones necesarias para su resolución o el tipo de respuesta que producen (ver tabla 2).

Respecto a los *elementos de un problema*, uno de los factores de clasificación es el número de elementos que aparecen en el enunciado. Así, cuanto mayor sea el número de elementos significativos que aparecen en el problema, mayor será el número de proposiciones que contiene, lo que incrementará la dificultad de comprensión y, quizá, el número de pasos necesarios para su resolución. No obstante, este criterio podría manipularse también con el objetivo de insertar información irrelevante en el enunciado e incrementar la necesidad de inhibición de respuesta de los participantes (Marzocchi et al., 2002; Re et al., 2016).

De modo similar, el *tipo de contexto* -familiar o no familiar- en el que están insertados los elementos que se describen en el problema puede afectar a la resolución del mismo, especialmente entre los alumnos más pequeños (Stern & Lehrndorfer, 1992; Vicente et al., 2008).

Tabla 2: Criterios de clasificación de los problemas aritméticos verbales (Duque de Blas et al. 2021).

Criterio	Descripción
Tipos generales de problemas	Problemas de cambio, combinación, comparación o igualación.
Elementos nombrados en el problema	
- Número de elementos	Incrementa la dificultad de comprensión, y se relacionaría con los pasos para resolverlo o la presencia de información distractora.
- Contexto de los elementos	Familiar o no familiar para quien resuelve el problema.
- Alineación semántica entre elementos	Alineación simétrica o asimétrica.
Cantidades relacionadas con los elementos	
- Magnitud	Grande o pequeña.
- Tipo de datos	Números enteros o decimales.
- Cómo se representan	Cardinales (entidades no ordenadas) u ordinales (entidades ordenadas).
- Cómo se expresan	De forma explícita y/o relacional.
- Términos que expresan cantidad	
* Adverbios de cantidad	"Más que"; "menos que", "igual a"; "tanto como".
* Verbos	“Ganar”, “perder”, “ingresar”, “extraviar”, etc.
* Sustantivos	“Ganancia”, “pérdida”, “ingreso”, “extravío”, etc.
- Sentido del término verbal insertado	Consistente o inconsistente con la operación sugerida.
Pregunta del problema	
- Número de preguntas	Una o más preguntas.
- Ubicación de la pregunta	Al principio o al final del problema.
- Tipo de pregunta	Referida al resultado global de alguna parte específica, o al conjunto de las partes implicadas.
- Datos solicitados	Datos numéricos o cualitativos.
Operaciones	
- Tipos	Suma, resta, multiplicación y/o división.
- Número o pasos	Desde 1 en adelante

Tabla 2 (continuación): Criterios de clasificación de los problemas aritméticos verbales (Duque de Blas et al. 2021).

Criterio	Descripción
Tipo de respuestas producidas	
- Modalidad de la respuesta	Generación de respuesta o selección de alternativas
- Precisión de la respuesta	Respuestas correctas, superficiales o respuestas incorrectas no-superficiales.

En la misma línea, si los elementos descritos en los problemas están *semánticamente alineados* entre sí (Ej. enunciar "margaritas" y "tulipanes" evoca englobarlos en un conjunto "flores", lo que sugiere una operación aditiva), su eficacia en la resolución se verá facilitada en comparación con aquellos problemas en los que hay una relación semántica desalineada entre los elementos (Bassok et al., 1998; Gros et al., 2020).

Otro criterio de clasificación haría referencia a las *cantidades descritas* en el problema. Este factor puede afectar de forma diferencial a la precisión en la resolución de los problemas en función de la edad de la persona a cargo de la resolución del problema, así como en el tipo de estrategia que utiliza para resolverlo (Thevenot & Oakhill, 2005).

Sobre el *tipo de datos* que se presentan, las magnitudes decimales serán una fuente de dificultad añadida a los problemas, en comparación con las magnitudes enteras (Bell et al., 1984; Harel et al., 1994). Del mismo modo, la cardinalidad u ordinalidad de las magnitudes presentadas constituye un criterio frecuentemente manipulado por algunos teóricos del campo de la semántica por su capacidad para sugerir una operación aritmética determinada para su resolución (ver Gros et al., 2019, 2020).

El *modo en que se explicitan las magnitudes* en el problema alude a si éstas son expresadas de forma directa (Ej. Juan tiene 5 canicas) o de forma relacional (Ej. Juan tiene el doble de canicas que de caramelos). Una magnitud expresada directamente no requeriría una operación aritmética adicional, por lo que la resolución del problema resultará menos exigente.

En relación a los términos que se utilizan para expresar cantidades y sus variaciones, podemos encontrar un conjunto variado de fórmulas léxicas para denotarlo. Una de ellas es mediante la utilización de *adverbios de cantidad, los verbos* que hay en el enunciado del problema (“más que”, “menos que”, “igual que”, “tanto como”) que pueden indicar una relación de cambio o comparación entre dos sujetos u objetos del enunciado. Este aspecto resulta de enorme trascendencia, ya que de su correcta interpretación teniendo en cuenta el contexto en que se encuentre, dependerá el tipo de operación aritmética necesaria para la resolución del problema. Así, podemos encontrar dos tipos de problemas en función del sentido de los adverbios utilizados: a) *problemas coherentes o consistentes*, aquellos en los que el adverbio de cantidad "más" o "menos" utilizado coincide con la operación de adición o sustracción, respectivamente, que debe ser realizada para la resolución; y b) *problemas no coherentes o inconsistentes*, aquellos en los que es necesario realizar la operación aritmética inversa a la sugerida literalmente por el adverbio de cantidad; esto es, realizar una resta en presencia del término "más", o una suma en presencia del término "menos". En consecuencia, los problemas inconsistentes requerirán un esfuerzo cognitivo mayor que los consistentes, lo que afectará a la precisión de respuesta de los participantes en la tarea.

Otra fórmula para expresar magnitudes y transformaciones de magnitudes es mediante verbos, como “ganar”, “perder”, “obtener”, “extraviar”, etc., que implican cambios en las cantidades expresadas en el enunciado del problema y que pueden ser formalizados en términos de operaciones aritméticas. No obstante, es posible expresar también dichos cambios mediante la sustantivación de estos verbos (“ganancia”, “pérdida”, “ingreso”, “extravío”, etc.). Este aspecto resulta relevante, ya que puede influir en el tiempo de reacción de los participantes a la hora de resolver el problema (Daroczy et al., 2020).

El siguiente criterio de clasificación de los AWP tiene que ver con *la pregunta que se realiza* en el problema. Los problemas pueden contener una o más preguntas que deben ser contestadas, y que pueden referirse a cantidades generales resultantes de una operación aritmética entre dos conjuntos (Ej. Juan tiene 3 canicas y Pepe 5. ¿Cuántas canicas tienen entre los dos?); o a resultados parciales (Ej. Juan tiene el doble de canicas que Pepe. Entre los dos tienen 9 canicas. ¿Cuántas canicas tiene Pepe? ¿Cuántas canicas tiene Juan?). En este segundo caso, el número de preguntas no redundantes del problema sería igual al número de pasos intermedios que requiere para su resolución. Este aspecto puede ser relevante desde el punto de vista experimental si lo que se desea es tener una información más exhaustiva de la eficacia de resolución del participante a lo largo de los distintos pasos intermedios que requiere el problema.

Respecto a la *posición en que se encuentra situada la pregunta* del problema, podría localizarse al principio o al final del enunciado. En caso de que el problema se iniciase con la pregunta, se impediría que persona que resuelve realice los cálculos intermedios que se requirieran a medida que va leyendo el enunciado, obligándole a mantener el contenido de la pregunta hasta el final (Thevenot & Oakhill, 2005). Este aspecto podría incrementar el efecto de recencia en la tarea.

Otro aspecto relacionado con la pregunta hace referencia al *tipo de datos que solicita* el problema. Es frecuente que los AWP demanden una respuesta numérica a los problemas. Sin embargo, también podrían solicitar, por ejemplo, una respuesta de tipo cualitativo (Ej. "Si Juan tiene 5 canicas y Pepe la mitad, ¿quién tiene más canicas?).

En relación con las *operaciones que requiere la resolución del problema,* estos podrían ser categorizados en función del tipo y número de operaciones aritméticas de adición, sustracción, multiplicación y/o división que pudieren requerir. El tipo de operaciones aritméticas necesarias para la resolución del problema podría afectar a la eficacia de la resolución, especialmente entre los participantes de menor edad. Un ejemplo de ello podría encontrarse en los problemas con operaciones sustractivas, que pueden ser resueltas por los adultos más eficientemente como resultado de llevar a cabo un procedimiento inverso aditivo (Campbell, 2008; Campbell & Alberts, 2010; Orrantia et al., 2012). Asimismo, el número de operaciones incrementaría la dificultad de los problemas, ya que implican una secuencia de computación más extensa, lo que conlleva un mayor número de "ventanas de oportunidad" para la comisión de errores de cálculo.

El último criterio general en que los problemas aritméticos verbales pueden clasificarse tiene que ver con la *respuesta de resolución producida por los participantes*. Así, en base a este criterio los AWP pueden diferir en dos aspectos principalmente. El primero tiene que ver con *modalidad de la respuesta* del problema, encontrándose dos posibles alternativas: la "selección" y la "generación" de respuesta. En la *selección de respuesta*, el problema ofrece a la persona a cargo de la resolución un conjunto discreto de alternativas de respuesta para que pueda identificar y seleccionar aquella que es correcta tras haber aplicado su particular secuencia de cálculo. En cambio, en la modalidad de respuesta de tipo *generativo*, el participante debe producir su propia respuesta a partir de las operaciones y los resultados obtenidos tras la secuencia de cálculo aplicada sin que le sean suministradas las posibles alternativas de respuesta en el problema.

La modalidad de respuesta requerida por el problema es un aspecto relevante, ya que puede afectar, posteriormente, a la proporción que se obtiene de cada tipo de respuesta en el problema y el tipo de estrategia utilizada (Katz et al., 2000). Así, mediante una modalidad de *selección de respuesta*, se acota el conjunto de respuestas plausibles al problema (2, 3, 5… alternativas), reduciendo la incertidumbre del participante. Este hecho es determinante en la orientación de la respuesta y la conducta del participante, ya que podría inducir un proceso de autorregulación de la respuesta en caso de que ésta no se adecuase a alguna de las soluciones ofrecidas por el problema, lo que iniciaría un nuevo ciclo de resolución que permita identificar la respuesta correcta entre las opciones que dispone. En tal caso, la modalidad de selección de respuestas reduciría, probablemente, la proporción de respuestas incorrectas cometidas por el participante, en favor de alguna de las alternativas, ya sea correcta o superficial. Por el contrario, en la modalidad de *generación de respuestas*, la proporción de respuestas incorrectas y su diversidad dentro del conjunto de respuestas incorrectas se verían probablemente incrementados.

El segundo y último criterio en que los AWP pueden clasificarse tiene que ver con la *precisión de la respuesta* que da el participante al problema. A este respecto, podremos identificar 3 tipos de respuestas principales que pudieren obtenerse en función de las características particulares del problema: respuesta correcta, respuesta superficial y respuesta incorrecta no-superficial.

- La respuesta correcta se identificaría con la solución precisa del problema planteado, lograda tras haber aplicado una secuencia aritmética ajustada y adecuada a las características y contexto del problema.
- La respuesta superficial sería un tipo de respuesta incorrecta que podría encontrarse en los problemas en los que el sentido del término verbal insertado está expresado de manera *inconsistente*. Una interpretación literal de los adverbios de cantidad ("más que - suma" o "menos que - resta") provocaría que el participante aplicara un procedimiento de cálculo de suma o resta *consistente* con lo sugerido por dicha partícula verbal, en lugar de la operación inversa requerida para la correcta resolución del problema ("más que – resta" o "menos que – suma"). Esto conduciría al participante a la comisión de un *error esperable* y relativamente frecuente denominado "respuesta superficial". El número total de respuestas superficiales que pueden existir en un AWP inconsistente estaría relacionado de manera exponencial con el número de operaciones inconsistentes de un problema (n). Dicha relación se expresaría del siguiente modo:

$$\textit{Número de respuestas superficiales} = (2^{n}) - 1$$

La base 2 del exponente de la función identifica las dos posibles alternativas a la hora de interpretar el sentido del adverbio de cantidad insertado en el enunciado del problema inconsistente: de manera "correcta-inversa" o de manera "incorrecta-literal". Con la sustracción de la unidad en la función, quedaría descartada la respuesta correcta del conjunto de respuestas, dando como resultado el número total de respuestas superficiales.

También es posible encontrar respuestas de tipo superficial en aquellos problemas aritméticos que, por la estructura y/o contenido de su enunciado, son capaces de generar una representación semántica inadecuada que induzca una respuesta errónea esperable y frecuente como consecuencia de la aproximación heurística del participante a la tarea. Un ejemplo de ello puede observarse en los problemas de reflexión cognitiva (Frederick, 2005).

- Las respuestas incorrectas no-superficiales se identificarían con cualquier elemento del conjunto indeterminado de respuestas que pudiere dar el participante y que no se corresponden ni con las respuestas correctas, ni con las respuestas superficiales.

1.2.2. Etapas de la resolución de problemas aritméticos

La resolución de problemas aritméticos es una tarea compleja en la que tienen lugar una gran diversidad de procesos secuenciales para lograr producir una respuesta a la pregunta planteada.

Mayer, en su libro *Thinking, problem solving, cognition* (1992), cita como una de las primeras descripciones del proceso de resolución de problemas la aportación realizada por el economista inglés Graham Wallas (1926), entendida como una secuencia ordenada en la que se distinguirían las siguientes fases:

1. *Preparación*: la persona encargada de resolver el problema inicia su actividad recopilando toda la información relativa al propósito y contenido del problema para efectuar una tentativa preliminar de solución.

2. *Incubación*: en esta fase, la persona que resuelve se desentiende momentáneamente de la tarea de resolución para llevar a cabo otro tipo de tareas, tales como pasear, dormir, trabajar, etc., mientras que de forma latente el proceso de resolución continua en su mente hasta que le sobreviene la siguiente etapa.

3. *Iluminación*: denominada también *insight,* en esta fase la clave del problema se revelaría ante la persona como la solución a la situación de incertidumbre que plantea.

4. *Verificación*: correspondería con la etapa en la que la respuesta revelada se revisa y comprueba su validez como solución al problema planteado.

De modo muy similar, pero esta vez basándose en la experiencia docente adquirida, Pólya (1945) describió la solución a un problema como un proceso en el que se describen las siguientes etapas:

1. *Comprender el problema*: haría referencia a un periodo en que la persona recopila toda información relacionada con el problema y se pregunta “qué es lo que se quiere saber” y “cuáles son los datos y restricciones”.

2. *Trazar un plan de resolución*: en esta etapa, la persona que resuelve utilizaría su experiencia pasada con el objetivo de encontrar una estrategia que, aplicada sobre la situación problemática presente, le sirviera como modelo para resolver el problema.

3. *Ejecutar el plan de resolución*: se efectúa la secuencia procedimental de cálculo determinada en la fase anterior, así como la supervisión de los resultados intermedios que se van obteniendo en cada paso.

4. *Comprobación del resultado o mirar atrás:* en esta etapa se verifica el resultado obtenido con el objetivo de comprobar si se ajusta como explicación razonable al problema planteado.

Otra propuesta más exhaustiva acerca de las etapas de resolución de problemas es la planteada por Mayer (1992), quien definiría el proceso como una sucesión de actividades mentales diferenciadas en base al tipo de conocimiento específico que se requiere de la persona para poder llevarlas a cabo. Estas actividades tendrían por objeto las siguientes tareas:

1. *Traducción del problema*: hace referencia a la recodificación de cada elemento del enunciado del problema en una representación mental. Requeriría el uso de un *conocimiento de tipo lingüístico* del idioma en que se describe el problema, así como un *conocimiento semántico general* acerca del funcionamiento de las cosas con el que poder generar una representación precisa de los vínculos relacionales que se establecen entre los distintos sujetos y objetos descritos del problema.

2. *Integración del problema*: mediante la cual se logra un modelo situacional del problema que permite caracterizarlo en función del tipo de problema específico que corresponda (cambio, igualación, comparación o combinación). Cada tipo de problema sugeriría un procedimiento concreto de resolución. Es por ello que el *conocimiento esquemático* que se posea de cada tipo de problema facilita la tarea de integración del modelo situacional y la resolución, dado que muchos de los errores que se comenten tienen que ver con la aplicación de un esquema inapropiado para el tipo de problema que se plantea. El conocimiento esquemático ayudaría también a fijar la atención hacia los rasgos estructurales del problema, una característica común que diferencia a los sujetos expertos de los inexpertos, que tenderían a prestar más atención a los rasgos superficiales, tales como las cantidades y la presencia de palabras clave dentro del problema.

3. *Planificación y monitoreo de la solución*: consiste en elaborar un plan para la resolución del problema, merced al *conocimiento estratégico* que dispone la persona que resuelve. Este proceso, como se anticipó parcialmente en las etapas descritas por Pólya, dependería de varios aspectos relacionados:

i. Disponibilidad de un problema semejante que pudiera servir de modelo inicial para la resolución.

ii. Reescritura del problema; una estrategia que ayudaría a explicitar todas las relaciones contenidas en el enunciado del problema y facilitar la comprensión.

iii. Descomposición del problema en submetas, de modo que permita una supervisión de cada uno de los pasos necesarios para la resolución.

4. *Ejecución de la secuencia de solución*: esta última etapa requeriría necesariamente un *conocimiento* de tipo *procedimental*, que correspondería con las habilidades de cálculo que dispone la persona para resolver el problema.

1.2.3. Modelos teóricos de resolución de problemas aritméticos verbales

Los problemas aritméticos verbales (AWP) han dado lugar a numerosos estudios y modelos explicativos acerca de los procesos implicados en su resolución, bien dentro del ámbito de la psicología, bien en el de la ciencia computacional. Entre los modelos más destacados se encontrarían dos modelos clásicos para la resolución de problemas, el modelo de la Gestalt y el modelo general de Newell y Simon para la resolución de problemas (1972); así como dos modelos específicos de resolución de problemas aritméticos, basados en el esquema y la representación del modelo situacional.

1.2.3.1. Modelo de la Gestalt para la resolución de problemas

La Gestalt (Koffka, 1922; Köhler, 1967; Wertheimer, 1938) fue una corriente psicológica opuesta al conductismo americano de principios del siglo XX. Uno de los aspectos diferenciadores entre la Gestalt y las corrientes asociacionistas es que consideraba que la representación mental generada a partir de la percepción de la realidad material no puede entenderse como una simple suma de las partes que la componen, del mismo modo que una pieza musical no constituye un simple conjunto discreto de notas y sus repeticiones, ya que a partir de las mismas notas pueden crearse infinitas piezas distintas (González-Labra, 2019). Es la estructura global de la pieza, generada a partir de las interrelaciones que establecen las notas entre sí y el lugar que ocupan, la que permite caracterizarla como una entidad global diferenciada. Siguiendo esta misma analogía, para la Gestalt la representación mental de la realidad tiene lugar porque el pensamiento aplica sobre el objeto o fenómeno percibido una estructura concreta. Esta estructura que permite interpretar la realidad se genera a partir de las experiencias que va adquiriendo la persona en su proceso interactivo con el ambiente, y que define el modo mediante la cual representamos las cosas o resolvemos un problema. Las experiencias predisponen a la mente del sujeto hacia una configuración específica de los elementos para su interpretación, de modo que su percepción, y no al revés, sería la consecuencia de haber aplicado una forma -o Gestalt- ya previamente establecida.

Un aspecto clave de la teoría de la Gestalt relacionado con la solución de problemas es la organización y reorganización de los elementos de un problema, un aspecto determinado por el *principio de pregnancia* que establece que las cosas tienden a percibirse de aquel modo que maximice su simplicidad y coherencia. Esta tendencia del sistema cognitivo permitiría establecer una secuencia procedimental que, partiendo de un estado inicial, permitiera llegar a un estado meta u objetivo. Sin embargo, en función de las características de los problemas, podría inducirse una estructura organizativa que limite o imposibilite la resolución del mismo, lo que requeriría que la persona que resuelve llevara a cabo un procedimiento de reorganización de sus elementos para que, percibidos desde un punto de vista estructural alternativo, permitiera encontrar una alternativa de resolución que fuera eficaz y creativa. Esta cualidad determinaría la diferencia, según Wertheimer, entre las dos actitudes o aproximaciones características que podría mostrar una persona frente a una tarea de solución de problemas; una, basada en el *pensamiento* *reproductivo* y automático, que se sustenta sobre las experiencias previas del individuo y sus conocimientos procedimentales. Estas estructuras predispondrían al sujeto hacia una *fijeza funcional* de su conducta que se operativizaría en una estrategia o secuencia procedimental o metodológica específica. La otra, basada en el *pensamiento productivo,* un proceso activado a voluntad gracias al cual, y tras un periodo de *incubación,* sería posible experimentar una vivencia de revelación de la clave del problema, también llamado *insight*, merced a la cual los elementos y restricciones del problema se *reorganizarían* en torno a una estructura diferente que surge tras haber comprendido en profundidad y de manera súbita la situación que se plantea y las relaciones dinámicas que se establecen entre los elementos (ver Duncker, 1945).

1.2.3.2. Teoría de Newell y Simon de resolución de problemas

Otra aproximación al proceso de resolución de problemas es la teoría, procedente de las ciencias de la información, de Newell y Simon. Los AWP, en tanto problemas que son, compartirían los mismos componentes y procesos descritos por Newell y Simon (1972) para la resolución de cualquier tipo de problema general. Estos autores considerarían que la resolución de un problema se sustenta, de forma análoga a un computador, en la habilidad para el manejo de símbolos de la persona encargada de resolver la tarea. Este hecho resulta de especial relevancia, pues pone de manifiesto que el proceso de resolución podría ser, en consecuencia, simulado mediante algoritmos en un computador. Para estos autores, todo problema consta de tres elementos o partes fundamentales:

1. *El sistema encargado de procesar el problema*: haría referencia a las capacidades intelectuales de la persona que lo resuelve y su habilidad para el manejo de símbolos con los que codifica la representación mental del problema.

2. *El ambiente de la tarea*: se refiere a la estructura propia del problema planteado, de modo que ésta podría determinar la eficacia de resolución del mismo.

3. *El espacio del problema*: se identificaría con la representación mental que la persona elabora del problema y sobre la que se operará hasta alcanzar la solución.

El proceso de resolución de problemas comprendería dos etapas diferenciadas en las que se llevarían a cabo las siguientes tareas. En primer lugar, un *proceso de comprensión* profunda de la situación de incertidumbre que plantea el problema y sus restricciones. Este proceso estaría determinado por el espacio del problema, es decir, por una representación a partir de la cual se establecería un estado inicial de partida y un estado final o meta. En segundo lugar, se llevaría a cabo el propio *proceso de solución* del problema, en el que la persona que resuelve determinaría las estrategias procedimentales necesarias para transitar desde un estado inicial hacia un estado final.

La secuencia de procedimientos que se establece podría no tener éxito como consecuencia de no disponer de una estrategia específica que aplicar para la resolución del problema concreto. Esta situación obligaría a la aplicación de una estrategia de tipo globalista o heurístico de respuesta, denominado *medios-fines,* mediante la cual se descompone la meta final en un conjunto de submetas diferentes para, a través de sucesivos ensayos y errores, aproximar su respuesta en cada iteración a la meta fijada de inicio. Si dicha estrategia fuera reiteradamente ineficaz, se reactivaría un nuevo ciclo de comprensión con el fin de reevaluar el espacio del problema e incorporar nuevos aspectos que le permitieran emprender un nuevo proceso tentativo de resolución más ajustado a las condiciones y el tipo de problema.

1.2.3.3. Teorías basadas en el esquema

Otro tipo de teorías sobre el proceso de resolución de problemas serían las teorías basadas en la generación, activación e ejecución de esquemas (Kintsch & Greeno, 1985; Rumelhart, 1980). Un esquema es una estructura de datos proposicionales almacenado en la memoria a largo plazo que se genera a partir de los repetidos encuentros con aquellos problemas que comparten una estructura proposicional semejante. La función de un esquema sería ofrecer un marco conceptual para la clasificación e interpretación de las representaciones mentales. Estas estructuras operativas, una vez generadas y ejecutadas, serían las responsables de proporcionar un algoritmo válido de resolución del problema. Con la lectura del enunciado del problema, la entrada verbal se transformaría en una representación conceptual de su significado en formato proposicional. Esta representación activaría un proceso de recuperación de la memoria a largo plazo de aquel esquema que compartiera la misma estructura proposicional del problema para, posteriormente, contextualizarlo con los datos numéricos concretos del enunciado y poder aplicar el proceso de resolución específico asociado al esquema recuperado.

Las teorías basadas en el esquema fueron muy populares en los años 80. No obstante, diferentes estudios han demostrado cómo pequeñas variaciones en los enunciados de los problemas (por ejemplo, que las cantidades que aparezcan sean de tres dígitos en lugar de dos) pueden producir cambios en las estrategias que se emplean en la resolución, afectando así a la eficacia con la que se resuelven. Este hecho indicaría, en consecuencia, que un tipo de esquema diferente estaría siendo aplicado, lo que constituye una demostración de que el esquema no sería, por tanto, el marco universalmente válido y aplicable a estructuras proposicionales idénticas que se preconiza (Thevenot & Oakhill, 2005). Del mismo modo, otros efectos de contenido que se producen en los problemas, como el que sucede con su reescritura (Vicente et al., 2008), tampoco encontrarían acomodo en las teorías basadas en esquema.

1.2.3.4. Teoría del modelo situacional y estrategias de resolución

En relación al modelo situacional, esta propuesta teórica supone la aplicación del modelo de construcción-integración de la comprensión de textos a la resolución de problemas aritméticos verbales (Kintsch, 1998). La teoría del modelo situacional pone de manifiesto que, del mismo modo en que las habilidades de cálculo son indispensables para la resolución de los problemas, también es necesario considerar las diferentes formas en que los enunciados pudieren entenderse para poder determinar el éxito del proceso. Esta observación vino fundamentada por la creciente evidencia de que los errores más frecuentes que se cometían en la resolución de los AWP eran debidos a una construcción inadecuada de la representación mental del problema, en lugar de errores causados por una incorrecta ejecución de la secuencia procedimental (Sternberg, 1996). La resolución de los AWP requeriría, por tanto, de una comprensión en profundidad del enunciado del problema, especialmente si los términos verbales que aparecen en los problemas establecen relaciones que son *inconsistentes* con las operaciones aritméticas que sugieren. Supongamos el siguiente problema:

> *En SportGood, una sudadera de la marca SP cuesta 50 euros. Esto es, 3 euros menos de lo que hay que pagar por la misma sudadera en Primark. Si necesitas comprar 4 sudaderas SP, ¿cuánto tendrás que pagar en Primark?*

La dificultad de este problema de una operación inconsistente y una multiplicación adicional radicaría en determinar la operación aritmética de suma o resta que hay que aplicar en el primer paso para obtener el precio de una sudadera en Primark. El término comparativo utilizado en el problema "menos" sugeriría realizar una operación de resta. Sin embargo, por el contexto en el que se encuentra insertado el adverbio de cantidad, la operación necesaria para determinar el valor de la sudadera en Primark es en este caso la inversa a la sugerida -una suma-, dando así lugar a la siguiente secuencia correcta de cálculo:

Precio de 4 sudaderas en Primark = (50 + 3) x 4

Estos problemas permitieron identificar que la persona que resuelve el problema puede emplear dos tipos estrategias de resolución diferenciadas. Por un lado, podría aplicarse una estrategia directa que, en palabras de Siegler & Jenkins (1989), podría resumirse en "calcular primero, pensar después". El proceso se iniciaría identificando los números de cada categoría con los que, posteriormente, se realizan las operaciones aritméticas estimadas como necesarias en base a las relaciones literales del enunciado del problema. Esta estrategia de traducción directa conduciría frecuentemente a la comisión errores en la resolución de los problemas debido fundamentalmente a dos motivos:

1. En primer lugar, por una comprensión deficiente de lo que representa el problema, centrada más en los aspectos cuantitativos que en los cualitativos (Chi et al., 1988; Mayer et al., 1992; Sternberg & Frensch, 1992).

2. En segundo lugar, por una estimación inadecuada de su dificultad, quizá como consecuencia de identificar el problema como un ejercicio rutinario para el cual ya se dispone de un modelo o plan de resolución previos. Esta estrategia estaría dominada por los procesos intuitivos y automáticos del Sistema1 de razonamiento (Evans & Stanovich, 2013a; Stanovich, 2006).

La segunda estrategia para la resolución de problemas aritméticos, denominada “estrategia del modelo del problema”, consta de varias etapas que, en conjunto, guían al sujeto hacia la construcción de una representación mental del problema basándose principalmente en sus aspectos cualitativos subyacentes (Hegarty et al., 1995; Mayer & Hegarty, 1996; Thevenot, 2010).

La primera etapa se inicia con la comprensión de las premisas del problema, un proceso en el que se transforma cada pieza local de información en proposiciones que representan los elementos descritos en el texto del problema. El resultado de dicha transformación generaría una red semántica latente denominada "base del texto", a partir de la cual se construiría, en una segunda fase, una representación o modelo situacional del problema (Kintsch, 1998). Dicho modelo es el resultado de la integración de los contenidos de la base del texto con los contenidos y experiencias relacionados almacenados en la memoria a largo plazo. De esta manera, las características de los objetos descritos en el texto, así como las palabras claves del problema, quedarían integrados en entidades diferenciadas formando una representación mental netamente más elaborada que la resultante de un proceso de traducción directa. En esta fase, cobra especial relevancia la habilidad para recuperar la información disponible de la memoria y relacionarla con los nuevos datos que se actualizan en la MO. Exige una capacidad para razonar de forma reflexiva e inhibir las respuestas automáticas que puedan generarse durante el proceso, pues su ausencia podría generar una representación inadecuada del problema.

La representación situacional permitirá iniciar una tercera fase que tendría como objetivo elaborar un plan que resuelva el problema planteado, determinando con exactitud las operaciones aritméticas que sean necesarias a partir de las relaciones entre los distintos objetos del problema. Este aspecto resulta especialmente relevante en los AWP, pues uno de los elementos mediante el cual se manipula la dificultad en dichos problemas es a través del uso de adverbios de cantidad "más", "menos" o "igual" que relacionan los objetos del problema y permiten inferir el sentido de las relaciones establecidas. De ese modo, se determinan las operaciones de adición, sustracción, multiplicación o división que sean necesarias en el proceso de cálculo de la respuesta. Esto denota la estrecha relación que mantiene la etapa de generación del plan de resolución y la etapa previa de transformación de la base del texto en un modelo situacional, ya que los adverbios de cantidad implicados en el enunciado del problema no permiten identificar por sí solos el cálculo aritmético necesario para la resolución del mismo, siendo posible una interpretación *a sensu contrario* que requeriría, en consecuencia, la inhibición de la respuesta automática que se genere. En esta fase, también mediaría una habilidad de orden superior que planifica la estrategia de resolución estableciendo la secuencia de operaciones necesarias para lograr una respuesta. Por último, el proceso de resolución finalizaría con una etapa de ejecución del plan previsto en la que es necesaria una supervisión consciente del proceso que minimice los errores de cálculo que puedan producirse.

Si bien pueden identificarse las mismas etapas en la estrategia directa y en la estrategia del modelo del problema, es necesario indicar que la diferencia fundamental entre ambas se encuentra en la segunda fase, es decir, en el modo en que se integra la información a partir de la base del texto. En el caso de la estrategia directa, dicha integración es limitada y centrada en los números y palabras clave del texto del problema, mientras que la otra estrategia orienta el modelo situacional hacia los objetos del problema, permitiendo la integración de las características relevantes de los mismos y el sentido de las relaciones que quedan establecidas por el enunciado del problema.

Tanto el número de relaciones descritas en el problema como la consistencia o no-consistencia de los adverbios que informan sobre el sentido de las mismas constituyen dos variables de enorme potencial experimental, debido a la estrecha relación que guardan con los procesos cognitivos básicos y su alta capacidad para modular la dificultad de la tarea.

1.3. Habilidades generales implicadas en la resolución de problemas

La resolución de problemas aritméticos verbales es una tarea exigente para la cual es necesario poner en funcionamiento un conjunto determinado de habilidades generales y ejecutivas que contribuyen de manera diferencial al procesamiento de la información y la respuesta, así como a la explicación de los resultados logrados. Entre las habilidades generales que influyen en la eficacia de resolución de los problemas aritméticos verbales podemos destacar principalmente dos tipos: las habilidades de tipo cognitivo y las habilidades metacognitivas. Como veremos más adelante, las habilidades cognitivas hacen referencia al conjunto de procesos intelectuales, automáticos y voluntarios, encargados del procesamiento de la información externa. Respecto a las habilidades metacognitivas, constituirían otro conjunto de habilidades relacionadas con el conocimiento del propio procesamiento de la información y el control estratégico y emocional que se ejerce a lo largo del proceso de resolución.

1.3.1. Habilidades cognitivas: la comprensión lectora y el razonamiento

Por su relevancia para la resolución de problemas aritméticos verbales, destacaremos principalmente dos habilidades cognitivas generales estrechamente relacionadas: la comprensión lectora y el razonamiento.

1.3.1.1. La comprensión lectora

En los epígrafes anteriores vimos como el ser humano es capaz de entender aquello que le rodea gracias a la capacidad que desarrolla para representar y clasificar en su mente una realidad externa material. A este acto simultáneo de extracción y construcción de significado (Snow & Sweet, 2003) lo denominamos *comprensión lectora* cuando la realidad sobre la que actúa es un texto escrito.

La comprensión lectora es una habilidad lingüística general constituida por un conjunto de procesos complejos de descodificación y manipulación de información. En tal sentido, puede entenderse como el resultado de un proceso multisistémico y multinivel que comprende una sucesión de etapas que interactúan entre sí y que están dirigidas hacia la formación de una representación del significado que contiene un texto. Estos subprocesos, así como el tipo de representación a la que dan lugar, son los siguientes (García-Madruga, 2006):

1. *El procesamiento léxico*: hace referencia al conjunto de procesos que, a partir de la entrada sensorial, permiten descodificar e identificar los grafemas utilizados para representar las palabras (vía léxica), así como asociarlos a sus fonemas (vía fonológica) y significado léxico correspondientes.

2. *El procesamiento sintáctico*: permite identificar las distintas estructuras sintácticas que relacionan, de acuerdo a las reglas del lenguaje en el que el texto sea representado, las distintas partículas sintagmáticas de una oración.

3. *El procesamiento semántico-pragmático*: tiene que ver con las relaciones de significado que se establece entre las palabras y que conformarían las ideas e intenciones que el autor pretende transmitir.

4. *El procesamiento referencial*: por medio del cual se elabora un modelo mental situacional a partir de la integración del conjunto de ideas que un texto contiene y los conocimientos y creencias que posee el lector. Este proceso aplicado a la resolución de AWP fue extensamente descrito en el epígrafe anterior.

El procesamiento léxico y sintáctico serían niveles superficiales de comprensión textual. Este tipo de habilidades se desarrollarían hasta alcanzar una madurez y automatización a lo largo de la educación primaria (Cuetos et al., 2004). Por contra, el procesamiento semántico-pragmático conformaría el conjunto de proposiciones contenidas en el texto base, su microestructura, según el modelo de construcción e integración de Kintsch y van Dijk (1978; Van Dijk & Kintsch, 1983). Estas proposiciones estarían jerárquicamente organizadas y representadas en distintos niveles que se van construyendo conforme el lector avanza en su lectura y va relacionando los distintos contenidos locales del texto durante el proceso. La integración o síntesis de ese conjunto proposicional basado en los aspectos cualitativos, sus relaciones significativas y los conocimientos del lector permitiría generar un conjunto de ideas generales y específicas que, clasificadas por orden de importancia, darían lugar a la macroestructura textual del texto, identificada también como el "modelo situacional" del texto. Esta macroestructura surge a partir de la aplicación de un conjunto de macrorreglas sobre el contenido proposicional de la base del texto que tendrían como objetivo (Van Dijk & Kintsch, 1983):

1. *Suprimir la información* proposicional que carezca de relevancia para la interpretación del texto escrito.

2. *Seleccionar la información* proposicional relevante, fruto de la cual se dota de contenido semántico global al texto.

3. *Generalizar* o extraer los aspectos que hay en común entre los distintos elementos heterogéneos del texto (lugares, personas, objetos, intenciones…) con el objetivo de clasificarlos en torno a una idea.

4. *Integrar* los contenidos del texto con el conocimiento previo que dispone lector sobre el funcionamiento de las cosas.

La aplicación de estas macrorreglas es un proceso dependiente de los conocimientos del lector. Estos conocimientos pueden hacer referencia a diversos aspectos, como por ejemplo los relacionados con el área temática del texto, el funcionamiento de los procesos mentales que facultan la capacidad de extraer el significado de ese texto (metacognición), o el conocimiento general acerca de la estructura retórica del texto, también denominada *superestructura*. Este esquema general organiza el texto de un modo característico y actuaría como una guía para el lector, facilitándole tanto el proceso de comprensión del texto, como el recuerdo de su contenido.

En relación a la última macrorregla, la integración de contenidos, resulta de enorme importancia por la posibilidad que ofrece al lector de realizar inferencias a partir de lo que lee, una habilidad por medio de la cual se extraerían conocimientos no explícitamente descritos en el texto a partir de las relaciones entre los distintos elementos que se describen y el conocimiento previo que dispone el lector. Así, el modelo situacional generado constituiría, en virtud del tipo y número de inferencias realizadas, una entidad con mayor contenido semántico que la suma de proposiciones contenidas en la base del texto. Estos procesos inferenciales desempeñarían un papel crucial en otro tipo de habilidades cognitivas relacionadas con la resolución de problemas: el razonamiento.

La comprensión lectora es una habilidad cognitiva general que ha sido relacionada en numerosos estudios con la eficacia en la resolución de AWP (Kintsch & Greeno, 1985; Pape, 2004). Los errores más frecuentes que se producen tienen que ver con una inadecuada o pobre comprensión del enunciado del problema (Cummins et al., 1988). En un estudio longitudinal realizado por Lerkkanen y colaboradores (2005), pudieron establecer en estudiantes de educación primaria que las habilidades matemáticas y la comprensión lectora se mantenían altamente relacionadas entre sí durante los dos años siguientes al inicio del estudio. Asimismo, el rendimiento matemático resultaba un factor más predictor de la comprensión, que ésta del rendimiento matemático, lo que justificaba la necesidad de otorgar un papel fundamental al conocimiento matemático desde los primeros años de enseñanza primaria.

De modo parecido, Vilenius-Tuohimaa y colaboradores (2008) identificaron que la resolución de AWP estaba altamente relacionada con las habilidades de tipo superficial de la lectura, es decir, del reconocimiento y descodificación de las palabras, así como con la habilidad para ajustar el tipo de estrategia de lectura y su velocidad en función de las características del enunciado de los problemas. Sin embargo, una vez controladas estas diferencias en las habilidades superficiales de la lectura, encontraron que la eficacia de la resolución de AWP continuaba estando relacionada con las habilidades de comprensión, lo que sugería para estos autores que ambas habilidades estarían también relacionadas a través de unas habilidades de razonamiento general. En la misma línea, otros autores han resaltado la relación entre las habilidades de comprensión de textos y el rendimiento en resolución de AWP (Boonen et al., 2016; Fuchs et al., 2018).

La capacidad para extraer y construir significado a partir del medio escrito es un complejo proceso estrechamente relacionado con la memoria operativa (Daneman & Carpenter, 1980; Gathercole & Alloway, 2006). La memoria operativa, como veremos en próximos epígrafes, puede definirse como el espacio mental de almacenamiento y procesamiento en el que se coordinan y tienen lugar los diversos procesos cognitivos que facultan las distintas habilidades del ser humano, entre ellos la comprensión lectora. En este espacio mental de trabajo, se transforma la información y se almacenan temporalmente los resultados parciales y temporales de los distintos niveles de procesamiento del discurso, a lo largo de una acción coordinada de que da lugar a la representación mental del significado. No es de extrañar, por tanto, que gran parte de las dificultades que pueden encontrarse en relación a la lectura y la comprensión tienen que ver con un funcionamiento inadecuado de los distintos componentes que forman parte de la memoria operativa, así como de las distintas funciones que realiza su sistema ejecutivo central. A este respecto, Carretti y colaboradores (2009) llevaron a cabo un metaanálisis con el objetivo de identificar las diferencias en funcionamiento ejecutivo entre lectores altamente competentes y poco competentes. Los resultados obtenidos indicaron que los lectores menos competentes tenían más dificultades para llevar a cabo tareas verbales complejas que requerían tanto procesamiento y almacenamiento de la información, debido a un funcionamiento menos eficiente de su sistema de control atencional. En cambio, no encontraron diferencias en rendimiento entre ambos perfiles lectores cuando la tarea era de tipo visuoespacial.

Otros estudios, como el realizado por Johan y colaboradores (2020), han mostrado recientemente que la MO es capaz de explicar, junto con la inhibición y la inteligencia fluida, las diferencias relativas a la velocidad lectora de los alumnos preadolescentes. De modo parecido, un estudio de Ober y colaboradores (2019) ha permitido resaltar que la MO ejerce tanto un efecto directo sobre la comprensión de textos, como un efecto indirecto mediado por la habilidad para recordar e inferir fragmentos del texto en población normotípica adolescente. Sin embargo, otras habilidades como el cambio de tarea estarían más relacionadas con los procesos de descodificación de las pseudopalabras, aunque también existen estudios que han podido encontrar asociación entre el cambio de tarea y un mejor rendimiento en comprensión lectora (Follmer, 2018).

Por último, el aprendizaje y uso de estrategias metacognitivas es un factor identificado como vía mediante la cual es posible mejorar el rendimiento en la comprensión lectora

(National Reading Panel, 2000). Estos aspectos serán abordados en mayor profundidad en el apartado correspondiente a las habilidades metacognitivas.

1.3.1.2. El razonamiento

El razonamiento puede definirse como el proceso mediante el cual operamos con una representación o modelo mental con el fin de inferir información no que no aparece explícita (Johnson-Laird, 2006). La adquisición de esta habilidad para manipular representaciones mentales constituye, probablemente, el mayor hito del desarrollo en la adolescencia, ya que junto con la abstracción permite al individuo desligarse de la necesidad de la presencia material y su contenido para poder operar y clasificar los modelos mentales generados en mente (Inhelder & Piaget, 1955).

Las habilidades racionales que son necesarias para resolver los problemas aritméticos verbales pueden ser diferentes de las utilizadas en la resolución de problemas geométricos o de razonamiento probabilístico. No obstante, es posible determinar la existencia de ciertos rasgos comunes que subyacen a todos los problemas matemáticos. En particular, cabe destacar la implicación de un tipo de aptitud cognitiva general basada en el razonamiento abstracto, la inteligencia fluida (Cattell, 1963; Engle et al., 1999; Fung & Swanson, 2017; Horn & Cattell, 1967; Jaeggi et al, 2008), que se identificaría con procesos relacionados con el razonamiento inductivo, deductivo y cuantitativo (Sternberg & Ben-Zeev, 1996).

En relación a estas dos clases de razonamiento, inductiva y deductiva, esta división se identificaría con los dos tipos de inferencia que el individuo realiza en su proceso manipulativo de la representación mental: las inferencias inductivas y las deductivas. El razonamiento inductivo haría referencia a un tipo de proceso mental que permite llegar a conclusiones generales a partir de unas premisas particulares sobre las que se apoya para obtener una conclusión cuya veracidad se entendería en términos probabilísticos. La inducción facultaría a la persona a realizar predicciones o a establecer la probabilidad de ocurrencia de un fenómeno o acontecimiento. Por contra, en el razonamiento deductivo la verdad de las premisas garantizaría que la conclusión que se obtiene fuera necesariamente verdadera, pues es la consecuencia lógica que se deriva de las anteriores. Este tipo de inferencia no necesitaría de ninguna información adicional para validar la veracidad de sus conclusiones, ni aportarían un contenido semántico extra a la persona, sino que su función no es otra que hacer explícita la información que subyace de las premisas (García-Madruga, et al., 2016).

Aunque el pensamiento matemático se ha relacionado con diversas habilidades de razonamiento, existe cierto consenso en que la deducción es uno de sus principales componentes. El razonamiento deductivo constituye un criterio a partir del cual se pueden establecer diferencias entre el comportamiento llevado a cabo por los expertos y por los novatos en una tarea (Sternberg y Ben-Zeev, 1996). La deducción permite categorizar las proposiciones como verdaderas o falsas con independencia de su contenido y, por tanto, actúa como factor de protección frente a la asunción de respuestas poco realistas, o incluso inverosímiles, durante una tarea de resolución de problemas (Wyndhamn & Säljö, 1997).

Los estudios sobre el razonamiento, especialmente aquellos encomendados a analizar los errores o sesgos de procesamiento que se producen debido -entre otros aspectos- al contenido de las premisas implicadas y las creencias de la persona, han permitido alcanzar un considerable acuerdo entre la comunidad científica respecto a la existencia de dos tipos de sistemas de razonamiento diferenciados (ver De Neys & Pennycook, 2019). Estos sistemas de procesamiento ejercerían su actividad a través de procesos muy diferentes; por un lado, los basados en heurísticos de respuesta (Sistema 1), y por otro, los de tipo analítico (Sistema 2; Evans, 1984, 2003; Evans & Over, 1996; Evans & Stanovich, 2013). En relación al Sistema

1, haría referencia a un tipo de razonamiento inmediato, automático y del que sólo su producto final es accesible a la conciencia. Este procesamiento heurístico sería una actividad preatencional que seleccionaría el tipo de información específica del problema que es relevante para su resolución, descartando aquella que se considere superflua, con el fin de simplificar la tarea encomendada y obtener una respuesta rápida de resolución. Así, el modo particular de funcionamiento del Sistema 1 podría definirse mediante el concepto propuesto por Kahneman y Tversky (1972; Tversky & Kahneman, 1974) de heurísticos; esto es, atajos o estrategias que permiten al sujeto acceder rápidamente a una respuesta que puede ser correcta. Sin embargo, el empleo de estos heurísticos produciría sesgos que llevarían al sujeto a la comisión de errores de respuesta cuando son aplicados sobre situaciones aparentemente similares. Por ejemplo, mediante el heurístico de representatividad tenderíamos a evaluar la probabilidad de que una persona sea médico, agricultor o piloto de líneas aéreas, por medio de la semejanza que existe entre la citada persona y el estereotipo de la profesión, ignorando las probabilidades previas de encontrarnos con alguien que verdaderamente desempeñe dicho oficio.

El otro sistema de razonamiento (Sistema 2) sería un tipo de procesamiento lento, cognitivamente costoso y que haría uso extensivo de la memoria operativa (Evans, 2003, 2008; Evans & Over, 1996; Evans & Stanovich, 2013a; Kahneman & Frederick, 2002; Sloman, 1996; Stanovich & West, 2000). Alter y colaboradores (2007) propusieron que los procesos del Sistema 2 se activarían por las experiencias metacognitivas de la persona. Para ello, estos autores manipularon las tareas con el fin de dirigir a sus participantes a experimentar mayor dificultad o falta de fluidez, forzando así que tuvieran que llevar a cabo un proceso analítico propio del Sistema 2. De ese modo, encontraron que los participantes respondían de manera más precisa en una tarea de silogismos cuando estos estaban impresos en un tipo de letra difícil de leer. Este resultado indicaría que la experiencia de la dificultad de los participantes actuaba como un mecanismo de alerta por medio del cual las personas activaban el Sistema 2 de razonamiento.

Fletcher y Carruthers (2012) han propuesto recientemente tres características principales del Sistema 2, que son particularmente relevantes para la metacognición en el razonamiento:

1) Está sujeto al control intencional.

2) Puede guiarse por creencias normativas acerca de los métodos de razonamiento adecuados.

3) Anula las respuestas irreflexivas aceptadas de modo automático por el Sistema 1.

Todas estas características de la evaluación de la propia razón como requisito previo, así como la capacidad de razonamiento en sí misma, son habilidades de desarrollo tardío, emergiendo por lo general durante la preadolescencia (Moshman, 2004; Pillow, 2002), por lo que podría predecirse que antes de esta edad los niños carecerían de un control total en el cambio del Sistema 1 al Sistema 2 de razonamiento.

Del mismo modo que sucede con otro tipo de competencias, las habilidades de razonamiento han sido estrechamente vinculadas a la capacidad y eficacia de los distintos procesos que tienen lugar en la memoria operativa durante el procesamiento de la información (Kyllonen & Christal, 1990; Oberauer et al., 2007; Süß et al., 2002). A este respecto, Evans y Stanovich (2013a, 2013b) sostendrían que una de las características a partir de las cuales puede diferenciarse las actividades del Sistema 2 es por el uso extensivo de los recursos de MO que realizan, en comparación con las actividades del Sistema 1. La capacidad de la MO ha sido destacada como un factor determinante del rendimiento resultante de tareas de razonamiento inductivo (Zamary et al., 2019), razonamiento deductivo (Beatty & Vartanian, 2015), razonamiento condicional (Rojas-Barahona et al., 2021), razonamiento analógico (Simms et al., 2018) o razonamiento probabilístico (Yin et al., 2020).

Como vimos anteriormente, la teoría de procesos duales basa su funcionamiento en la existencia de un proceso de cancelación de respuesta como mecanismo mediante el cual el Sistema 2 toma el control sobre las actividades de pensamiento anulando las respuestas superficiales o heurísticas producidas por el Sistema 1 (Kahneman, 2011). En consecuencia, la eficacia con la que los procesos automáticos de respuesta pasan a ser sustituidos por procesos estratégicos bajo control consciente estaría determinada por las diferencias individuales que pudieran existir respecto a la capacidad de inhibición de respuestas prepotentes y la capacidad para refrescar -actualizar- el contenido de la MO con otras tentativas de respuesta más elaboradas. De ese modo, autores como De Neys y Franssens (2009) demostraron que la inhibición jugaría un papel destacado en la explicación de los sesgos debidos a las creencias en tareas de razonamiento, aquellos que se producen cuando se aceptan como correctas aquellas afirmaciones que coinciden con las creencias de la persona, en lugar de determinar la veracidad de las conclusiones a través del análisis de la validez de las premisas o la probabilidad de ocurrencia de un evento. Otros autores vinculan la inhibición con un mejor rendimiento en razonamiento fluido (Wang et al., 2020) o una cualidad fuertemente asociada al pensamiento crítico (Li et al., 2021). De modo parecido, las habilidades de actualización y cambio atencional también han sido asociadas a algunas habilidades como el pensamiento crítico (Li et al., 2021) o el control cognitivo durante la realización de una tarea (Gathmann et al., 2017).

1.3.2. Habilidades metacognitivas

Las habilidades metacognitivas hacen referencia al conocimiento que dispone la persona sobre cuándo usar, cómo coordinar y cómo supervisar sus habilidades y recursos para obtener una respuesta correcta (Mayer, 1998). Es, en síntesis, el conocimiento acerca del propio conocimiento (Flavell, 1979). Esta habilidad ha sido también definida por otros autores como el acto de escoger y planificar lo que se va a hacer mientras se supervisa lo que se está haciendo, esto es, procesos que juegan un importante papel en las actividades mentales que son dirigidas hacia la aplicación de algoritmos y heurísticos de respuesta (Garofalo & Lester, 1985).

La metacognición, como clase de conocimiento que es, debe entenderse como un concepto multidimensional que responde a las distintas formas en que está representado en la memoria, o metamemoria; esto es, como conocimiento de carácter declarativo y referido al propio sistema cognitivo, llamado *conocimiento metacognitivo*; y como conocimiento de tipo procedimental encargado de la coordinación y regulación de los propios procesos mentales, también llamado *control metacognitivo* (Brown, 1987). Tanto los procesos declarativos, como los procedimentales de la metacognición son elementos que interactúan entre sí y que se interrelacionan con otros aspectos del procesamiento de la información (Gutiérrez-Martínez & Vila, 2013). Así, las habilidades metacognitivas estarían detrás de la diferencia que se establece entre los procesos automáticos y los procesos controlados de la conducta (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977), ya que son las encargadas de regular el comportamiento durante una tarea de resolución de problemas.

En relación al primer componente metacognitivo, el *conocimiento metacognitivo* representaría el conocimiento que posee la persona acerca de su propio conocimiento, lo que implica unas nociones acerca del funcionamiento del sistema cognitivo y sus componentes particulares (memoria, atención, etc.), así como de las actividades mentales que puede o debe desarrollar para llevar a cabo una tarea. Es, por tanto, un conocimiento de tipo declarativo en el que pueden distinguirse tres clases de contenidos:

1. *Conocimientos relativos a la persona*: constituiría el corpus de conocimientos que posee la persona sobre sí misma y sobre los demás en relación a una tarea específica. Este aspecto permitiría asignar tareas a las diferentes personas en función de las diferencias y

capacidades individuales que posean, así como una cierta consciencia de las propias limitaciones para alcanzar la meta deseada.

2. *Conocimientos relativos a la tarea*: hacen referencia al tipo de habilidades y procesos que requiere una tarea para poder realizarse. A partir de este contenido la persona podría estimar la dificultad que entraña una tarea y asignar un mayor número de recursos atencionales o estratégicos para lograr el objetivo.

3. *Conocimientos relativos a las estrategias* que son necesarias o facilitan la realización de una tarea, lo que ayuda a que el proceso de resolución de problemas pueda llevarse a cabo de un modo que resulte cognitivamente más "económico".

A partir del conocimiento metacognitivo, la persona determinaría las demandas cognitivas que requiere una tarea y asignaría convenientemente los recursos necesarios para llevarla a cabo. El conocimiento metacognitivo es considerado por algunos autores como una habilidad *offline*, ya que el efecto que ejerce es previo a la ejecución de la tarea (Desoete et al., 2003).

El segundo componente de la metacognición, el *control metacognitivo*, se refiere al tipo de proceso mental que tiene por objeto regular el conocimiento y la conducta de la persona mientras realiza la tarea. Se trata, por tanto, de un contenido de tipo procedimental que ejercería su actividad regulatoria de la conducta por medio de los siguientes subprocesos:

1. *Planificación*: hace referencia a la conducta previa que se lleva a cabo antes de la realización de la tarea que tiene por objeto organizar y orientar la secuencia de la acción y los recursos que se disponen hacia la consecución del objetivo propuesto.

2. *Supervisión*: correspondería al proceso mediante el cual se evalúan los resultados parciales en cada tramo de la secuencia de la acción con el fin de comprobar si se adecúan a lo esperado inicialmente.

3. *Evaluación*: permitiría valorar si el resultado final obtenido a partir de la estrategia seleccionada se adecúa a los objetivos y metas propuestos.

Tanto la supervisión, como la evaluación se considerarían procedimientos metacognitivos *online*, ya que su presencia se manifestaría durante la ejecución de la tarea (Desoete, 2008). Además de estas tres habilidades de control metacognitivo, algunos autores incluirían entre ellas la *predicción* de resultados, es decir, la capacidad de anticipar correctamente el éxito o el fracaso de las actuaciones en la tarea (Desoete, 2008; Lucangeli & Cornoldi, 1997).

La metacognición es una habilidad que experimentaría un gran desarrollo desde la infancia y la adolescencia (ver Schneider & Löffler, 2016; Weil et al., 2013). Como ya introdujimos parcialmente en epígrafes anteriores, los estudios sobre metacognición han permitido resaltar la importancia de esta variable como un factor determinante del rendimiento mostrado por un individuo en la realización de tareas tales como la comprensión de textos. En relación a esta habilidad, algunos autores, como Fuentes (1998), han podido establecer un perfil diferencial de tipo metacognitivo que permite describir el comportamiento característico de los *lectores competentes* y los *lectores poco competentes* a lo largo de las distintas etapas de la lectura que tienen lugar durante la resolución de AWP:

1. Antes de la lectura:

i. Los lectores competentes activarían su conocimiento previo sobre el tema en cuestión antes de iniciar la tarea, mientras que los lectores poco competentes comenzarían a leer sin llevar a cabo una preparación previa.

ii. Los lectores competentes son capaces de entender la tarea y establecer el propósito, mientras que los lectores poco competentes comienzan a leer sin saber por qué.

iii. Los lectores competentes eligen la estrategia apropiada para la tarea, mientras que los lectores poco competentes leen sin tener una noción sobre cómo aproximarse al material de lectura.

2. *Durante la lectura:*

i. Los lectores competentes son capaces de focalizar su atención, mientras que los lectores poco competentes se distraen con facilidad.

ii. Los lectores competentes son capaces de anticiparse y predecir resultados durante la lectura, mientras que los lectores poco competentes "leen para terminar".

iii. Los lectores competentes son capaces estratégicamente de completar una información cuando se produce una falta de comprensión (por ejemplo, mediante un proceso de inferencia), mientras que los lectores poco competentes desconocen qué hacer.

iv. Para los lectores competentes, la estructura del texto, o superestructura, resulta de utilidad para la comprensión del mismo, mientras que los lectores poco competentes no perciben ninguna estructura en el texto.

v. Los lectores competentes organizan e integran la nueva información que leen, mientras que los lectores poco competentes sólo añaden información adicional sin integrarla con los conocimientos previos que disponen.

vi. Los lectores competentes monitorizan su propio proceso de comprensión sabiendo lo que está ocurriendo en cada momento, mientras que los lectores poco competentes no se dan cuenta de que no están entendiendo lo que leen.

3. *Después de la lectura:*

   i. Los lectores competentes reflexionan sobre la lectura, mientras que los lectores poco competentes terminan su actividad conforme termina la lectura.

   ii. Los lectores competentes consideran que el éxito en la tarea es producto de su esfuerzo, mientras que los lectores poco competentes consideran que el resultado es producto de su suerte en la tarea.

   iii. Los lectores competentes son capaces de sintetizar las ideas principales de un texto y buscar información sobre ello en fuentes externas, habilidades de las que carecen los lectores poco competentes.

De modo similar, autores como Oakhill y cols. (2005) desarrollaron un estudio en relación a las actividades de monitoreo que se efectuarían durante la lectura, en particular, aquellas mediante las cuales una persona evalúa su propia comprensión con el fin de comprobar si es necesario regular la lectura para resolver los problemas que va encontrando y facilitar así el entendimiento. Los resultados encontrados fueron diversos. Por un lado, pudieron confirmar que el monitoreo de la lectura es una actividad que los lectores más competentes realizarían con un grado de precisión mayor que aquellos lectores menos competentes, especialmente cuando se enfrentan a tareas en las que existen inconsistencias insertadas en el texto. Asimismo, la detección múltiple de inconsistencias separadas entre sí que realizan los lectores competentes tiende a ser más efectiva que la de los lectores poco competentes. Sin embargo, apenas habría diferencias en la detección múltiple de inconsistencias cuando éstas están localizadas adyacentemente, lo que indicaría que ambos perfiles lectores pueden llevar a cabo los procesos integración de contenidos de un modo similar bajo unas condiciones específicas. En cualquier caso, ninguno de los resultados obtenidos por estas autoras pudo explicarse por una competencia diferente en relación con las actividades de procesamiento superficial de la lectura, variable que fue convenientemente controlada.

Otros estudios longitudinales, como el realizado por Cain y colaboradores (2004) con alumnos entre 8 y 11 años fueron capaces de demostrar que la comprensión lectora estaría estrecha y moderadamente relacionada en ese tramo de edad con las habilidades metacognitivas de los lectores (entre $r = .40$ y $r = .49$), en particular, con la supervisión que se realiza durante el proceso de lectura. También, estudios más recientes resaltan las habilidades de monitoreo durante la lectura como una de las características determinantes del grado de comprensión alcanzado por los lectores adolescentes (Mañá et al., 2009; Vidal-Abarca et al., 2010).

El despliegue de los distintos procesos metacognitivos que tienen lugar durante la comprensión de textos es una actividad dependiente de los recursos de memoria operativa que tienen disponibles los lectores. Un estudio llevado a cabo por Linderholm y van den Broek (2002) determinó que las personas con mayor competencia lectora son capaces de asignar mayor cantidad de recursos de memoria operativa para la supervisión de la comprensión de textos, en comparación con los lectores menos competentes.

A la vista de tales consideraciones, no es de extrañar que a lo largo de largo de las últimas décadas hayan aparecido numerosos intentos por mejorar la competencia de las distintas habilidades generales o específicas a través de la instrucción de estrategias metacognitivas (Carretti et al., 2014; Carriedo-López, 1992; Mateos, 1991). Sin embargo, también la metacognición se ha significado como una vía a través de la cual poder mejorar las habilidades de razonamiento (Alonso-Tapia et al., 1991; García-Madruga et al., 2022) y la resolución de problemas (Cornoldi et al., 2015).

En relación a las Matemáticas, algunos de los procesos metacognitivos más determinantes a partir de los cuales pueden explicarse las diferencias en rendimiento entre alumnos adolescentes son la predicción de resultados, la planificación, el monitoreo y la evaluación de los resultados en las distintas etapas del problema (Widodo-Yulianto, 2020). Un reciente metaanálisis indicó que las habilidades metacognitivas estarían correlacionadas de forma moderada y positivamente ($r =$

.37) con el rendimiento matemático en la adolescencia (Muncer et al., 2021). Otros estudios, como el realizado por Desoete (2008) indican que las habilidades metacognitivas, junto con una inteligencia, serían capaces de explicar entre el 53% y el 76% de la varianza de resultados en el rendimiento matemático en la escuela primaria.

Los programas de mejora de la competencia matemática y la resolución de problemas a través de la instrucción en habilidades y estrategias metacognitivas han permitido cosechar resultados diversos en función de la habilidad metacognitiva sobre la cual el programa focalizaba principalmente sus esfuerzos (ver de Boer et al., 2018; Desoete et al., 2003). Por ejemplo, recientes estudios han permitido comprobar que la mejora en la percepción de la autoeficacia de los alumnos a través de estos programas tendría un efecto positivo en el rendimiento académico gracias a que se logra incrementar la motivación intrínseca del alumno hacia la tarea (Tian et al., 2018). Este aspecto nos llevaría a considerar el modo diverso en que las habilidades metacognitivas ejercen un efecto positivo sobre el rendimiento académico en Matemáticas (William & Maat, 2020).

Por ejemplo, Lucangeli y Cornoldi (1997) demostraron con alumnos de primaria que aquellos que obtenían mejor rendimiento en Matemáticas disponían de un conocimiento más claro de las reglas útiles para evaluar la corrección de una respuesta y de los pasos necesarios para lograr sus objetivos. Esta relación resultaba más evidente en aquellas tareas matemáticas menos dependientes de procedimientos automatizados, o heurísticos, como las de resolución de problemas. En el cálculo de operaciones aritméticas, por el contrario, el efecto del control metacognitivo resultaba menos determinante, ya que la eficacia dependía en mayor medida de si el alumno disponía o no del algoritmo de computación por medio del cual se realizan las operaciones. En la misma línea, Legg y Locker (2009) comprobaron que los alumnos que contaban con más habilidades metacognitivas lograban moderar o regular el impacto emocional de la tarea y, por tanto, los niveles de ansiedad ante las Matemáticas, lo que ocasionaba que la confianza acerca de la respuesta que daban en una tarea de resolución de problemas matemáticos fuera mayor.

En relación al conocimiento metacognitivo, Swanson (1990) demostró que los niños de primaria con grandes habilidades metacognitivas alcanzaban un rendimiento mayor en tareas de resolución de problemas que aquellos que tenían menor nivel. Además, aquellos con mayor nivel eran más propensos a recurrir a estrategias hipotético-deductivas de tipo condicional, así como a efectuar comprobaciones sobre la idoneidad de una hipótesis, en comparación con aquellos alumnos con una puntuación más baja en metacognición. Otros autores como Tian y colaboradores (2018) han confirmado recientemente la relación entre el conocimiento metacognitivo y un mejor rendimiento matemático en adolescentes. Esta relación estaría, además, mediada por otro tipo de variables, como la percepción de autoeficacia de la persona y su motivación intrínseca, cuyos efectos también son determinantes del rendimiento en Matemáticas.

Asimismo, existirían diferencias significativas en el modo en que hombres y mujeres hacen uso de su conocimiento metacognitivo en la resolución de problemas matemáticos difíciles, encontrándose un uso más extensivo de dicho conocimiento en hombres que en mujeres, lo que contribuiría a explicar parte de las diferencias de rendimiento en las tareas de resolución de problemas.

1.4. La memoria operativa y el funcionamiento ejecutivo

Históricamente, el término memoria ha sido definido de múltiples maneras atendiendo básicamente a dos criterios fundamentales: persistencia o duración de la huella en la memoria (corto plazo vs largo plazo), y capacidad de almacenamiento (limitado vs ilimitado). Denominamos memoria operativa (MO) al sistema responsable del almacenamiento temporal y la manipulación activa de la información. Este elemento se ha convertido en uno de los conceptos claves en la explicación de los procesos cognitivos subyacentes a cualquier actividad intelectual. Durante el desarrollo normal se produce un incremento de su capacidad y eficiencia que explica, en parte, las diferencias en la competencia académica del individuo (Daneman & Carpenter, 1980; Gathercole & Alloway, 2006; Swanson, 1994; Swanson et al., 1990).

A lo largo de las últimas décadas, se han propuesto distintos modelos explicativos de funcionamiento de la memoria que constituyen los antecedentes históricos del actual concepto de memoria operativa (véanse, por ejemplo, Atkinson & Shiffrin, 1968; Craik & Lockhart, 1972). Sin embargo, uno de los modelos que mayor repercusión ha generado ha sido el modelo de Baddeley y Hitch (1974), junto con sus sucesivas revisiones (Baddeley, 1986, 1992, 2000). Este modelo, a diferencia de los anteriormente propuestos, concebiría la memoria no sólo como un sistema de almacenamiento de información, sino también como un espacio mental de procesamiento en el cual la información proveniente de los distintos formatos se codifica, integra y elabora relacionándola con las experiencias y conocimientos previos que se disponen.

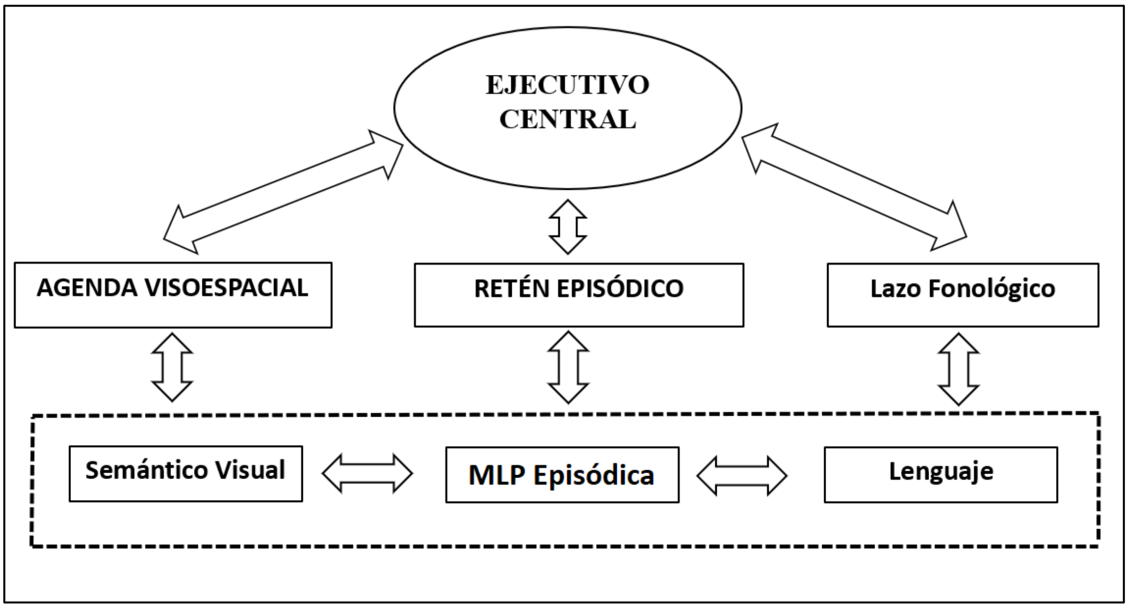
1.4.1. Modelo de memoria operativa de Baddeley

La MO, según el actual modelo de Baddeley, constaría de 3 subsistemas básicos esclavos o subsidiarios de un componente directivo denominado sistema ejecutivo central. (ver figura 1). Estos componentes son:

a) *Agenda visoespacial*: encargada de la integración, manipulación y almacenamiento de los contenidos visuales y espaciales. Haría posibles actividades tales como la interpretación de mapas o dibujos técnicos, así como el encargado de la rotación mental de figuras para la resolución de puzles o acoplamiento de piezas, etc. Esta habilidad de rotación ha sido frecuentemente relacionada con el cálculo aritmético y el cálculo mental (Cheng & Mix, 2014; Thompson et al., 2013), siendo más predictiva del rendimiento en Matemáticas cuanto menor es la edad de los participantes (Holmes et al., 2008). Un reciente metaanálisis realizado por Allen y colaboradores (2019) ha permitido confirmar la relación existente entre la agenda visoespacial y el rendimiento en Matemáticas.

b) *Lazo fonológico:* encargado de la integración, manipulación y almacenamiento de los contenidos fonológicos y auditivos, y su mantenimiento mediante repaso subvocal. El lazo fonológico desempeñaría un papel primordial en la adquisición del lenguaje y la lectura (Baddeley et al., 1998), así como una mayor relevancia en las tareas aritméticas en las que el cálculo implica el almacenamiento de información temporal (Fürst & Hitch, 2000). No obstante, su contribución a la explicación del rendimiento matemático no es concluyente y, en todo caso, limitada (Passolunghi et al., 2007; Passolunghi & Costa, 2019).

Figura 1. Modelo multicomponente de la Memoria operativa (Baddeley, 2000)



c) *Retén episódico*: actúa como interfaz entre los distintos componentes de la MO. Estaría encargado de unificar los contenidos de los distintos formatos perceptivos en un espacio mental para su elaboración y almacenamiento activo. Además, permitiría la recuperación de los contenidos de la memoria a largo plazo (MLP) para que fueran combinados y elaborados con la información entrante.

d) *Ejecutivo central (EC):* que coordinaría toda la actividad de los distintos subsistemas y el control atencional del sujeto para la transferencia del producto resultante a otras áreas cognitivas. Es el principal componente de la MO y responde, en esencia, al "sistema de atención supervisora" descrito por Norman y Shallice (1986). La actividad directiva y de control atencional del EC, responsable de la regulación de la conducta del sujeto, estaría asociada al desarrollo del córtex prefrontal (Diamond, 2013; Yuan & Raz, 2014), aunque también a otras regiones corticales y subcorticales (Alvarez & Emory, 2006; Clark et al., 2021; Heyder et al., 2004). Su actividad puede ser descompuesta en una serie de *funciones,* llamadas *ejecutivas* (FE), que describirían unos procesos relativamente independientes, aunque relacionados entre sí, por medio de los cuales el individuo: 1) ajustaría sus expectativas y se anticiparía a las consecuencias; 2) planificaría la secuencia de su acción; 3) llevaría a cabo sus propósitos de manera flexible y; 4) se autoevaluaría estableciendo, también, un control emocional frente a la tarea (Luria, 1966).

Las distintas tareas que desarrolla el EC fueron objeto de una profusa actividad investigadora, con resultados dispares, en pacientes con lesiones prefrontales y no prefrontales para determinar la unicidad o diversidad de su funcionamiento (ver una revisión en Garcia-Barrera, 2019; Karr et al., 2018).

1.4.2. Modelos de funcionamiento ejecutivo

Desde los primeros años 50 del siglo XX hasta la actualidad, diferentes modelos de funcionamiento ejecutivo han sido propuestos como explicación a la facultad que posee el individuo para manejar y regular su comportamiento dirigido hacia la consecución de objetivos. Muchos de estos modelos surgen como consecuencia del incremento de las evidencias obtenidas a partir del estudio de fenómenos con pacientes que padecen algún tipo de lesión o deterioro cognitivo, así como por el desarrollo y empleo de nuevas técnicas y procedimientos en el campo de la Psicología y la Neurociencia. Algunos de los modelos propuestos hasta la fecha son el modelo de procesos automáticos y controlados (Broadbent, 1958), el modelo de control cognitivo (Posner & Snyder, 1975), el modelo de procesamiento atencional automático y controlado (Shiffrin & Schneider, 1977), el modelo del sistema atencional de supervisión (Shallice, 2002), o el modelo en cascadas (Banich, 2009), entre otros. Una revisión más exhaustiva puede encontrarse en Goldstein y cols. (2014), o Chai y cols. (2018).

A continuación expondremos dos de los modelos que han sido tomados como referencia en la elaboración de este trabajo de tesis: el modelo de funcionamiento ejecutivo de Miyake y cols. (2000), y el modelo de García-Madruga y cols. (2016).

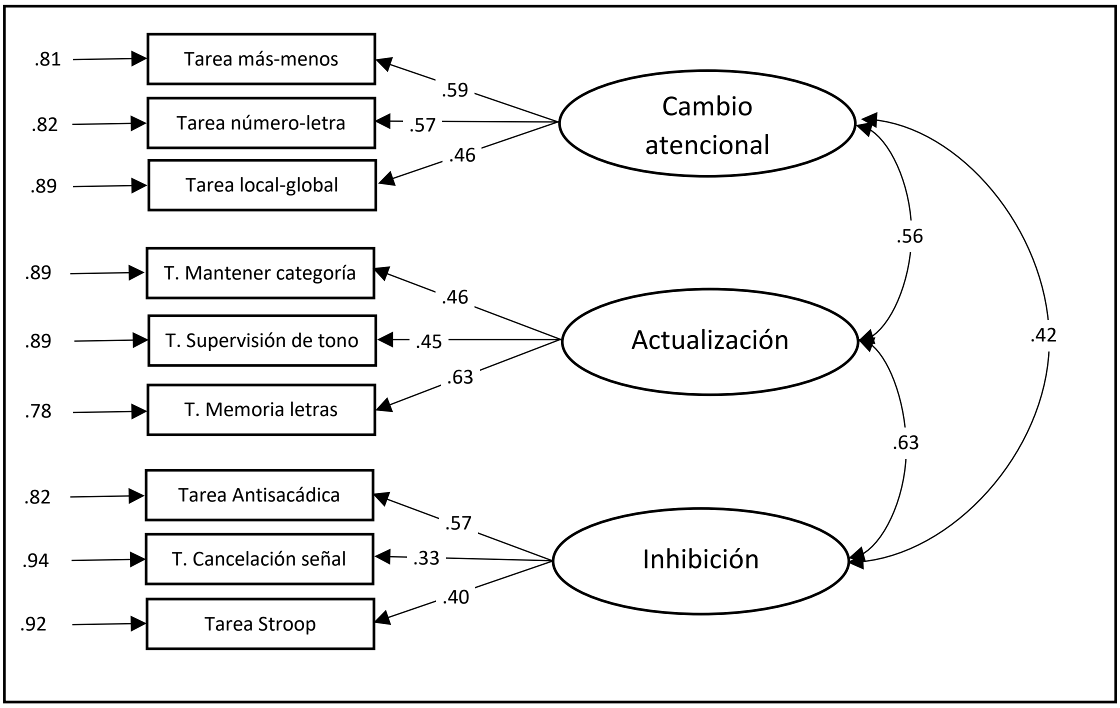
1.4.2.1. Miyake y cols. (2000)

Uno de los trabajos más relevantes e influyentes en el estudio de las FE fue el desarrollado por Miyake y cols. (2000). Estos autores fueron capaces de demostrar matemáticamente la existencia de tres entidades latentes independientes, representadas dentro de una elipse (ver figura 2), que corresponderían a las funciones ejecutivas que el EC llevaría a cabo para el control y procesamiento de la información. Como veremos más adelante, el modelo tripartito de funciones ejecutivas propuesto no negaría la existencia de subprocesos diferenciados que, en conjunto, constituirían cada una de las funciones del ejecutivo central. Tal es el caso, por ejemplo, con la capacidad del individuo para inhibir sus actividades de pensamiento. Estudios posteriores (Friedman & Miyake, 2004) inspirados en otras aportaciones teóricas relevantes (Dempster & Corkill, 1999; Harnishfeger, 1995; Nigg, 2000) pudieron determinar que la inhibición debía ser entendida como un conjunto diverso de procesos inhibitorios que, aunque compartieran parte de la varianza de sus resultados, implicaban factores y, por tanto, subrutinas ejecutivas distintas del mismo factor general latente.

Las funciones ejecutivas que identifica el modelo de Miyake y colaboradores son las siguientes (ver figura 2):

1. *Cambio atencional*: tendría que ver con la alternancia del foco de la atención durante las distintas operaciones o estrategias que pueda requerir la secuencia de resolución de una tarea (Monsell, 1996), así como la supervisión flexible de todo el proceso para determinar el mejor modo de llevarlas a cabo (ver Diamond, 2013). Jugaría un papel importante a la hora de entender los fallos de control cognitivo en pacientes con daños cerebrales en las tareas de laboratorio que requieren un cambio de tarea (Monsell, 1996).

Figura 2: Modelo de estimaciones de tres factores (Miyake et al., 2000)



La alternancia entre tareas distintas y/o novedosas provocaría un incremento de la demanda de los recursos de la MO que podrían aumentar la probabilidad de cometer errores en la respuesta (Rogers & Monsell, 1995). Esta habilidad ha sido relacionada con el rendimiento en la resolución de problemas matemáticos (Schwaighofer et al., 2016).

2. *Actualización*: correspondería a la habilidad para renovar y monitorizar las representaciones mentales de la MO, o dicho en palabras de Morris y Jonnes (1990) "el acto de modificar el estatus actual de la representación de un esquema en la memoria para acomodarse a una nueva entrada de información" (p.112). Esta función requeriría, por tanto, la realización de dos subtareas: a) el control y la codificación de la información entrante para determinar su relevancia para la tarea en curso y; b) la revisión adecuada de los elementos mantenidos en la memoria de trabajo mediante la sustitución de la información antigua no relevante por información más nueva y pertinente (Miyake et al., 2000). Su acción, por tanto, no sería meramente pasiva (almacenamiento), sino que también ejercería un papel activo en la manipulación de los contenidos de la MO.

A este respecto, Ecker y colaboradores (2010) llevaron a cabo un estudio para determinar los componentes o subfunciones que integrarían el constructo de actualización. Estos autores encontraron que el rendimiento en actualización estaría explicado por la presencia de tres factores diferenciados que constituirían los subprocesos de esta función:

i. <u>La recuperación</u> de contenidos declarativos o procedimentales de la memoria a largo plazo que fueran pertinentes para la realización de la tarea.
ii. <u>La transformación</u> de la información mantenida en la MO, en base, también, a los contenidos anteriormente recuperados, lo que permite la elaboración de la representación mental.
iii. <u>La sustitución</u> de los contenidos en la MO que ya no resultan de utilidad por otros que sí lo son. Este último subproceso sería el único específico de la función de actualización. Las subrutinas de recuperación y transformación de la información

también serían funciones compartidas y relacionadas con la capacidad temporal de almacenamiento de la MO.

Algunos autores han podido identificar cambios evolutivos en los procesos de actualización de la MO. Linares y colaboradores (2016) identificaron una reducción del tiempo que invertían los individuos en la recuperación, transformación y sustitución de contenidos en la MO con la edad, logrando alcanzar, además, una mayor precisión en la ejecución de dicho proceso. Estos cambios asociados a la edad se producirían, particularmente, en los procesos de recuperación de la información. Del mismo modo, Carriedo y colaboradores (2016) encontraron que los subprocesos de recuperación y transformación propios de la actualización se desarrollarían hasta la adolescencia, siendo la sustitución un subproceso que aún experimentaría un desarrollo creciente hasta la edad adulta.

La actualización es una habilidad frecuentemente relacionada con el rendimiento académico. Por ejemplo, Gathercole y colaboradores (2004) comprobaron que el rendimiento en Lengua y Matemáticas en niños de 7 años estaba determinado por el rendimiento obtenido en pruebas de MO, particularmente en las tareas de amplitud. Sin embargo, el factor capacidad de la MO dejaba de ser explicativo del rendimiento en Lengua a partir de los 14 años, aunque no para las Matemáticas y otras asignaturas del área de ciencias.

3. *Inhibición*: haría referencia a la función mediante la cual una tendencia dominante, automática o predominante de respuesta es inhibida, de forma deliberada, cuando la situación lo requiere (Miyake et al., 2000). La inhibición ha sido asociada a habilidades cognitivas de muy diverso tipo, así como a otros factores sociales y comportamentales (Diamond, 2013). Esta habilidad básica dotaría al individuo de una capacidad para autorregularse y actuar de forma flexible frente a las distintas condiciones que requiera

una tarea (Gutiérrez-Martínez et al., 2017). Del mismo modo que sucede con la actualización, no se puede considerar un único factor que se identifique con la inhibición (Friedman et al., 2008; Friedman & Miyake, 2004), sino que existiría un conjunto variado de *procesos de inhibición* que proporcionarían la capacidad para regular tanto el pensamiento como la conducta. Aunque los procesos de inhibición difieran en algunas de sus características, estarían relacionados y controlados por las mismas estructuras neurológicas frontales (Friedman & Miyake, 2004; Miyake et al., 2000) y prefrontales (Munakata et al., 2011; Ridderinkhof et al., 2004).

Los subprocesos que integrarían el constructo de inhibición serían los siguientes:

i. <u>Inhibición de las respuestas prepotentes</u>: facultaría a la persona de una capacidad de suprimir deliberadamente las respuestas dominantes, automáticas o preponderantes de una tarea.

ii. <u>La resistencia a la interferencia que puedan generar los distractores</u>: capacidad de resistir o resolver las interferencias de la información del entorno externo que es irrelevante para la tarea en cuestión. Estaría asociada a la atención focalizada o al refuerzo selectivo de los estímulos objetivo. En este subproceso, el estímulo objetivo y la interferencia se presentarían simultáneamente.

iii. <u>La resistencia a la interferencia proactiva</u>: dotaría a la persona de la capacidad para resistir las intrusiones en la memoria de información que antes era relevante para la tarea, y ya no lo son. En este subproceso, el estímulo que actúa como interferencia se presenta antes que el estímulo objetivo.

La inhibición es una FE que sufre cambios y transformaciones con la edad. Los niños serían capaces de mostrar algunas conductas inhibitorias desde los 9 meses de edad (Amso

& Johnson, 2008; Frank et al., 2009). Un estudio reciente elaborado por Constantinidis y Luna (2019) acerca de los procesos inhibitorios revelaría que durante la adolescencia se produce una maduración de las estructuras neurológicas subyacentes a esta facultad que estaría, además, vinculada a una maduración funcional de dichos procesos. Las respuestas conductuales inhibitorias de los adolescentes serían similares a la de los adultos, con la salvedad de que dicha respuesta no sería siempre consistente. Un aspecto muy relevante de este estudio es la afirmación de que la maduración del control inhibitorio se lograría a través de las mejoras en las conductas que se dirigen a la preparación de la respuesta, el procesamiento de errores y las respuestas planificadas.

St Clair-Thompson & Gathercole (2006) demostraron que las habilidades de inhibición estarían asociadas significativamente con el logro académico general, una vez que son controladas las capacidades de actualización de los participantes. Esta asociación fue, incluso, más explicativa que la obtenida por la capacidad de adquisición de conocimientos en un dominio específico. Por otro lado, Lubin et al., (2013, 2016) señala la importancia que tiene la inhibición en la cancelación de algunas estrategias *sobreaprendidas* por los estudiantes cuando se enfrentan a la resolución de problemas aritméticos verbales (Hegarty et al., 1995; Mayer et al., 1996; Shum & Chan, 2020). La inhibición actuaría anulando la secuencia de respuesta heurística, que sería más accesible en la memoria.

Adicionalmente, diversos estudios realizados con población preadolescente (10-11 años) diagnosticada con trastorno por déficit de atención e hiperactividad (TDAH) muestran una dificultad significativamente mayor de estos alumnos, en comparación con un grupo normativo, a la hora de seleccionar la información relevante con la que calcular las operaciones necesarias para la resolución de problemas matemáticos, así como un número mayor de intrusiones cuando el enunciado del problema contiene información no relevante para la resolución (Marzocchi et al., 2002; Re et al., 2016).

1.4.2.2. Modelo de funcionamiento ejecutivo de García-Madruga y cols. (2016)

Existen otras propuestas complementarias de organización funcional del sistema ejecutivo. Una de las más recientes es la realizada por García-Madruga y colaboradores (2016). Estos autores discriminarían entre tres tipos de funciones ejecutivas:

1. *Funciones ejecutivas nucleares (on-line) de la MO*: estas funciones participarían durante la ejecución y monitoreo de la tarea (on-line) cuando su uso sea requerido. Estarían integradas por las tres funciones clásicas descritas por Miyake y colaboradores anteriormente expuestas (cambio atencional, activación y actualización de representaciones mentales, e inhibición de respuestas o información no relevante) y una última función complementaria: la focalización y mantenimiento de la atención durante la tarea. La integración de la función de focalización dentro de este modelo nuclear de FE estaría fundamentada en las aportaciones de Pascual-Leone (1987, 2000b, 2000a) en el campo de la atención. Para este autor, la atención es concebida como un sistema modular compuesto por una capacidad mental-atencional limitada (espacio metal M) que permite mantener en la mente diferentes elementos o esquemas de información que son relevantes para la resolución de la tarea intelectual; y un mecanismo que interrumpe la atención mediante el cual se inhiben activamente los esquemas que no son relevantes para la tarea. Estos dos componentes atencionales básicos, la capacidad de activación mental y la capacidad de inhibición mental, incluirían además dos componentes ejecutivos: el desplazamiento y la actualización. Estudios como el realizado por Agostino y colaboradores (2010) mostraron la relevancia y contribución de estos cuatro componentes en la habilidad de resolución de problemas aritméticos que implicaban multiplicación.

2. *Funciones ejecutivas de orden superior (off-line) de la MO*: posibilitarían que las habilidades intelectuales más complejas, como el razonamiento y la resolución de problemas, pudieran llevarse a cabo en acción conjunta con las funciones ejecutivas

nucleares. Se conciben como funciones "off-line" porque, a diferencia de las anteriores, no estarían presentes durante la ejecución y monitoreo de la tarea, sino que su acción comienza, bien antes, o bien después de que curse la tarea. Tampoco serían procesos necesarios para la ejecución de la tarea, aunque sí un cierto grado de planificación de la secuencia de respuesta debe estar siempre presente para la realización de la misma (García-Madruga et al., 2022). Ambas serían funciones de desarrollo tardío. Se diferenciarían, por tanto, dos tipos de funciones ejecutivas:

i. Planificación del comportamiento futuro: definida por Luria (1966) como la capacidad de organizar la conducta para alcanzar un objetivo específico que puede ser desglosado en una serie de submetas o pasos intermedios. Su acción se manifiesta antes de que se inicie la conducta del sujeto. Esta función estaría implicada en habilidades intelectuales complejas, como el razonamiento (Andrews et al., 2014; Handley et al., 2002) o la resolución de problemas (Eichmann et al., 2019; Unterrainer & Owen, 2006); y también relacionada con otras funciones ejecutivas, como los procesos de inhibición (D'Antuono et al., 2017; Humes et al., 1997) o la memoria operativa (Gilhooly et al., 2002; Injoque-Ricle & Burin, 2011). Sikora y colaboradores (2002) encontraron un rendimiento significativamente peor en la tarea de la Torre de Londres (TOL) en un grupo con dificultades aritméticas, que en un grupo con dificultades de lectura o un grupo sin dificultades.

ii. Revisión de la ejecución de la tarea: correspondería con la acción de verificación que se realiza sobre el producto resultante cuando la tarea ha finalizado. En palabras de Pólya (1945), se identificaría con un "mirar atrás". Permitiría la detección de inconsistencias realizadas en la ejecución de la tarea mediante la búsqueda de contraejemplos o la supervisión exhaustiva de las alternativas de respuestas, habilidades indispensables en tareas como el razonamiento y la resolución de problemas.

3. *Procesos emocionales*: subyacen a cualquier tipo de tarea, independientemente de su dificultad o novedad. Estos procesos incluirían el control emocional de la conducta, o lo que es lo mismo, la habilidad para domeñar la respuesta emocional que suscite la tarea e imponer el pensamiento racional. Estos procesos están ligados al desarrollo metacognitivo del individuo, una facultad ampliamente desarrollada durante la adolescencia (Schneider & Löffler, 2016; Veenman et al., 2006) y que ha demostrado su influencia en las tareas de resolución de problemas (Abdullah et al., 2017; Gómez-Chacón et al., 2014; Vula et al., 2017). En este sentido, algunos autores cuantifican que las habilidades metacognitivas serían responsables de un 37% de la varianza de resultados en la tarea de resolución de problemas matemáticos (Desoete et al., 2001).

1.4.3. Diferencias evolutivas en la memoria operativa

Las explicaciones en torno al desarrollo de la memoria operativa (MO) pueden clasificarse en base a dos criterios que harían referencia a la naturaleza del cambio que se produce en dicho sistema cognitivo. Así, un primer grupo estaría conformado por aquellos autores que consideran que los cambios que se producen en la MO son consecuencia del desarrollo de su estructura, que conllevaría un incremento de su capacidad; mientras que el otro grupo estaría formado por aquellos autores que consideran los cambios cualitativos en la MO como la consecuencia de un incremento en la eficiencia del funcionamiento de sus componentes.

Uno de los autores más relevantes de la visión estructuralista del desarrollo de la MO es la teoría de los operadores constructivos de Pascual-Leone (1970, 1987; Pascual-Leone & Johnson, 2011). Esta teoría permitiría predecir el desarrollo de los subestadios propuestos por Piaget, ya que considera que estos estadios son alcanzados a partir del incremento de la capacidad central de procesamiento (espacio mental M) que se correspondería con el constructo de MO de otros modelos. Este incremento de espacio mental, determinado por el desarrollo neurológico del individuo, permitiría que el individuo fuera capaz de procesar e integrar simultáneamente un mayor número de esquemas, dando lugar a los cambios cualitativos en el desarrollo que describe Piaget gracias a la acción conjunta del mecanismo atencional activador. Así, entre los 3 y los 15 años, se produciría un incremento gradual de la "energía mental" del sujeto, llegando a procesar un máximo de 7 esquemas simultáneos, lo que conformaría la capacidad de almacenamiento temporal de la MO.

Por otro lado, existiría una visión funcionalista del desarrollo de la MO (Case, 1974; Case et al., 1982) que consideraría que la capacidad total de la MO no experimentaría cambios estructurales significativos desde la infancia, sino que los cambios en la MO vendrían dados por la mejora de la eficiencia en los procesos y estrategias que subyacen a la realización de las tareas hasta el grado de automatización de algunos de ellos. Esto permitiría liberar recursos cognitivos que podrían emplearse en la elaboración de una representación metal más compleja. La mejora de la eficiencia en las tareas de procesamiento se produciría como consecuencia del número de experiencias del sujeto con la tarea, lo que incrementaría su especialización y mejoraría su rendimiento.

La mejora de las técnicas de neuroimagen, unida a la enorme variedad de estudios sobre las mejoras evolutivas en las funciones ejecutivas, invita a pensar que ambas posturas podrían ser complementarias (Camos & Barrouillet, 2018; Chai et al., 2018; Spencer, 2020).

Aunque a lo largo de este capítulo se han hecho brevemente referencia a algunas de las relaciones que se establecen entre los principales componentes y procesos cognitivos revisados, y la resolución de problemas aritméticos verbales, estas relaciones serán abordadas con mayor detalle a lo largo de la exposición teórica introductoria de cada uno de los experimentos.

2. OBJETIVOS E HIPÓTESIS

2.1. Descripción

El objetivo principal de esta tesis fue profundizar en el estudio de los procesos cognitivos y metacognitivos implicados en la resolución de problemas aritméticos verbales (AWP) en la adolescencia.

Como hemos visto en epígrafes anteriores, la adolescencia es un periodo del desarrollo en el que se producen cambios intelectuales significativos como consecuencia de la maduración de las estructuras cerebrales y el incremento de la eficiencia con que se realizan las actividades de pensamiento y otros procesos subyacentes a las tareas. Estos aspectos dotan al individuo de una mayor capacidad para aprender y reflexionar, lo que proporciona un conocimiento más extenso acerca del funcionamiento de las cosas y de su propio sistema cognitivo. Durante la adolescencia, un periodo que se inicia en los últimos años de la Enseñanza Primaria (EP) y continúa en la Enseñanza Secundaria Obligatoria (ESO), comienzan a construirse y definir las vocaciones profesionales de las personas a partir de las experiencias que se van adquiriendo en los diferentes entornos y condiciones a los que se expone. El resultado que obtiene la persona a partir de su proceso interactivo con dicha experiencia podría estimular un interés creciente sobre un área de conocimientos determinada, especialmente si además siente que es capaz de rendir competentemente en ese terreno. Es por ello que, para el despertar de las vocaciones científicas, resulte necesario determinar cuáles son las habilidades de tipo cognitivo y metacognitivo que, desarrolladas normalmente, permiten que una persona pueda emprender de manera exitosa actividades tales como la resolución de problemas aritméticos verbales (AWP), una competencia curricular básica que resulta imprescindible en cualquier ámbito de la vida.

La resolución de problemas aritméticos verbales es un extenso y prolífico campo de investigación en el que han contribuido, entre otros, autores provenientes de las distintas ramas de la Psicología, la Educación y las Ciencias, en general. El análisis de las aportaciones al campo de estudio nos permite comprobar cómo una gran cantidad de esos trabajos fueron llevados a cabo teniendo como objetivo la determinación del tipo de estrategias que despliegan los participantes durante el proceso de resolución de los problemas, así como el análisis de sus errores característicos. Este aspecto ha permitido identificar y fundamentar algunas de las claves educativas e instruccionales sobre las que actuar para facilitar la adquisición de los conocimientos matemáticos y la resolución de problemas. Sin embargo, durante los últimos 20 años, y gracias a la consolidación de los distintos modelos de funcionamiento ejecutivo, se ha producido también un incremento paulatino -aunque todavía limitado- de publicaciones relacionadas con el estudio de algunas de las variables cognitivas y ejecutivas implicadas en el proceso de resolución. Estos estudios han contribuido notablemente al incremento del conocimiento sobre la materia destacando, por ejemplo, la importancia de ciertas habilidades de tipo general, como la comprensión de textos, que resultan esenciales para que el proceso de resolución de un problema se desarrolle de manera eficaz. La naturaleza narrativa de los AWP exige de la persona a cargo de la resolución que genere una representación o modelo situacional del problema que sea adecuado a las características y relaciones que se establecen en el enunciado, pues de modo contrario una representación superficial conduciría a la toma de decisiones y soluciones equivocadas. Así, comprobamos cómo las aportaciones de los teóricos provenientes del campo de la comprensión han supuesto un incuestionable enriquecimiento en el estudio de esta materia, logrando ofrecer explicaciones del proceso de resolución de AWP con un nivel de detalle mayor y, también, complementario con los modelos teóricos previamente postulados. Sin embargo, este mismo fenómeno de transferencia no se ha visto replicado con otro tipo de habilidades que, presumiblemente, también deberían estar implicadas en la eficacia de resolución de AWP, como son las habilidades de razonamiento. Tal es así que apenas es posible encontrar estudios en los que se analice la contribución específica de esta habilidad en la resolución de AWP, ni siquiera en relación con el procesamiento dual, pese a que podrían explicar parte de las diferencias en la eficacia con que se resuelven los problemas de tipo consistente e inconsistente y algunos sesgos característicos que se encuentran. Este será uno de los objetivos a considerar en el estudio 1 de este trabajo de tesis.

En la misma línea, también es frecuente encontrar estudios que evalúan la actuación de los participantes en relación con alguna función ejecutiva de la MO en un grupo poblacional específico, siendo más habituales aquellos en los que el rendimiento de los participantes se justifica en base a sus capacidades de inhibición o de actualización de contenidos en la MO. Sin embargo, tampoco nos ha sido posible encontrar apenas estudios en los que, por ejemplo, se analizara el papel que desempeñan los procesos de planificación en la resolución de AWP, unas habilidades necesarias para el establecimiento de una secuencia apropiada de resolución del problema. Además, pese a que de modo individual estas medidas de funcionamiento ejecutivo sí han mostrado su significativa contribución a la explicación de la precisión de respuesta lograda por los participantes en la resolución de AWP, el número de trabajos en los que se analiza de manera combinada las relaciones existentes entre las habilidades generales, los procesos ejecutivos y la eficacia de resolución es también reducido. Consideramos, por tanto, que establecer como objetivo un análisis combinado de todas estas habilidades cognitivas puede ayudar a proporcionar una representación más completa de las condiciones y recursos presentes que interactúan durante la tarea, lo que dotaría a este trabajo de mayor nivel de detalle. Estos aspectos serán considerados en los estudios 2 y 3 de este trabajo de tesis doctoral.

Otra cuestión importante que motiva la realización de este trabajo tiene que ver con el tipo de población sobre la que, por término general, se han efectuado gran parte de trabajos de investigación en el área de la resolución de AWP. A este respecto, podemos identificar dos grandes criterios en que podrían clasificarse las distintas aportaciones experimentales en este campo y que aluden a la tipicidad de la muestra y al grupo de edad analizado. En relación al criterio de *tipicidad de la muestra*, podríamos dividir los estudios en dos grupos poblacionales mayoritarios: población normotípica y población atípica. Del mismo modo que sucede con otras áreas de la Psicología o la Medicina, gran parte de lo que conocemos sobre los procesos y funciones que se activan durante la resolución de problemas proviene del estudio de los factores que explican un comportamiento o actuación ineficaz en poblaciones que presentan algún tipo de dificultad de aprendizaje, generalmente relacionada con la adquisición de conocimientos Matemáticos, u otro tipo de trastornos de tipo atencional y/o comportamental (TDA o TDAH). En cambio, la descripción de la progresión y, por tanto, del perfil que describe un estudiante normotípico en su proceso de adquisición de la competencia de resolución de AWP continúa adoleciendo de un limitado número de estudios que, además, suelen centrarse en la adquisición y despliegue de las distintas estrategias que pueden emplearse en el proceso de resolución de los problemas, más que en las habilidades cognitivas generales y ejecutivas que las posibilitan. Una descripción precisa de este patrón de desarrollo podría resultar de enorme interés educativo por las implicaciones instruccionales que pudieren derivarse, especialmente en relación a la baja eficiencia académica.

En relación al segundo criterio de clasificación de los estudios en esta área, tanto en los estudios sobre población normotípica, como en aquellos realizados con población atípica, encontramos dos grandes grupos de edad claramente mayoritarios que protagonizan la mayor parte de las investigaciones. Estos grupos estarían principalmente formados por niños de hasta 8 o 9 años, o por adultos universitarios, siendo el número de estudios sobre población adolescente ciertamente escaso y, en todo caso, limitado a las condiciones anteriormente expuestas. Muy pocos estudios han evaluado la eficacia de los participantes normotípicos adolescentes en la resolución de AWP, en dos niveles escolares diferentes, en un periodo diverso en el cual tienen lugar cambios tan trascendentales en el desarrollo intelectual, y llevando a cabo un análisis multivariado de los procesos cognitivos que subyacen a la resolución de AWP. Por esta razón, las distintas muestras de participantes que forman parte de los estudios contenidos en este trabajo de tesis doctoral han sido seleccionadas tomando en consideración su pertenencia al grupo normotípico adolescente.

Como hemos tratado de exponer a lo largo de los epígrafes anteriores, la resolución de problemas aritméticos verbales es un proceso complejo en el que intervienen un gran número de habilidades y procesos cognitivos. Esta resolución está asociada a la construcción de un modelo situacional de los problemas que permita comprender en toda su extensión el estado inicial del que se parte para poder deducir, a partir de las relaciones que se establecen entre los elementos del problema, los procedimientos de cálculo requeridos en el problema. Consideramos, pues, que este ejercicio de comprensión aplicado al ámbito de las Matemáticas podría resultar de utilidad para la explicación y, quizá, el pronóstico del rendimiento académico de los estudiantes. Dada la transversalidad de las habilidades que se requieren para la resolución de AWP, creemos que la eficacia mostrada por los estudiantes podría resultar de utilidad como medida predictiva del rendimiento en áreas de conocimiento no necesariamente relacionadas con las Matemáticas o las ciencias, sino también en aquellas en las que se requiera un uso intensivo de las habilidades de comprensión y la reflexión crítica de los contenidos.

Este aspecto nos lleva a considerar el siguiente objetivo propuesto, que tiene que ver con el estudio de las diferencias entre dos tipos de problemas aritméticos: los AWP clásicos y los problemas de reflexión cognitiva (CRT). Consideramos que ambos problemas tienen características comunes, especialmente entre los problemas AWP inconsistentes y los CRT, dada la capacidad compartida de inducir una tendencia automática de respuesta superficial. Asimismo, los problemas CRT han demostrado poseer cierta capacidad para predecir el rendimiento académico. Es por ello que otro de los objetivos a considerar en este estudio consistirá en analizar los aspectos diferenciales que aporta cada tipo de problema en términos de precisión de respuesta y en relación al tipo de errores característicos que generan. Estas cuestiones serán abordadas en profundidad en el estudio 3.

Todas estas consideraciones constituyen, pues, las aportaciones novedosas en las que pretende contribuir este trabajo de tesis doctoral. A lo largo de los diferentes estudios, se analizará el papel que desempeñan algunas de las habilidades cognitivas generales, implicadas en la resolución de AWP en la adolescencia, como la cognición fluida, la comprensión lectora y el razonamiento, así como un conjunto variado de funciones ejecutivas de la MO (planificación, inhibición y actualización) y metacognitivas; en particular, la experiencia de dificultad percibida por los participantes. Además, se determinará la contribución específica de los procesos ejecutivos subyacentes a la resolución de los distintos tipos de AWP, lo cual podría ayudar a determinar las causas por la que muchos alumnos normotípicos yerran en el proceso de resolución de estos problemas.

Los objetivos e hipótesis particulares contemplados en cada uno de los estudios que forman parte de esta tesis doctoral serán descritos continuación:

2.2. Estudio 1

Arithmetic Word Problems Revisited: Cognitive Processes and Academic Performance in Secondary School

2.2.1. Objetivos

1. Analizar la relación entre distintas variables cognitivas generales (inteligencia fluida, comprensión y razonamiento deductivo) y la eficacia en la resolución de una tarea de problemas aritméticos verbales (AWP).

2. Comprobar la eficacia diferencial en la resolución de los distintos tipos de AWP: problemas consistentes, inconsistentes, de una y dos operaciones, más una multiplicación final.

3. Determinar en qué en qué grado estas medidas son capaces de explicar los resultados académicos obtenidos por 135 estudiantes de 2º y 3º de Educación Secundaria Obligatoria (ESO) en las asignaturas de Matemáticas e Historia.

2.2.2. Hipótesis

1. Esperamos diferencias significativas en el rendimiento de los AWP entre dos cursos, 2º y 3º de la ESO.

2. También esperamos diferencias en función del tipo de problema. Los AWP inconsistentes serán significativamente más difíciles que los problemas consistentes. Del mismo modo, los problemas aritméticos con dos operaciones de suma/resta serán significativamente más difíciles que los de una sola operación. El efecto de no consistencia de los prob-lemas aumentará la dificultad más que la del número de operaciones.

3. Las medidas cognitivas de razonamiento y comprensión lingüística se correlacionarán significativa y moderadamente con las medidas de eficacia en la resolución de los AWP. Las correlaciones con las medidas cognitivas serán mayores con problemas inconsistentes, y con los problemas de dos operaciones, debido a su mayor demanda cognitiva; en cambio, estas correlaciones serán más bajas en los problemas consistentes y los de una sola operación. Todas las medidas cognitivas, incluida la resolución de problemas aritméticos, correlacionarán positivamente con el rendimiento académico.

4. Dada la relación anterior, esperamos confirmar una capacidad de asociación entre las variables cognitivas medidas (razonamiento y comprensión) y la puntuación obtenida en la resolución de problemas aritméticos. Asimismo, las distintas variables cognitivas y, en particular, la puntuación global de los problemas aritméticos, mostrarán su capacidad explicativa significativa el rendimiento académico en Historia/Geografía y Matemáticas.

2.3. Estudio 2

The role of working memory updating, inhibition, fluid intelligence and reading comprehension in explaining differences between consistent and inconsistent arithmetic word problem solving performance.

2.3.1. *Objetivos*

1. Estudiar el papel de las medidas cognitivas utilizadas (la actualización, inhibición, la comprensión y la inteligencia fluida) en la explicación de las diferencias en la eficacia de resolución de AWP consistentes e inconsistentes en una muestra de 182 alumnos 5º curso de Educación Primaria (10 años).

2.3.2. *Hipótesis*

1. Basándonos en la bibliografía (Hegarty et al., 1995; Lewis y Mayer, 1987; Pape, 2003), esperamos encontrar un efecto de consistencia que diera lugar a un mayor porcentaje de respuestas correctas en los problemas consistentes.

2. La precisión con la que los participantes resolverán los problemas consistentes frente a los inconsistentes podría estar relacionada, no con el número de operaciones aritméticas en sí, sino con la construcción de la representación semántica del AWP.

3. La representación semántica del AWP requiere la reflexión y la aplicación controlada de procesos tales como la inhibición y la actualización de la información contenida en el AWP.

4. Consideramos que puede haber influencia específica de la actualización, la inhibición y la inteligencia en la precisión de la resolución del AWP en el mismo modelo unitario, así como el papel mediador de la comprensión lectora, teniendo en cuenta el tipo de problema

5. Por último, la hipótesis de que los AWP C-I (una operación consistente y una inconsistente) serán más difíciles de resolver (efecto del orden de la consistencia) que los AWP I-C (una operación inconsistente y una consistente), porque exigen un mayor esfuerzo de actualización, ya que la segunda operación inconsistente (más difícil) tiene que ser resuelta cuando la WM ya está sobrecargada con el resultado de la primera operación consistente.

2.4. Estudio 3

Reflection and intuition in arithmetic word problems: cognitive and executive processes in secondary school students.

2.4.1. Experimento 1

2.4.1.1. Objetivos

1. Comparar la eficacia en la resolución de dos tipos de problemas matemáticos, AWP y CRT en una muestra de 81 estudiantes de 4º de la ESO y explorar sus relaciones con la inteligencia fluida y la comprensión lectora.

2. Analizar el tipo de respuestas dadas por los participantes (correctas, superficiales e incorrectas) en ambos tipos de problemas con el objetivo de inferir un patrón de respuestas asociado a la tarea.

2.4.1.2. Hipótesis

1. Del mismo modo que en los estudios precedentes, la respuesta más frecuente en los AWP serán las respuestas correctas, y los errores más frecuentes serán las respuestas erróneas no superficiales. En cambio, esperamos encontrar una mayor frecuencia de respuestas superficiales en CRT. También, esperamos una menor proporción de respuestas correctas y respuestas superficiales en CRT que en los AWP Inconsistentes.

2. En cuanto a los dos tipos de AWP, esperamos encontrar un número significativamente mayor de respuestas correctas en los problemas consistentes que en los inconsistentes. También esperamos encontrar una dificultad mayor en los problemas de dos operaciones que en los de una operación.

3. Dada la necesidad común de construir un modelo situacional de los problemas para obtener una solución correcta, esperamos encontrar intercorrelaciones positivas entre la

precisión de respuesta de AWP inconsistente y CRT, así como entre las respectivas respuestas superficiales de ambos problemas. También esperamos encontrar una relación moderada de las medidas cognitivas, inteligencia fluida y comprensión lectora, con el AWP y el CRT. En consecuencia, los análisis de regresión informarán de un poder de asociación significativo de los problemas con las medidas cognitivas.

2.4.2. Experimento 2

2.4.2.1. Objetivos

1. Confirmar los resultados obtenidos en el experimento anterior en relación a la precisión de las respuestas en AWP y CRT en una muestra de 82 estudiantes de 3º y 4º de Educación Secundaria Obligatoria.
2. Relacionar la precisión de respuesta en los problemas con la experiencia metacognitiva de dificultad de los participantes autoevaluada.
3. Explorar las relaciones entre la precisión de respuesta en AWP y CR con nuevas medidas de funcionamiento ejecutivo (inhibición, actualización y planificación), así como con las medidas cognitivas anteriores (inteligencia fluida y comprensión lectora).
4. Comprobar la capacidad de AWP y CRT, junto con las variables cognitivas, en la explicación del rendimiento académico de los estudiantes.

2.4.2.2. Hipótesis

1. Esperamos confirmar los principales resultados obtenidos en el experimento 1. La respuesta más frecuente en los AWP será la correcta. De las dos variables que afectan a los problemas de AWP, la consistencia/inconsistencia será un factor más relevante en el incremento de la dificultad que el número de operaciones. El porcentaje de respuestas correctas será mayor en AWP que en CRT y el porcentaje de respuestas superficiales será mayor en CRT que en AWP.

2. Esperamos encontrar diferencias significativas en la ratio de dificultad percibida por los participantes en los distintos tipos de AWP. Los problemas consistentes serán calificados como más fáciles que los inconsistentes. La dificultad de los problemas CRT será subestimada, es decir, se considerarán más fáciles de lo que realmente son.

3. Habrá correlaciones positivas entre las medidas cognitivas y ejecutivas, y los problemas AWP y CRT. Este patrón de correlaciones será diferente según el tipo de problema. El AWP inconsistente y el CRT también estarán interrelacionados.

4. Esperamos una asociación relevante de la actualización, la inhibición, la planificación, la inteligencia fluida y la comprensión lectora en el rendimiento de las tareas AWP y CRT. Los análisis de senderos (path analyses) confirmarán estas asociaciones.

5. Esperamos una contribución significativa de las dos habilidades cognitivas básicas (comprensión lectora e inteligencia fluida), así como de AWP y CRT en la explicación de los resultados en ambas medidas de rendimiento académico (calificación global en Matemáticas y Media académica global a final de curso).

3. ESTUDIOS EMPÍRICOS

3.1. ESTUDIO 1

Duque de Blas, G., Gómez-Veiga, I., & García-Madruga, J. A. (2021). Arithmetic Word Problems Revisited: Cognitive Processes and Academic Performance in Secondary School. *Education Sciences*, *11*(4). https://doi.org/10.3390/educsci11040155

education sciences

MDPI

Article

Arithmetic Word Problems Revisited: Cognitive Processes and Academic Performance in Secondary School

Gonzalo Duque de Blas [1,*], Isabel Gómez-Veiga [2] and Juan A. García-Madruga [2]

1 La Escuela Internacional de Doctorado de la UNED (EIDUNED), Department of Developmental and Educational Psychology, Faculty of Psychology, National Distance Education University (UNED), Calle de Juan del Rosal, 10, 28040 Madrid, Spain
2 Department of Developmental and Educational Psychology, Faculty of Psychology, National Distance Education University (UNED), Calle de Juan del Rosal, 10, 28040 Madrid, Spain; igveiga@psi.uned.es (I.G.-V.); jmadruga@psi.uned.es (J.A.G.-M.)
* Correspondence: gduquedeblas@gmail.com or gduque8@alumno.uned.es

Abstract: Solving arithmetic word problems is a complex task that requires individuals to activate their working memory resources, as well as the correct performance of the underlying executive processes involved in order to inhibit semantic biases or superficial responses caused by the problem's statement. This paper describes a study carried out with 135 students of Secondary Obligatory Education, each of whom solved 5 verbal arithmetic problems: 2 consistent problems, whose mathematical operation (add/subtract) and the verbal statement of the problem coincide, and 3 inconsistent problems, whose required operation is the inverse of the one suggested by the verbal term(s). Measures of reading comprehension, visual–spatial reasoning and deductive reasoning were also obtained. The results show the relationship between arithmetic problems and cognitive measures, as well as the ability of these problems to predict academic performance. Regression analyses confirmed that arithmetic word problems were the only measure with significant power of association with academic achievement in both History/Geography (β = 0.25) and Mathematics (β = 0.23).

Keywords: arithmetic word problems; reading comprehension; reasoning; academic performance; secondary school

check for updates

Citation: Duque de Blas, G.; Gómez-Veiga, I.; García-Madruga, J.A. Arithmetic Word Problems Revisited: Cognitive Processes and Academic Performance in Secondary School. *Educ. Sci.* **2021**, *11*, 155. https://doi.org/10.3390/educsci11040155

Arithmetic Word Problems revisited: cognitive processes and academic performance in secondary school

Abstract: Solving arithmetic word problems is a complex task that requires individuals to activate their working memory resources, as well as the correct performance of the underlying executive processes involved in order to inhibit semantic biases or superficial responses caused by the problem's statement. This paper describes a study carried out with 135 students of Secondary Obligatory Education, each of whom solved 5 verbal arithmetic problems: 2 consistent problems, whose mathematical operation (add/subtract) and the verbal statement of the problem coincide; and 3 inconsistent problems, whose required operation is the inverse of the one suggested by the verbal term(s). Measures of reading comprehension, visual-spatial reasoning and deductive reasoning were also obtained. The results show the relationship between arithmetic problems and cognitive measures, as well as the ability of these problems to predict academic performance

Keywords: Arithmetic word problems; reading comprehension; reasoning; academic performance; secondary school.

3.1.1. Introduction

"Understanding a problem and solving it is practically the same thing" – Rumelhart (1980)

Mathematics is a type of language through which we construct models that represent the qualities and relationships of measurable reality. Considered the "language of science", the learning of its elementary rules is essential for the development of mathematical thought and constitutes one of the fundamental objectives of formal education. This learning process, not without its difficulties (Nelson & Powell, 2018; Peng et al., 2019), often determines the academic and future employability of its students, see (Caspi et al., 1998; Marjoribanks, 2005; Orrantia, 2006; Sánchez-Pérez et al., 2015; Watt et al., 2012), something well known to students whose level is, in general terms, below the European average according to the Organisation for Economic Co-operation and Development (OECD, 2019).

The present paper addresses the relationship that reading comprehension and reasoning processes in arithmetic word problems (AWPs) have with academic achievement in secondary school, focusing not only on classic abstract reasoning, but also on new cognitive measures of verbal deduction. We argue that deduction and reading comprehension are involved in arithmetic word problem solving and most of the complex learning tasks students commonly face at school and constitute an important higher-order cognitive ability that underlies academic achievement.

Different theoretical approaches have been proposed to explain the components and processes involved in AWP solving. Among the most important theories, schema-based models (Kintsch & Greeno, 1985; Rumelhart, 1980), situational models (Kintsch, 1998; Reusser, 1989), and contributions from the semantic field (Bassok et al., 1995, 1998; Gros et al., 2020; Thevenot, 2010) stand out. This paper is focused on the schema and situational contributions, Kintsch and Greeno (Kintsch & Greeno, 1985) and Kintsch's comprehension model (Kintsch, 1998), as well as dual process theories of thinking and reasoning (Evans, 2007; García-Madruga et al., 2007). Its aim is double. First, to analyze the relationship that AWPs have with reading comprehension and abstract reasoning. To this end, we deepen our understanding of cognitive processes involved in solving a particular kind of mathematical reasoning problem: consistent and inconsistent change problems with two or three arithmetic operations. The second objective is to check the capacity of students to solve these AWPs to predict academic performance in two basic and different subject matters: Mathematics and History/Geography.

Although some of the skills needed to solve these kinds of problems may be different from those used in solving geometric problems or those of probabilistic reasoning, it is possible to determine the existence of certain common features that underlie all mathematical problems. In particular, it is worth highlighting the implications of a kind of general cognitive aptitude based on abstract reasoning, fluid intelligence (Cattell, 1963; Engle et al., 1999; Fung & Swanson, 2017; Horn & Cattell, 1967; Jaeggi et al., 2008), that is identified with processes related to inductive reasoning (the search for rules or patterns that relate stimuli in order to establish more general conclusions), deductive reasoning (those that draw logical conclusions from a sequence of previous premises), and quantitative reasoning (those that involve concepts of a quantitative type, such as AWPs; (Sternberg & Ben-Zeev, 1996).

Arithmetic Word Problems (AWPs)

AWPs must be solved by combining the numbers mentioned in the text using basic arithmetic operations (addition, subtraction, multiplication, and division). Solving AWPs is a complex task that demands individuals to activate the underlying working memory executive processes in order to inhibit non-relevant information, semantic biases, overlearned procedures (heuristics) or superficial responses induced by the statement of the problem (de Koning et al., 2017; Jiang et al., 2020; Lubin et al., 2016; Passolunghi & Pazzaglia, 2004, 2005; Passolunghi & Siegel, 2001; Shum & Chan, 2020).

There are a diversity of mistakes that students commit when solving arithmetic problem (Abdullah et al., 2015; Hasanah et al., 2017; Hegarty et al., 1992; Rohmah & Sutiarso, 2018). Two of them are (1) errors in the calculation required to find the solution (execution phase) and (2) errors derived from a poor understanding of the text and the relationships described. By studying these errors, we can delineate two key aspects regarding these kinds of problems. On the one hand, AWPs represent not only a demonstrative exercise of the calculation skills necessary, but also a way to test a student's textual comprehension and reasoning, as well as an associated measure of student academic performance. This is the main hypothesis considered in the present study. Secondly, given the variety of basic cognitive processes involved in their resolution, AWPs could be used as a screening tool to detect signals of possible learning difficulties in reading and Mathematics.

There is a wide variety of AWPs, which can be classified according to values they can adopt on a set of variables. We have tried to synthetize the most relevant ones in Table 1.

These criteria would affect to the global extension of AWP and the difficulty solving them. The present study was carried out with 5 "change" AWPs (see Materials section for more details), in which only the number of operations and the direction of verbal term used (consistent vs. inconsistent) were manipulated to modulate the difficulty of the task.

Table 1. Criteria and values used to classify arithmetic word problems (AWPs).

Criteria	Values
General types of AWP	Change, Combine, Compare or Equalizing problems (Riley et al., 1984)
Items named in the problem	
- Number of elements	The more significant elements, the greater number of steps needed for solving.
- Context of elements	Familiar or non-familiar for solver (Bassok et al., 2008; Guthormsen et al., 2016)
- Semantic Align between elements	Symmetric or asymmetric align (e.g.: content vs container) [12]
Quantities related to elements	
- Magnitude	Large or small value size (Thevenot & Oakhill, 2005)
- Data type	Integers or decimal magnitudes (Bassok et al., 1998)
- How they are represented	Cardinals (unordered entities) or Ordinals (ordered entities) (Gros et al., 2019)
- How they are expressed	Explicitly (e.g.: x = 10) and/or relationally (e.g.: y = x + 2) (Gros et al., 2019)
- What they are connected with	Verbal terms: "More than"; "Less than", "Equals to"; "As many as" (Riley et al., 1984)
- Sense of inserted verbal term	Consistent or inconsistent with the suggested operation (Hegarty et al., 1992; Riley et al., 1984)
Problem question	
- Number of questions	From 1, to above
- Location	At start or end of the problem (Thevenot et al., 2007)
- Type of question	Referring to the overall outcome of any specific part, or to the whole of the parts involved
- Demanding data	Numeric or qualitative data (comparative; e.g. "Who has more marbles?")
Operations	
- Types	Addition, subtraction, multiplication and/or division (Fischbein, 1989)
- Number or steps	From 1, to above

AWP and Reading Comprehension

AWPs have led to numerous studies and explanatory models regarding the processes involved in their resolution. Among the most outstanding is the work of Kintsch and Greeno (1985) and Kintsch (1998), who applied the notion of schema and the construction–integration model of text comprehension to solving AWPs. These authors showed that, in the same way calculation skills are indispensable for solving these problems, it is equally necessary to consider the different ways these problems can be understood in order to be able to determine the way success occurs. This observation came from the growing evidence that the most frequent errors committed in the resolution of AWPs resulted from the inadequate construction of the mental representation of the problem (Shum & Chan, 2020; Sternberg, 1996), which could also be derived from semantic biases induced by the statement of the problem (Gros et al., 2020).

Along this same line, numerous authors (Hegarty et al., 1992, 1995; Lewis & Mayer, 1987; Riley et al., 1984) have studied the resolution of AWPs through the use of non-routine problems. The resolution of an AWP requires an in-depth understanding of what is said in the text in order to realize what the question refers to, as well as to determine the meaning or direction of the verbal terms used to connect the objects with the quantities and to show the changes or variations narrated in the statement. Both are key conditions to allow the arithmetic formalization stage to be executed following the reliable sense of the relationships reported in the problem statement. However, participants might use two different types of resolution strategies on their way to reach it. The first strategy is the *direct strategy*, one best summarized in the words of Siegler and Jenkins (1989): "calculate first, think later." In the same line, other authors related a "compulsion to calculate" (Stacey & MacGregor, 1999) due to children's prior experience with arithmetic (Khng & Lee, 2009). This process would begin by identifying the numbers of each category with which participants, subsequently, would perform the arithmetic operations they deem necessary based on the literal relationships described within the problem. This strategy of direct translation often leads to mistakes in the resolution of problems due to two reasons or errors: (1) that of understanding, and (2) that of metacognition. The first is due to a poor understanding of what the problem represents, that is, by focusing more on its quantitative aspects than on its qualitative ones (Boote & Boote, 2018; Chi et al., 1988; Sternberg & Frensch, 1992). The second is due to an inadequate estimation of the problem's difficulty, perhaps as a consequence of identifying the problem as an ordinary exercise for which a previous model and resolution plan are already available. This direct strategy is, therefore, a strategy largely dominated by intuitive and automatic Type 1 processes, whereas Type 2 reflexive processes appear absent.

The second strategy for solving arithmetic problems is called the *problem model strategy*. It consists of several stages that, together, guide the participant toward the construction of a mental representation of the problem based mainly on its underlying qualitative aspects (Bassok et al., 1998; Chi et al., 1988; Kahneman, 2011). The first stage begins with understanding the premises of the problem, a process in which each local piece of information is transformed into propositions that represent the elements described in the text of the problem. The result of this transformation forms a latent semantic network known as a "text base", from which a situational model of this problem is then constructed in a second phase (Kintsch, 1998). This model is the result of integrating the content of the text base with that of related content and experiences stored in long-term memory. In this way, the characteristics of the objects described in the text, as well as the key words of the problem, are integrated into differentiated entities, forming a mental representation that is clearly more elaborate than that resulting from a direct translation process (Kintsch, 1998). In this phase, therefore, the ability to recover information available from memory and relate it to newly acquired data in working memory (WM) takes on special relevance. This phase requires a capacity to reason reflexively and inhibit any automatic responses that may be generated during the process, given that their absence could generate an inadequate representation of the problem.

The situational representation allows a participant to initiate a third phase whose objective is to elaborate a plan to solve the problem posed, determining with exactitude the arithmetical operations that are necessary given the relations between the different objects in the problem. This is an especially relevant component in AWPs, given that one of the elements by which problem difficulty is manipulated is through the use of adverbs of quantity, such as “more”, “less” or “equal”. It is these adverbs that interrelate the objects of the problem and allow people to infer the meaning of the established relationships. In this way, the operations that are necessary in the process of calculating the response to the problem—addition, subtraction, multiplication, and/or division—are determined. This demonstrates the close relationship between the stage of generating a plan for resolving the problem and the previous stage during which the text base within a situational model is transformed. That is, the adverbs of quantity within the problem statement cannot in themselves serve to identify the arithmetic calculations required for its resolution, a contrary interpretation being similarly possible, the result of which would consequently require the inhibition of any generated automatic response.

Regarding the problems used in this study, change problems are those in which the problem statement narrates an initial state in which the objects/subjects described are associated with some quantities, a transformation process that increases or decreases the quantities described, and a final question that motivates the purpose of the problem and requires the calculation of final quantities. Looking at the sense of the verbal terms inserted, we can distinguish two kinds of compare problems: (1) problems with a coherent or consistent statement, when the adverb “more” or “less” coincides with the operation of addition or subtraction, respectively, that must be performed to enact resolution; and (2) non-coherent or inconsistent problems, when the adverb suggests performing an inverse arithmetic operation to the literal one. That is, performing subtraction in the presence of the term “more”, or addition in the presence of the term “less”. In this phase, a higher-order ability would mediate a resolution strategy/plan that establishes the sequences of operations necessary to respond to the problem posed. Ultimately, the resolution process ends with a planned execution stage during which one consciously supervises the process to minimize any calculation errors that may occur.

Although the same stages can be identified in both the "direct strategy" and the "strategy of the problem model", the fundamental difference between the two is found in the second phase, that is, in the way in which information is integrated from the text base. In the "direct strategy", this integration is limited and focuses on the numbers and keywords in the problem text. Conversely, in the other strategy, the situational model is oriented toward the objects of the problem, and this allows for the integration of its relevant characteristics as well as the sense of the relationships established within the problem statement.

In the arithmetic problems of our study, the verbal problem statement may contain two adverbs of quantity when describing the price of the objects in three different places. The description of a greater number of relationships between the terms necessarily involves an increase in the number of propositions in the text base that must be integrated, as well as a greater number of steps or operations that must be planned and executed in the resolution. There are, therefore, two types of problems based on this criterion: (1) problems of a single operation, in which a relationship requires the operation of addition or subtraction in order to solve it (e.g., "Juan has 5 cookbooks. Isabel has two books less than Juan. How many cookbooks does Isabel have?"); and (2) problems of two operations, which describe two chains of relationships that require partial calculations to obtain the final result (e.g., Isabel has two books less than Juan. Marta has three more books than Isabel. How many cookbooks does Marta have?") (Mayer & Hegarty, 1996).

The number of relationships described in the problem and the consistency or non-consistency of the adverbs that provide information regarding their meaning are both variables of enormous experimental potential. This is due to the close relationship they have with basic cognitive processes and their high ability to modulate the difficulty of the task.

AWP and Reasoning

Neither intelligence nor mathematical thought are limited to deduction. However, this reasoning process has been correctly considered the essence of mathematical thought. Deduction implies a process of sequential reflection through a series of steps that lead to a conclusion. This conclusion can be considered the consequence of consolidating or representing a mental model of the premises of the deductive task (Johnson-Laird, 1983). The tendency to use deductive reflection, as opposed to intuition in the process of problem solving, constitutes for many theorists one of the fundamental differences between expert and novice arithmeticians (Sternberg & Ben-Zeev, 1996). Numerous studies have shown the differences between the intuitive and analytical modes of thought, giving rise to a dual theory of processing in which two types of systems are distinguished (Evans, 2003, 2007; García-Madruga et al., 2007). The first system, Type 1, is intuitive, guided by data, and is therefore automatic. It acts quickly, and only its final product is accessible to consciousness (Kahneman, 2011). Thus, the particular modus operandi of the Type 1 system tends to provide answers based on heuristics (Kahneman & Tversky, 1972; Tversky & Kahneman, 1973) that reduce the complexity involved in evaluating possibilities as well as the need for more cognitive resources, a fact demonstrated by the low correlation this process has with measures of intelligence and working memory (WM; Stanovich & Toplak, 2012). The Type 2 system, on the other hand, is slow, costly both cognitively and motivationally, and guided by will and conscience: it is based on WM performance (Evans, 2003, 2007; Evans & Over, 1996; Kahneman & Frederick, 2002; Sloman, 1996; Stanovich & West, 2000). The fundamental characteristic of the Type 2 system is its ability to generate a cognitive decoupling in the process of solving problems by means of which the automatic responses generated by Type 1 are inhibited and the most characteristic abilities of the human being are enabled; that is to say, hypothetical deductive thought, or mental simulation executed via the generation of different mental models of the problem, and decision making itself (Evans & Stanovich, 2013a; Stanovich & Toplak, 2012). The Type 2 system therefore requires an intensive use of WM executive functions, such that tasks which necessitate its use tend to correlate clearly with measures of WM and fluid intelligence.

Fletcher and Carruthers (Fletcher & Carruthers, 2012) propose three main characteristics that define Type 2 reasoning: (1) it is subject to intentional control, (2) it can be guided by normative beliefs about appropriate reasoning methods (see, Gómez-Chacón et al., 2014), and (3) it cancels out, or in its case confirms, the unreflective responses automatically generated by the Type 1 system. These characteristics describing the supervision of inferential processes are skills that occur later in development, emerge during preadolescence, and culminate their development in youth (Moshman, 2004; Pillow, 2002; Santamaría et al., 2013). Before its acquisition, children lack the ability to activate the Type 2 system and control the interaction between both systems.

Aims

The purpose of this paper is to help to clarify the nature of mathematical thinking by analyzing the relationship between some of the cognitive variables that underlie it and to determine to what extent they are related to the academic achievement of Secondary Obligatory Education. To do this, we compare 5 arithmetic problems of increasing difficulty (consistent or inconsistent, and of either one or two arithmetic operations, plus a multiplication across all cases), as well as two measures of reasoning (Kaufman brief intelligence test KBIT; Kaufman & Kaufman, 2013) and a deductive reasoning test (DRT; García-Madruga, et al., 2014) and two of reading abilities (reading processes assessment battery (PROLEC-SE; Ramos & Cuetos, 1999) and The Spelling test for secondary school (Spelling-SE; Duque de Blas & Gómez-Veiga, 2018). A detailed description of the problems used can be seen in Table 2 and in the Supplementary Materials.

Table 2. Description of problems in terms of number of consistent or inconsistent operations (working memory load), inconsistency, and need of inhibition of superficial responses.

	WM Load Number of Operations (Add/Subtract + Mult.)	Inconsistency	Superficial Responses
Problem-1	1 + 1	No	No
Problem-2	1 + 1	Yes	Yes
Problem-3	2 + 1	No–No	No
Problem-4	2 + 1	No–Yes	Yes
Problem-5	2 + 1	Yes–Yes	Yes

The first two problems require two operations, one addition or subtraction, and a multiplication. Problem 2 differs in that the operation that must be applied (addition or subtraction) is inconsistent with the adverb of quantity in the literal statement ("less" or "more", respectively). Consequently, Problem 2 will require the inhibition of a surface representation in the text base created from the literal statement. When this inhibition does not happen, it will produce a superficial response that we could define resulting from the correct performance of a mistaken plan. This inhibition is needed in order to construct a representation that allows the participant to apply the inverse arithmetic operation of that indicated by the adverb of quantity (Shum & Chan, 2020). Apart from correct and superficial responses, any other response is also erroneous and involves either arithmetical mistakes or erroneous answers probably produced by a low commitment of the participant to the task. Problems 3, 4, and 5 are problems of three arithmetic operations, because they describe two relationships between the terms involved, which implies a greater demand on WM resources, and a final multiplication. These problems involve two addition or subtraction operations and may be consistent (Problem 3), require a consistent and an inconsistent operation

(Problem 4), or require two inconsistent operations (problem 5). Consistent problems have no superficial responses, inconsistent problems 2 and 4 have one possible superficial response, while inconsistent problem 5 has three possible superficial responses: in the first operation, in the second or in both.

The number of operations is a source of problem difficulty (Castro-Martínez & Frías-Zorilla, 2013; Hegarty et al., 1992; Viterbori et al., 2017), and it would be necessary to clarify whether the increase in the number of operations implies any qualitative increase in the problem difficulty, and what kind of mistakes are the most common, superficial or other error responses. In all cases, the same response schema is applied, with each having one additional operation, involving just a temporary increase in WM's updating process required to solve the problem. In the same way, the final multiplication required of all problems is nothing less than trivial, the calculation of which, in principle, poses few difficulties. In this way, Problem 3 should be more difficult than Problem 1, and Problem 4 more difficult than Problem 2.

When solving an AWP, we consider four possible scenarios as a response to these problems, which define the three types of mistakes that solvers would incur. First, the ideal scenario consists of correctly solving a correct plan; second, when solvers determine correctly the arithmetic embedded in the problem, but makes computational mistakes, we say that they are wrongly solving a correct plan; the reverse case, when a solver is able to compute correctly the operations, but the type of operations are wrongly determined, we says the solver is solving correctly a wrong plan; and last, when they are unable to determine the arithmetic of the problem and its required computation, solvers are wrongly solving a wrong drawn plan.

In terms of cognitive abilities, these mistakes can be caused, fundamentally, by executive functions' poor combined performance, in particular, the inhibition of superficial responses and the updating of base of text along the whole process. Thus, we might identify four kinds of solvers' patterns of performance, whose outcomes can be also identified by a particular scenario and kind of response (see Table 3).

Table 3. Solvers profiles attending their Executive Functions (EF) performance, scenario, and outcomes.

Profiles	EF performance		Scenarios		Response
	Inhibition	Updating	Plan	Execution	
1	Good	Good	Correct	Correct	Correct
2	Good	Bad	Correct	Incorrect	Error
3	Bad	Good	Incorrect	Correct	Superficial
4	Bad	Bad	Incorrect	Incorrect	Error

An additional profile can be considered when participants are not willing to face the task. In that case, results would be contaminated by a higher number of errors due to a randomized response.

If our conception is correct, the main difficulty with the problems comes from the process of reflection and inhibition involved in solving their non-consistency, as well as the need to update the base of text with the partial results of each problem according to a new interpretation of the relational magnitudes. Given the scenarios described above and a good participant engagement with the task, it seems reasonable to think that there is a higher probability of finding "errors" as a response (scenarios 2 or 4) than "superficial" responses (3), due to the greater number of them (double) in which an "error" response can be recorded as an outcome. However, we do not know whether such a claim can be empirically substantiated. In any case, the increase in the difficulty in the inconsistent compared to the consistent problems is an effect that has been confirmed by Verschaffel (1994), who in his study of 10–11-year-old children found a 71% success rate in the resolution of inconsistent problems, versus 82% in consistent problems. Likewise, Lewis and Mayer (1987) and Hegarty et al. (1992) confirmed in adults that the non-consistency of the statements caused a greater number of reversibility errors with two-step problems.

Our hypotheses were:

1. We expect significant differences in AWP performance between two grades, 2nd and 3rd grade of Compulsory Secondary School courses.

2. We also expect differences as a function of type of problem. Inconsistent AWPs should be significantly more difficult than consistent problems. In the same way, the arithmetic problems with two add/subtract operations should be significantly more difficult than those of a single operation. The non-consistency effect of the problems should increase the difficulty more than that of the number of operations.

3. The cognitive measures of reasoning and linguistic comprehension should correlate significantly and moderately with measures of effectiveness in the resolution of AWPs. The correlations should be higher between the cognitive measures and those of the inconsistent problems and the problems of two operations, due to their greater cognitive demand, than with the consistent problems and those of a single operation. All cognitive measures, including the resolution of arithmetic problems, should correlate positively with academic performance.

4. Given the previous relationship, our aim was to confirm the capacity of association of the cognitive variables measured (reasoning and comprehension) on the score obtained in solving arithmetic problems. Likewise, the various cognitive variables and, in particular, the global arithmetic problems' score, should show their capacity for association with academic performance in History/Geography and Mathematics.

3.1.2. Method

3.1.2.1. Participants

This study was carried out with a sample of 135 native Spanish speaking students (53 female) of compulsory secondary education, without learning difficulties, of a public institute (Madrid), of which 70 2nd grade students (24 female) and 65 3rd grade students (29 female) voluntarily participated. The average age in the 2nd grade was 13.10 (SD = 0.85) and in the 3rd grade 14.18 (SD = 0.80).

3.1.2.2. Task and Measures

Arithmetic Word Problems task (AWP)

To measure the ability to solve arithmetic problems, we used a new test based on the problems of Hegarty et al. (1992). It is formed by 5 AWPs: two consistent and three inconsistent. The problems involve elementary arithmetic calculations consisting of addition or subtraction, as well as multiplication, and varying their number in each problem (one addition or subtraction operation in problems 1 and 2, two addition or subtraction operations in problems 3, 4 and 5, plus a final multiplication operation across all problems). In order to control the possible semantic biases derived from the statement, the elements within the same problem were always the same (semantic align), but varying some of their characteristics: origin, size, color, material, price, and commercial areas where they can be acquired. The question of each problem was always related to the global price that solvers must pay (cardinality) for a specific quantity of objects in a particular store, so we can consider this context as familiar to solvers. The quantities described in the statement were some explicit and some relational. The problem's questions were always located at the end of the statement. The result of the operations necessary to calculate the requested response were always integers and never bigger than three digits. Correct and error responses were collected in all problems, as well as superficial responses for inconsistent problems. Two parallel versions of the problems were used, where the adverbs of quantity "more" and "less" were reversed, as were the place and object names. See the Appendix for an example of the 5 problems employed. Its measurement range is between 0 and 5. Cronbach's alpha for internal consistency was 0.75.

Reading processes assessment battery (PROLEC-SE)

The participants completed the task number five of Reading processes assessment battery (Ramos & Cuetos, 1999). This evaluates reading processes for secondary students and consist of two expository texts. Questions are divided in two kinds: those with enquiries regarding literal aspects of text and those which require making inferences. Their measurement range is between 0 and 20. Cronbach's alpha for internal consistency was 0.74.

Kaufman Brief Intelligence Test (KBIT)

Intelligence was evaluated through the matrix subtest of the Kaufman Intelligence Brief Test (Kaufman & Kaufman, 2013), in its Spanish version. This test provides a measure of abstract and visuospatial reasoning, and fluid (non-verbal) intelligence. Its measurement range is between 0 and 48. Cronbach's alpha for internal consistency was 0.81.

Deductive reasoning test simplified (DRTs)

The DRTs is a simplified version of the DRT (García-Madruga, et al., 2014) adapted to suit preadolescent participants (12–15 years old). It consists of nine propositional and syllogistic deductive and meta-deductive reasoning problems, divided into four types of inference problems: propositional deductive, propositional meta-deductive, syllogistic deductive, and syllogistic meta-deductive. The propositional deductive inferences include two inclusive disjunction problems (one affirmative and one negative). Participants must evaluate the possible conclusions of these two inferences. Both require the construction of multiple models. The propositional meta-deductive inferences include three truth-table problems in which participants have to analyze the consistency of three problems, each consisting of a conditional statement and an assertion. In the first problem, the assertion matched the first initial model of the conditional. The second problem requires a participant to construct the second conditional model in which antecedent and consequent are negated. Finally, the last problem requires the construction of the third and most difficult conditional model in which the antecedent is negated but the consequent is affirmed. The syllogistic deductive task requires participants to generate and write the solution to one categorical syllogism. Categorical syllogisms include the combinations of the four kinds of premises: universal affirmative, universal negative, the particular affirmative, and particular negative. Categorical syllogisms can be very difficult, often too difficult for preadolescents. Thus, we used only an easy, single model categorical syllogism. Finally, in the syllogistic meta-deductive necessity/possibility task, reasoners have to decide whether a given conclusion in three syllogistic problems is necessarily true, possible, or impossible. Its measurement range is between 0 and 9. Cronbach's alpha for internal consistency was 0.73.

Spelling-SE for secondary school

The spelling test for secondary school (Duque de Blas & Gómez-Veiga, 2018)is a lexical choice test in which participants have 40 words represented orthographically, each next to two pseudo-homophones. For example: abeja, abega, aveja ("abeja" means "bee"; the other two are meaningless words but would be pronounced very similarly). The objective is to select the correct orthographic form for each word. This test is based on the spelling subtest included in the Reading Assessment Battery (BEL; López-Higes et al., 2002). Distinct levels of usage frequency (high, medium, and low), described in Lexesp -a computerized lexicon of Spanish (Sebastián et al., 2000)- were taken into account when selecting the words contained in this test. Its measurement range is between 0 and 40. Cronbach's alpha for internal consistency was 0.76.

Academic achievement

The final grades in Mathematics and History/Geography (from 1 to 10) at the end of the school year were taken as a measure of academic achievement. As for the latter, it is a discipline whose study requires the student to think critically about events, as well as to contextualize and relativize them in terms of past culture and social organizations in a way that allows drawing reasonable conclusions after evaluating alternative points of view. The growing effort to promote such critical reflection skills in history teaching is remarkable (McLaughlin & McGill, 2017; Reed & Kromrey, 2001; Williams & Worth, 2001; Yang, 2007) because they can be used as a predictor of academic performance. Moreover, participation of an adolescent sample seemed particularly relevant, as reasoning skills are increasingly important during this developmental period (Barrouillet, 2011), when learning activities become more complex. The final scores are the average result obtained during the three trimesters making up the school year. These scores reference the aptitude and behavioural characteristics of the students.

3.1.2.3. Procedure

Each of the two symmetrical versions of the AWPs was randomly assigned to each of the participants. The participants had to solve the tasks in two sessions distributed over a week. The order of test presentation for the first session was PROLEC-SE, DRTs, and arithmetic, while in the second session, they performed KBIT and Spelling-SE, in that order. All the tests were completed digitally via online forms in the computer room during school hours. The protocols of this study were approved by the ethics committee of the National University of Distance Education (UNED).

3.1.2.4. Data Analyses

The analyses were carried out using the statistical package for social sciences, SPSS v25. A winsorization procedure (Field, 2013; Wilcox, 2011) on global and partial scales of the AWP task, as well as on the global scores of the tasks, was performed in order to control the outliers. Descriptive results on the tasks, ANOVAs, bivariate Pearson correlation coefficient and hierarchical regressions among cognitive measures, AWPs, and academic achievement were run in order to test the hypothesis regarding the relationship between the cognitive measures with the performance on AWP, and the power of association of these variables on the academic achievement.

3.1.3. Results

Descriptive Statistics and Comparisons

The first analysis aimed to determine if there was any difference between versions (more or less) of AWPs. The results confirmed the non-existence of statistically significant differences as a function of which arithmetic test version was carried out (F (1, 133) = 0.365, p = .547).

The results of the tasks, including mean number of correct responses, standard deviation, range, number of participants, and differences between scholar grades, are shown in Table 4. The most difficult task in terms of percentage of correct responses was the AWP task.

Table 4. Global descriptive statistics (mean of number of correct responses, range, standard deviation, and mean differences between 2nd and 3rd grades).

N = 135	Mean	Range	Percentage	SD	Difference between 2nd (N = 56) and 3rd (N = 57)			
					M (SD)–M (SD)	*F*	*p*	*Cohen's d*
AWP-Global	2.22	0–5	44	1.47	1.66 (1.18)–2.83 (1.52)	25.41	.001	0.80
PROLEC-SE	11.18	0–20	56	3.89	10.01 (4.38)–12.43 (2.81)	14.3	.001	0.62
Spelling-SE	30.74	0–40	77	2.68	30.34 (2.68)–31.17 (2.64)	3.26	.073	0.31
KBIT	35.10	0–48	73	4.55	34.24 (4.91)–36.02 (3.96)	5.28	.023	0.39
DRTs-Global	4.74	0–9	53	1.37	4.60 (1.36)–4.89 (1.38)	1.54	.217	0.21

Likewise, we can observe that the differences in mean correct response between the two academic years were statistically significant, except for the DRTs and Spelling-SE. Results in KBIT and PROLEC-SE were as expected for their age and educational level.

The descriptive statistics and results obtained for each of the AWPs, including mean of correct and superficial responses for each problem, can be seen in detail in Table 5.

Table 5. Global descriptive statistics (range, mean, standard deviation) of arithmetic word problem (AWP) measures (correct, superficial, and error responses).

	Range	2nd (N = 70) M (SD)	3rd (N = 65) M (SD)	Total (N = 135) M (SD)
AWP-1	0–1	0.63 (0.49)	0.82 (0.39)	0.72 (0.45)
AWP-2	0–1	0.27 (0.45)	0.45 (0.50)	0.36 (0.48)
AWP-2-sup.	0–1	0.17 (0.38)	0.22 (0.41)	0.19 (0.40)
AWP-2-error	0–1	0.56 (0.50)	0.34 (0.48)	0.45 (0.50)
AWP-3	0–1	0.41 (0.50)	0.66 (0.48)	0.53 (0.50)
AWP-4	0–1	0.24 (0.43)	0.52 (0.50)	0.38 (0.49)
AWP-4-sup.	0–1	0.10 (0.30)	0.11 (0.31)	0.10 (0.31)
AWP-4-error	0–1	0.66 (0.48)	0.37 (0.49)	0.52 (0.50)
AWP-5	0–1	0.20 (0.40)	0.38 (0.49)	0.29 (0.46)
AWP-5-sup. (any)	0–1	0.09 (0.28)	0.12 (0.33)	0.10 (0.31)
AWP-5-error	0–1	0.71 (0.46)	0.49 (0.50)	0.61 (0.49)
AWP-one-op.	0–2	0.9 (0.73)	1.26 (0.64)	1.07 (0.71)
AWP-two-op.	0–3	0.77 (0.82)	1.57 (1.07)	1.16 (1.03)
AWP-Consistent	0–2	1.04 (0.73)	1.48 (0.66)	1.25 (0.73)
AWP-Inconsistent	0–3	0.61 (0.79)	1.35 (1.19)	0.97 (1.06)
AWP-Superficial	0–3	0.36 (0.61)	0.45 (0.73)	0.40 (0.67)
AWP-Error	0–3	2.03 (1.60)	1.20 (1.08)	1.63 (1.27)
AWP-Global	0–5	1.66 (1.18)	2.83 (1.52)	2.22 (1.47)

Sup.: superficial response. (**any**): all superficial responses (first, second or both operations).

The results showed the expected pattern, depending on the number of operations involved and the consistency or non-consistency of the verbal statement. Likewise, results confirm an increase in the number of correct answers in 3rd grade students compared to that of students in the 2nd grade. The results showed a high standard deviation on each problem, typical of a categorical variable where hit or miss was recorded, and similar to other studies on AWP. A mixed ANOVA for repeated measures was performed in order to check the differences among the problems' difficulty considering, also, the students' grades. The five AWPs were englobed in a factor called "difficulty of problems", while school grade was introduced as an intersubject factor. Results confirmed an overall significant difference in the difficulty of the problems (F (3.6, 135) = 24.12, MSE = 4.39, p < .001, $\eta 2p$ = 0.15), as well as a significant difference in performance regarding the student's grade, easier for 3rd grade than 2nd grade (F (1, 135) = 18.86, MSE = 7.77, p < .001, $\eta 2p$ = 0.24). However, factors did not interact significantly (F (3.6, 135) = 0.44, MSE = 0.08, p = .764). A post-hoc analysis of the problems (Table 6) identified differences in difficulty among all the problems (Bonferroni p < 0.05) except when comparisons were between AWP-2 (one inconsistent operation) and AWP-4 (consistent then inconsistent operations), AWP-2 and AWP-5 (two inconsistent operations), and AWP-4 and AWP-5; that is, there were no significant differences among inconsistent problems in terms of the number of operations.

Comparison between AWP-1 with AWP-2 problems revealed the differences in performance between one consistent operation and one inconsistent operation problems. This result could also be confirmed with the comparison between the problems with two operations. Thus, as expected, a consistent two-operation problem (AWP-3) was less difficult to solve than an inconsistent two-operation problem (AWP-5). Regarding the number of operations, we confirmed also significant differences in difficulty between one operation problems and two operations problems, only when they were consistent operations (AWP-1 with AWP-3).

Table 6. Post-hoc (Bonferroni) for performance comparison in AWP tasks.

(I) AWP	(J) AWP	Mean Differences (I-J)	Dev. Error	Sig.	*Cohen's d*
1	2	0.363	0.052	< .001	0.77
	3	0.184	0.053	.007	0.40
	4	0.339	0.051	< .001	0.72
	5	0.430	0.051	< .001	0.94
2	3	−0.179	0.057	.019	0.35
	4	−0.024	0.044	> .050	
	5	0.066	0.045	> .050	
3	4	0.155	0.052	.032	0.30
	5	0.246	0.051	< .001	0.50
4	5	0.091	0.039	> .050	

Regarding mistakes, we can observe that the proportion of superficial answers in inconsistent problems 2, 4, and 5 is always lower than the proportion of other erroneous responses in each course. Likewise, the decrease in age and school level for the other erroneous responses, but not the superficial ones, is remarkable (F (9, 135) = 2.06, MSE = 0.48 p = .038, $\eta 2p$ = 0.13). Additionally, we looked at the differences among superficial responses in problem 5. An ANOVA for repeated measures revealed non-differences regarding the probability to commit a superficial mistake in the first step/operation of the problem (superficial 1), or in the second step/operation of the problem (superficial 2) or both (F (1.8, 135) = 0.99, MSE = 0.04, p = .366)

Interrelationships among variables

All of the intercorrelations between the four cognitive measures are significant, even if slight (see Table 7). The correlations between the global mean for arithmetic problems and that of the cognitive variables were significant. A clearly increasing pattern of correlations between the cognitive measures and the one- and two-operation problems can be observed, as well as between the consistent and inconsistent problems, respectively. The correlations are higher in the two-operation and inconsistent measures—for example, with KBIT and the two-operation measures, as opposed to one-operation, and with inconsistent problems versus consistent ones. We can find a similar effect with PROLEC-SE and two-operation problems, as opposed to one-operation, and inconsistent versus consistent.

Table 7. Pearson correlations among academic achievement (History/Geography, Mathematics), reading comprehension (PROLEC-SE), reading decoding processes (Spelling-SE), visuospatial reasoning (KBIT), Deductive Reasoning Test simplified (DRTs), and arithmetical word problems (AWPs; one and two-operation problems, consistent and inconsistent problems, and global score).

	1	2	3	4	5	6	7	8	9	10	11
1. History-Geo	1	.61 **	.33 **	.32 **	.22 **	.04	.26 **	.27 **	.14	.35 **	.33 **
2. Mathematics		1	.30 **	.19 *	.39 **	.22 **	.34 **	.33 **	.24 **	.36 **	.38 **
3. PROLEC-SE			1	.26 **	.38 **	.23**	.20 **	.35 **	.24 **	.32 **	.35 **
4. Spelling-SE				1	.17 *	.11	.19 *	.23 **	.20 **	.23 **	.27 **
5. KBIT					1	.20 *	.22 **	.35 **	.19 *	.34 **	.34 **
6. DRTs						1	.21 **	.21 **	.08	.27 **	.24 **
7. AWP-one-op.							1	.42 **	.58 **	.65 **	.76 **
8. AWP-two-op.								1	.61 **	.82 **	.90 **
9. AWP-Consistent									1	.32 **	.73 **
10. AWP-Inconsistent										1	.88 **
11. AWP-Global											1

** $p < .01$; * $p < .05$. PROLEC-SE: reading processes battery test; Spelling-SE: spelling test for secondary education; KBIT: Kaufman Brief Intelligence Test; DRTs: Deductive Reasoning Test simplified; AWP: arithmetic word problem.

Regarding academic performance, Spelling-SE and KBIT showed slight correlational relationships with performance on History/Geography, whereas DRTs did not reach significance with academic performance. Academic performance on History/Geography significantly correlated with the diverse AWP measures, except consistent problems; particularly, it correlated with inconsistent problems and the global mean measure. As for Mathematics, the highest correlations were found with the global measure of arithmetic problems and the fluid intelligence measure of KBIT. Academic achievement in Mathematics and the AWP measures showed a relevant relationship, especially on the global measure and the inconsistent problems. In addition, it was confirmed that Arithmetic problems require both comprehension and reasoning abilities in the same mathematical domain, and KBIT requires an abstract visuospatial reasoning ability also demanded in mathematical learning.

Once the relationship between cognitive measures and academic achievement on the sample had been confirmed, a linear regression analysis was run to check the power of association of the cognitive measures on AWP (see Table 8). Results confirmed that cognitive tasks were able to explain about 20% of the variance of the performance in the AWP-Global in this sample. KBIT, PROLEC-SE, and Spelling-SE, in order of higher to lower explanatory capacity, were the significant variables included in the model. DRTs was not significant.

Similarly, a new regression analysis allowed determining which variables were associated with academic performance. Regression analysis on History/Geography was performed by including the previous cognitive measures with the inconsistent AWPs. The results indicated that the model was able to explain the 20% of the variance in History/Geography, of which AWP-Inconsistent, Spelling-SE, and PROLEC-SE, in order of higher to lower explanatory power, were significant. KBIT and DRTs were not significant. When looking at academic performance in Mathematics, the model explained 21% of the variance in the Mathematics performance for both secondary grades levels; only KBIT and AWP-Global, in this order, were significant.

Table 8. Simple linear regression among AWPs, cognitive measures (Spelling-SE, PROLEC-SE, AWP Global or AWP Inconsistent, DRTs and KBIT), and academic achievement in secondary school (direct regression).

Dependent Variable	R^2	R^2 Adj	*F*	*B*	*Beta*	*t*	*p*
AWP Global	0.22	0.20	9.101 ***				
KBIT				0.07	0.22	2.56	.012
PROLEC-SE				0.07	0.19	2.24	.027
DRTs				0.14	0.13	1.64	.103
Spelling-SE				0.09	0.17	2.10	.038
History & Geography	0.23	0.20	7.774 ***				
KBIT				0.02	0.04	0.42	.673
PROLEC-SE				0.11	0.21	2.34	.021
AWP-Inconsistent				0.27	0.25	2.92	.004
DRTs				−0.16	−0.11	-1.30	.195
Spelling-SE				0.16	0.22	2.68	.008
Mathematics Achievement	0.24	0.21	8.261 ***				
KBIT				0.11	0.25	2.93	.004
PROLEC-SE				0.04	0.09	1.01	.314
AWP Global				0.31	0.23	2.67	.009
DRTs				0.13	0.09	1.15	.254
Spelling-SE				0.04	0.05	0.61	.541

*** $p < 0.001$

3.1.4. Discussion

The goal of this work was to dig more deeply into the relationship between reasoning and reading comprehension processes in AWPs and their relationship to academic performance. The results confirm our first hypothesis: We found a reliable increase in the number of correct responses between 2nd and 3rd grade participants. Participants in the 2nd year of secondary school experienced greater difficulty solving the problems than did those of 3rd, who obtained a number of hits more than double that of younger students across the various problems posed. As expected, adolescents are in the process of acquiring necessary curriculum knowledge in diverse subjects at school while developing basic cognitive and formal thinking abilities, as well as reading comprehension abilities.

As has been shown in other studies (Castro-Martínez & Frías-Zorilla, 2013; Hegarty et al., 1992; Mayer & Hegarty, 1996), the number of calculation operations influences the difficulty in solving AWPs. This is because the increase in the number of steps or operations required in solving the problem involves constructing a text base with a greater number of propositions and a more complex situational model. The process of understanding two-operation problems is therefore quantitatively more complex than those of one consistent operation. This process requires, therefore, a greater number of resources dedicated to updating and integrating any new contents that constitute the situational representation of the problem (Bull & Lee, 2014; Friso-van den Bos et al., 2013; Passolunghi & Pazzaglia, 2005). However, contrary to our predictions, the effect on the difficulty generated by two-operation problems disappears in the presence of inconsistences, reinforcing the idea that the difficulty core falls on the way solvers interpret the key elements inserted on the problems. That is why Mathematics instruction for problem solving must be focused not only on computational procedures (quantitative skills), but also on reasoning abilities (qualitative skills; Mayer et al., 1992).

Our data also confirm the importance of comprehension processes that determine the arithmetic operations that must be applied when resolving AWPs. Thus, the inconsistent problems, in which the linguistic comparative does not coincide with the arithmetic operation necessary for its resolution, are significantly more difficult to solve than problems in which the linguistic expression and the operation coincide. This increase in difficulty is related to the need to inhibit the propositional microstructure referring to the comparative and replacing it with the inverse term. This requires explicit reflection during the process of constructing the situational model of the problem. The greater the number of non-consistencies contained in the problem statement, the greater the number of interpretation errors that can lead to an incorrect arithmetic formalization and, therefore, errors in calculation during the resolution phase. The results of our study are similar to those found in other studies on consistent and inconsistent AWPs (Hegarty et al., 1992, 1995; Lewis & Mayer, 1987; Pape, 2003; Passolunghi et al., 2021; Verschaffel, 1994). Inconsistent problems require a greater number of sentence readings and, therefore, greater dedication in terms of time and resources to be able to integrate the propositions relating the elements of the text, and to inhibit the initial representational models that arise from a superficial reading of the statements. As other studies have pointed out (Bull & Lee, 2014; Friso-van den Bos et al., 2013; Shum & Chan, 2020; St Clair-Thompson & Gathercole, 2006), a deficit in inhibitory control capacity during the process of solving arithmetic problems predisposes a person to adopt inadequate strategies for solving arithmetic problems.

We found that inconsistent one-operation problems were significantly more difficult to solve than the same consistent one-operation problems, and even more than consistent two-operation problems. These results confirm the importance of inhibition and updating processes in reasoning (Evans & Stanovich, 2013; Fletcher & Carruthers, 2012; García-Madruga et al., 2007; Shum & Chan, 2020; Stanovich & Toplak, 2012). Inconsistent problems require, unlike consistent problems, the resolver's ability to suppress one's intuitive, automatic, and superficial response tendency that is guided by the literal representation of utterances, and which leads to the construction of an erroneous situational representation.

The analysis of mistakes allows us to confirm that the most common are not superficial errors, defined as the result of the right performance of a mistaken plan, but other error responses. We also observe that the number of superficial responses is low and shows a stable ratio in both ages. As a matter of fact, correct responses increase in the same proportion that error responses decrease, whereas superficial responses keep constant. Other errors constitute a set of responses formed by calculation errors and other wrong answers due to various factors, such as a low commitment of the participants with the task, and the inability to understand the statement of the problem and difficulties in updating the base of text. It will be very interesting in future studies to check participants' metacognitive assessment of problems difficulty.

Regarding superficial responses in inconsistent problems, they were never higher than error responses, which were always the most frequent answer in the whole sample. Although there is a significant increase in response accuracy between grades, this result is evidence that even in 14-year-olds, there is a poor performance in the skills needed to inhibit superficial responses and those that allow updating of the base of text content as the problem-solving process takes place. Given the type of response recorded by the participants, it is not possible to determine which of the two factors, inhibition or updating, contributed more to the performance produced. However, recent studies would point to a greater effect of updating as a main cause of poor performance for solving these problems (Passolunghi et al., 2021).

With regard to our second hypothesis, we have confirmed the relationship between reasoning (KBIT and DRTs) and reading comprehension (PROLEC-SE and Spelling-SE) measures with arithmetic problems. We can observe, also, that the highest correlations are found in two-operation and inconsistent problems, as opposed to single operation and consistent problems. This confirms our prediction about the relationship between the degree of complexity of a task and the processing requirements necessary to carry it out successfully. The inconsistent problems require reflection and the controlled application of WM executive processes. That is, they require the activation of Type 2 reasoning processes. The processes of reflexive reasoning when solving inconsistent arithmetic problems take on a special importance, and this is revealed in their relationship with measures of deductive (DRTs) and visuospatial (KBIT) reasoning, both clearly superior with inconsistent problems than with consistent ones.

The relationships between cognitive measures, arithmetic problems, and academic performance are especially evident in the results obtained through regression analysis, which allowed us to establish the associative capacity of the various variables (hypothesis 4). Thus, we verified that the reasoning and text comprehension measures explain about 20% of the variance toward the effective resolution of AWPs. The measure with the most explanatory capacity was intelligence with KBIT, followed by the reading comprehension test PROLEC-SE. These data confirm the importance of the cognitive processes of reading comprehension and reasoning underlying the generation of situational models which enables the discovery of correct answers in arithmetic problems.

On the other hand, we found that AWPs show a significant ability to explain academic performance not only in Mathematics, but also in History/Geography. The measures of performance in History/Geography were Spelling-SE, PROLEC-SE, and AWP-Inconsistent; that is, two measures of reading abilities, one of superficial decoding and the other of comprehension, and a measure of deep comprehension and reasoning in arithmetic. These results confirm the basic linguistic nature of this subject and also show the sensitivity and usefulness of AWPs to investigate in areas of knowledge not directly related to Mathematics.

The most significant measure in terms of explaining performance in Mathematics was KBIT, followed by AWP-Global—that is, a measure of abstract reasoning and a measure of comprehension, reasoning, and calculation in Arithmetic. This KBIT beta value is somewhat lower than other studies on the predictive capacity of Mathematics performance in primary school (García-Madruga, et al., 2014), and similar to others in secondary education using RAVEN as an intelligence measure ($\beta = 0.26$; Gómez-Veiga et al., 2018).

A result deserving some comment is the lack of a significant involvement of DRTs in both regression models. In fact, DRTs correlations are low with all variables. The highest DRTs correlation is with AWP-Inconsistent ($r = 0.27$). Its correlation with Mathematics performance is significant but low, and with History/Geography is very low and not significant. A likely related result is the lack of reliable developmental differences between 2nd and 3rd grade students in DRTs: Deductive problems are equally difficult for students in both grades. These results pose some doubts regarding the use of this simplified version of DRT to assess preadolescents' linguistic reasoning.

On the whole, regression analysis results account for the diverse but related nature of the academic subjects, as well as the function of cognitive abilities, different in History/Geography than in Mathematics, required to be successful in both subjects. It is important to emphasize that the AWP measures were the unique ones present significantly in the regressions of both academic achievement measures. This interesting result corroborates the capability and usefulness of arithmetic word problems at school.

A possible limitation of our work arises from the use of teacher evaluations as a measure of academic performance. Being a summary of student academic achievement, it may prove a biased measure when reporting progress made by the students during the school year, thus compromising its validity and reliability (Allen, 2005). However, the numerical scores the educational center provided were individualized data obtained by students in each of the subjects (History/Geography and Mathematics) as a result of the achievements obtained over a prolonged period in time (i.e., an entire school year), and which included other characteristics of the student themselves, such as their motivation regarding learning or their behavior in the classroom. For this reason, we consider it a good measure of academic performance (Roth et al., 2015), something that we can also confirm based on its widespread use in other studies (Gómez-Veiga, et al., 2018; Kuhn & Holling, 2009).

There were some more limitations in this study. One of them stems from the number of problems the test contains. This measure was designed to be a short screening test. However, it would be advisable to consider increasing the number of problems in future studies in order to give greater statistical strength to the measure provided by this test. On the other hand, in problem number 4 (three operations and one inconsistency), we did not take into account whether the performance of the participants could vary due to the fact that the inconsistency was located in the second operation of the three necessaries to solve it, instead of the first as in the present study. Therefore, these limitations should be taken into account not only for future research projects, but also for the educational implications they might have for student learning.

Concerning educational implications, the different stages of the resolution process involve the need for relevant updating and inhibition resources. The results of this study invite us to consider the AWP task as a valid and apparently simple instrument to measure comprehension and reasoning aptitudes of secondary school students as a valid screening tool, particularly in the field of Mathematics. However, for this purpose, it would be necessary to clarify in detail the role played by working memory (WM) and the executive functions that arbitrate this kind of task and introduce in future studies specific measures of updating, inhibition, and WM span. These measures have frequently been related to academic success and its associative capacity in both Mathematics and other subjects (García-Madruga et al., 2013; García-Madruga, et al., 2014; Passolunghi & Mammarella, 2010; Re et al., 2016).

As we have demonstrated, the intelligence measure KBIT provides a relevant but limited explanatory role, and its explanatory capability is similar to the arithmetic problems in Mathematics, and not reliable in History/Geography. This fact holds special relevance, serving as an example to show that, apart from other contextual, motivational, and emotional considerations, the academic success of students depends on cognitive processes that are not reduced to classic fluid intelligence. Our AWP task might be included in the educational world as a particularly useful and complementary tool for measuring general capacity or screening learning difficulties. Its use would detect specific deficits in the cognitive abilities involving comprehension and reasoning.

3.1.5. Conclusions

The resolution of AWPs continues to be an educational milestone that must be acquired throughout a student's academic life. In addition to being an educational mechanism for the training and development of mathematical skills, AWPs stand out as an excellent tool with enormous experimental potential due to the high and diverse number of cognitive skills involved in solving them. This paper contributes to showing some of them. Another aim of this work was to demonstrate how these problems could also be used as predictors of academic performance, not only in closely related subjects such as Mathematics, but also in other apparently distant subjects such as History/Geography. However, it would be interesting to explore other psychoeducational applications as a measure for detecting learning difficulties in order to prevent students' academic failure.

3.2. ESTUDIO 2

Passolunghi, M. C., Duque de Blas, G., Carretti, B., Gómez-Veiga, I., & García-Madruga, J. A. (2021). The role of working memory updating, inhibition, fluid intelligence and reading comprehension in explaining differences between consistent and inconsistent arithmetic word problem solving performance. *Manuscript Submitted for Publication*.

Journal of Experimental Child Psychology

The role of working memory updating, inhibition, fluid intelligence and reading comprehension in explaining differences between consistent and inconsistent arithmetic word problem solving performance

--Manuscript Draft--

Manuscript Number:	
Article Type:	Full Length Article
Keywords:	inhibition; updating; word problem solving; fluid intelligence; working memory
Corresponding Author:	Maria Chiara Passolunghi University of Trieste, Italy Trieste, Italy
First Author:	Maria Chiara Passolunghi
Order of Authors:	Maria Chiara Passolunghi
	Gonzalo Duque De Blas
	Barbara Carretti
	Isabel Gomez-Veiga
	Juan Antonio Garcia-Madruga
Abstract:	Children's performance in arithmetic word problems (AWP) predicts their academic success, and also their future employment and earnings in adulthood. Understanding the nature and difficulties of understanding and solving AWP is important for theoretical, educational and social reasons. We investigated the relation between primary-school children's performance in different types of AWPs and their basic cognitive abilities (reading comprehension, fluid intelligence, inhibition and updating processes). The study involved 182 fourth- and fifth-graders . Participants were administered an AWP-solving task and other tasks assessing fluid intelligence, reading comprehension, inhibition, and updating. The AWP-solving task included comparison problems incorporating the adverbs "more than" or "less than", which demand consistent or inconsistent operations of addition and/or subtraction. The results showed that consistent problems were easier than inconsistent ones. Efficiency in solving inconsistent problems related to inhibition and updating. The order of presentation of the consistent or inconsistent demands affected the children's performance. Moreover, our results seem to indicate that the consistency order effect is related to the efficiency of updating processes. Path analyses showed that reading comprehension was the most important predictor of AWP-solving accuracy. Moreover, both executive functions - updating and inhibition - had a distinct and significant effect on AWP accuracy. Fluid intelligence had both direct and indirect (mediated by reading comprehension) effects on the overall measure of AWP performance. These domain-general factors are important factors in explaining children's performance in consistent and inconsistent arithmetic word problem solution.

The role of working memory updating, inhibition, fluid intelligence and reading comprehension in explaining differences between consistent and inconsistent arithmetic word problem solving performance

Abstract

Children's performance in arithmetic word problems (AWP) predicts their academic success, and also their future employment and earnings in adulthood. Understanding the nature and difficulties of understanding and solving AWP is important for theoretical, educational and social reasons. We investigated the relation between primary-school children's performance in different types of AWPs and their basic cognitive abilities (reading comprehension, fluid intelligence, inhibition and updating processes).

The study involved 182 fourth- and fifth-graders**.** Participants were administered an AWP-solving task and other tasks assessing fluid intelligence, reading comprehension, inhibition, and updating. The AWP-solving task included comparison problems incorporating the adverbs "more than" or "less than", which demand consistent or inconsistent operations of addition and/or subtraction. The results showed that consistent problems were easier than inconsistent ones. Efficiency in solving inconsistent problems is related to inhibition and updating. Moreover, our results seem to indicate that the consistency effect is related to the efficiency of updating processes. Path analyses showed that reading comprehension was the most important predictor of AWP-solving accuracy. Moreover, both executive functions - updating and inhibition - had a distinct and significant effect on AWP accuracy. Fluid intelligence had both direct and indirect (mediated by reading comprehension) effects on the overall measure of AWP performance. These domain-general factors are important factors in explaining children's performance in consistent and inconsistent arithmetic word problem solution.

Keywords: inhibition, updating, word problem solving, fluid intelligence; working memory

3.2.1. Introduction

Mathematical knowledge is a core element of contemporary human cultures and societies and - along with reading and writing - it is one of the basic components of a literate mind. It is therefore unsurprising that arithmetic word problems (AWPs) are so important at school. Children's performance with AWP predicts their academic success (Duque de Blas et al., 2021), and also their future employment and earnings in adulthood (Gross et al., 2009; Murnane et al., 2001). Shedding light on the nature and difficulties of understanding and solving AWP is therefore important for both educational and social reasons.

Investigating how children cope with AWP is also of theoretical relevance. In order to solve an AWP, children have to combine their linguistic and mathematical knowledge in a process characterized by a close relationship between understanding a problem and finding a solution. Solving AWP involves applying some basic arithmetical operations (addition, subtraction, multiplication and division), but also - and more importantly - deciding which kind of operation to apply at a given moment. This demands a correct "problem representation" which relies on a thorough understanding of the information contained in a complex text that includes some numerical quantities (De Corte et al., 1985; Lee et al., 2009).

A large number of studies has investigated the factors underlying the difficulties in solving AWPs (Daroczy et al., 2015; Fuchs et al., 2018; Lin, 2021; Pongsakdi et al., 2020) and the use of key words to identify operations and solve different types of word problems (eg. Jitendra, 2002; Powell, 2011). It has been clearly demonstrated that several types of the mistakes made in solving AWP stem from comprehension issues, as, for example when the problem demands the inhibition of keyword heuristic for solving (Mayer et al., 1996; Shum & Chan, 2020). This is particularly true for the so called compare problems, which are typically more challenging than other simple AWPs (Boonen & Jolles, 2015; Morales et al., 1985; Riley et al., 1984; Riley & Greeno, 1988; Schumacher & Fuchs, 2012).

In the present study we investigate the relation between primary-school children's performance in compare AWPs and basic cognitive variables (reading comprehension, fluid intelligence, inhibition and updating processes). Below is an example of a compare AWP considered in this study:

> "A pair of Adidas sneakers costs 30 Euros at Walmart. At Decathlon the same pair of sneakers costs 6 Euros more. At Decathlon the pair of Adidas sneakers costs 5 Euros less than a pair of Nike sneakers. How much will you have to pay for a pair of Nike sneakers at Decathlon?"

As we can see, this is a multi-step arithmetic word problem since it takes two operations to reach the solution (while a one-step arithmetic word problem takes only one operation to reach the solution). The multi-step arithmetic word problem of our example it takes two additions to find the solution: 30 + 6 + 5 = 41. The first addition is a "consistent" operation because the adverb "more" fits, while the second is "inconsistent" because the adverb "less" does not. This seemingly easy consistent-then-inconsistent (C-I) problem is therefore deceptive. It demands a thorough understanding of a complex instruction in the text, and a complete representation of the problem. According to Kintsch and Greeno (1985), this representation is on two levels: a first propositional level that accounts for the explicit information in the text; and a situational model that involves a complete integration of this information with the reader's previous knowledge. It is only after they have constructed an integrated situational model that readers realize that Nike sneakers at Decathlon are more expensive than Adidas sneakers.

Many authors have also analysed comparison problems that included the terms "more than" or "less than" (as in our example), and involved applying consistent or inconsistent additions and/or subtractions (Jarosz & Jaeger, 2019; Jiang et al., 2020; Mayer & Hegarty, 1996; Pape, 2003). Students can solve problems demanding consistent operations based merely on a "direct translation strategy", relying on a direct translation of the information in the text. They simply search for linguistic markers and keywords and associate "less" with subtraction and "more" with addition (Hegarty et al., 1995). On the other hand, solving AWP that demand inconsistent operations requires a "problem model strategy". Students built an integrated situational mental model of the problem and plan their solution on the basis of this model (Thevenot, 2010; Thevenot & Barrouillet, 2015).

These theoretical approaches have been corroborated by various studies showing that even if the arithmetic operations and linguistic structure of compare problems are very easy, not only young children but also older students and adults make errors the inconsistent compare problems (see Lubin, et al., 2013; Lubin et al., 2016). More importantly, the most common mistakes made in solving inconsistent AWPs are reverse errors, that is, to apply the wrong operation suggested by the adverb in the text (see, Boote & Boote, 2018; Lewis & Mayer, 1987; Mayer et al., 1996; Shum & Chan, 2020; Stern, 1993; Verschaffel et al., 1992; Verschaffel, 1994).

Working memory's executive functions and fluid intelligence

Working memory. Multiple processes are involved in solving an AWP. Problem solvers must read the problem, identify and retain relevant information, construct a mental model, plan how to arrive at the solution, and carry out calculations based on this plan. WM resources are involved in all these phases (Friso-van den Bos et al., 2013; Fung & Swanson, 2017; Kintsch, 1998). A large body of research has investigated the role of WM in problem solving by referring to the well-known tripartite WM model introduced by the seminal work of Baddeley and Hitch (1974). Several studies have shown that one of the components of this model, the central executive system, is needed to solve an AWP (Andersson, 2007; Fuchs et al., 2010; Lee et al., 2004; Swanson, 2006; Swanson & Sachse-Lee, 2001). Swanson (1994) reports that the central executive contributes to accuracy in AWP solving after controlling for phonological processing ability, fluid intelligence, and reading comprehension.

Executive functions. According to Miyake et al. (2000), the central executive component of Baddeley's model is related to three main executive functions: inhibition, updating, and shifting (see also Miyake & Friedman, 2012). Inhibition involves the ability to suppress dominant or prepotent responses; shifting involves the ability to shift strategies when attending to multiple tasks or mental processes; and updating involves the ability to replace outdated and irrelevant information, retaining only a limited set of elements in WM.

Importantly, for the purposes of this paper, several studies have demonstrated the role of these executive processes in AWP solving, and clarified that this process is supported by updating and inhibition. Agostino et al. (2010) investigated the extent to which inhibition, updating, shifting, and mental-attentional capacity (M-capacity) contributed to children's ability to solve multiplication word problems. Using a structural equation model, they demonstrated that updating had a more important role than age in predicting performance for multiple-step problems, while age and updating were equally important predictors for one-step problems.

The above-mentioned study confirmed findings previously reported by Passolunghi and Pazzaglia (2005), who produced evidence of the updating function being involved in the process of solving AWP. They demonstrated that children who fare better in math also have a higher updating performance. Other authors have likewise suggested and proved that updating could be a key cognitive process in solving AWP (Blessing & Ross, 1996; Hammerstein et al., 2019; Iglesias-Sarmiento et al., 2014; Kotsopoulos & Lee, 2012; Lee et al., 2018). Mori and Okamoto (2017) wrote that updating is a nuclear executive function involved in efficient integration processes that, as suggested by Kintsch and Greeno (1985), are needed to translate each verbal statement into a situational representation in order to solve the problem. The authors found that individuals with a high updating function construct a model of the problem that only retains task-relevant verbal information, while those with less effective updating abilities also consider extraneous information. Their results were interpreted taking the view that updating is an underlying executive function that is essential to activating only the information actually needed to construct an appropriate model of a problem. Constructing an integrated situational model is particularly important to understanding a problem and applying the "model" strategy (Mayer & Hegarty, 1996). An updating failure would give rise to errors in the integration process, resulting in an inappropriate model and a consequently incorrect solution to the problem (see Kotsopoulos & Lee, 2012; Re et al., 2016).

Only few studies have investigated the role of shifting in mathematics performance and the outcomes are mixed (Passolunghi & Costa, 2019). Moreover, intelligence showed stronger associations with math performance than shifting ability (Yeniad et al., 2013). However, several studies have shown a relationship between inhibition and mathematical ability (Bull & Scerif, 2001; Khng & Lee, 2009; Passolunghi & Siegel, 2001; Swanson & Fung, 2016) and support the idea that inhibition might be a core process in solving AWP. Its contribution was highlighted by Lubin and colleagues (2013, 2016), who showed that successful inconsistent problem solving relies on the ability to inhibit a misleading or overlearned arithmetical strategy, such as "add if more, subtract if less". This misleading strategy interferes with AWP-solving performance in childhood and adolescence, and even in adulthood (though experts become more efficient at inhibition and problem solving than non-experts). In other words, problem solving demands the inhibition of a superficial propositional representation of the problem resulting from a direct translation approach, which would lead to reasoning errors (Hegarty et al., 1992, 1995; Mayer & Hegarty, 1996). Consistently, Lemaire and Lecacheur (2011) found that children with better inhibitory control made more use of efficient strategies to solve arithmetical problems than children with lower levels of inhibitory control. Further evidence indicated that an increasing efficiency in solving inconsistent AWP from childhood to adulthood (Lewis & Mayer, 1987) is related to the gradual development of cognitive inhibitory control (Lubin et al., 2013, 2016).

Fluid intelligence. The role of fluid intelligence in AWP solving is also considered in the present study. Fluid intelligence embraces processes related to inductive and deductive reasoning, as well as quantitative reasoning (Sternberg & Ben-Zeev, 1996). Previous findings showed that fluid intelligence correlates strongly with measures of executive functions (Conway et al., 2003; Duncan et al., 2008; Kane & Engle, 2002; Salthouse, 2005). To be more specific, numerous studies found moderate-to-strong relations between intelligence and updating function (Friedman et al., 2006). Some studies produced empirical support for the importance of intelligence as a predictor of AWP solving. For example, Lee et al. (2004) reported that 10-year-old children scoring higher on intelligence, reading skills and vocabulary were also better at solving AWP. Fung & Swanson (2017) found an indirect effect of fluid intelligence on AWP solving in third-graders, and reading and calculation also mediated the influence of the WM executive system on AWP solving in children aged 6-10 years. A meta-analysis conducted by Peng and colleges (2019) revealed that the relations between fluid intelligence and mathematics tend to increase with age, and that fluid intelligence showed stronger relations to complex mathematical skills than to initial and basic mathematical skills.

The present study

Our main objective was to study the role of executive processes (particularly updating and inhibition), comprehension and fluid intelligence in primary-school children solving consistent and inconsistent AWP. To achieve this aim, participants were administered AWP tasks in which the wording of the problem did or did not match the operation needed to solve the problem. In other words, in consistent problems involving one operation, the adverbs “more” or “less” used in describing the problem fitted the arithmetical operation required, whereas in inconsistent problems they did not. We also included problems involving two operations, with a consistent request followed by an inconsistent one (C-I) in some, in vice versa (I-C) in the others.

The novelty and the main aim of this study was to examine the specific influence of updating, inhibition, and fluid intelligence on AWP-solving accuracy in the same unitary model, as well as the mediating role of reading comprehension, taking into account the different types of consistent and inconsistent compare arithmetic word problems.

Our hypotheses can be outlined as follows.

1. Based on the literature (Hegarty et al., 1995; Lewis & Mayer, 1987; Pape, 2003), we expected to find a *consistency effect* that gives rise to a higher percentage of correct answers for consistent problems (Hypothesis 1).
2. The accuracy with which participants solved consistent vs inconsistent problems might relate not to the number of arithmetical operations involved *per se,* but to the semantic representation of the AWP (Hypothesis 2).
3. The semantic representation of AWP requires reflection and the controlled application of processes, such as inhibition and updating of the information contained in the AWP (Hypothesis 3, see Passolunghi & Pazzaglia, 2005; Iglesias-Sarmiento et al., 2015; Thevenot & Barrouillllet, 2015).
4. We also aimed to examine the specific influence of updating, inhibition, and fluid intelligence on AWP-solving accuracy in the same unitary model, as well as the mediating role of reading comprehension, taking into account the type of problem (Hypothesis 4).
5. Finally, we hypothesized that C-I AWP would be more difficult (*consistency order effect*) because they demand a greater updating effort, since the second, inconsistent (harder) operation has to be solved when the WM is already loaded with the result of the first, consistent operation (Hypothesis 5).

The last three hypotheses represent the main novelty of the present study.

3.2.2. Method

3.2.2.1. Participants

The study involved 203 students between 10 and 11 years of age (mean age: 10 years and 6 months), attending the 4th and 5th grades at a primary school in a city in north-east Italy. From this initial sample, students with intellectual disabilities (5 children) or learning disabilities (8 children), and those with Italian as their second language (8 children) were excluded, leaving a final sample of 182 participants (95 females). Arithmetic abilities, as measures by standardized test, was on average level for grade.

3.2.2.2. Tasks and measures

Arithmetic Word Problem (AWP) task.

In this task, 12 different hypothetical problem situations were presented in two twin versions (see Table 1). The task included: four "consistent" problems, in which the adverbs "more" or "less" suggested an operation of "addition" or "subtraction", respectively; and eight "inconsistent" problems, in which the inverse arithmetical operation to the one suggested in the description of the problem was needed to solve it. Another variable that can be used to increase the difficulty of the problems is the number of relationships described between the elements of the problem: more relationships mean that more operations are needed to solve the problem. Four of the AWP involved one operation and eight required two. All other components of the word problems (number of sentences, vocabulary and syntax) are consistent and uniform. AWP problems were similar to those experienced by students of the considered school grade. However, very often the teachers do not emphasize the possibility of making mistakes in the inconsistent compare problems.

Table 1. Different types of AWPs tested in the present study.

#	Relational term	Type	Number of operations (WM load)
1	Plus	Consistent	1
2		Inconsistent	1
3		Consistent	2
4		Inconsistent-Consistent	2
5		Consistent- Inconsistent	2
6		Inconsistent	2
1	Minus	Consistent	1
2		Inconsistent	1
3		Consistent	2
4		Inconsistent-Consistent	2
5		Consistent- Inconsistent	2
6		Inconsistent	2

The AWP were equally divided into two groups (called "plus" and "minus" versions) depending on the type of adverb ("more" or "less") used in describing them. Each version of the AWP was presented at two sessions on different days to avoid tiring participants. Half of the participants were administered the "plus" version at the first session and the "minus" version at the second, while the versions were presented in reverse order for the other half. Problems varied in number of operations (1 or 2) and the type of relationship (consistent or inconsistent) between the elements. Cronbach's α for this task in the current sample was .77.

Reading comprehension.

This was measured using two expository texts (appropriate for fourth and fifth graders) drawn from the Italian standardized battery for the assessment of reading ability (Cornoldi et al., 2017). Participants were asked to read the text silently and then answer 12 multiple-choice questions (and the text remained available while they did so). The reliability (Cronbach's alpha) is .69 and .71, for grades 4 and 5, respectively.

Fluid intelligence.

The Scale 2, form A, of the Cattell Culture Fair Intelligence Test (Cattell & Cattell, 1963) was used. This consists of 4 reasoning subtests involving visuospatial material, each to be completed in 2.5-4 minutes. In the subtest named “Matrices”, for example, participants are shown 12 incomplete matrices of 4 to 9 cells containing abstract figures and shapes, plus an empty cell and 6 options from which to choose the one needed to complete each matrix correctly. The final score is the sum of the correct answers (max = 36). The test-retest reliability coefficient reported in the manual is .84 and .80 for fourth- and fifth-graders, respectively.

Updating.

This task, taken from Carretti et al. (2014), consisted of six lists of eight nouns each. After listening to each list, participants were asked to remember the three smallest items on each list, in the correct order of presentation. All the words were very familiar to the children and referred to objects that were easy to compare in terms of size. The number of correctly recalled words was the dependent variable (maximum score 18). Cronbach’s α for this task is .68.

Prepotent response inhibition.

This was measured with a version of the classic *Color Stroop task,* as used in previous research (Borella et al., 2010). The task was administered in a pencil-and-paper version, and consisted of 15 trials. Each trial included 20 stimuli printed on a separate sheet. In the first five trials, participants had to name the colors of strings of Xs printed in capital letters (control-color condition). Then there were five trials in which participants were asked to name the color of the ink used for words that were names of colors, where the ink and name colors were in conflict. In this condition, to perform the task efficiently, participants needed to inhibit the prepotent response prompted by reading the words in favour of the appropriate but non-dominant response prompted by focusing on the color of the ink (incongruent condition). In the last five trials, participants had to name the color of the ink used for words that were names of colors, but in this case the ink and name colors were consistent (congruent condition). For each trial, completion times and accuracy were recorded. An interference index was calculated for the response times from the differences between the control-color and the incongruent conditions: higher scores indicated greater difficulty in controlling the prepotent responses in the incongruent condition. The task showed a good reliability (control-color: $\alpha = 0.82$; incongruent: $\alpha = 0.73$; congruent: $\alpha = 0.85$).

3.2.2.3. Procedure

The tasks, except the updating task, were administered collectively at two sessions lasting about an hour each. During the first session, participants completed the AWP-solving task in the plus or minus version (balanced between participants) and the fluid intelligence test. In the second session, one or two days later, they were administered the other version of the AWP-solving task, the reading comprehension test, and the prepotent response inhibition task. The updating task was administered individually, and took about 10 minutes.

3.2.3. Results

Table 2 shows descriptive statistics for the measures included in the study. There was no difference in accuracy between the two versions (plus and minus) of the AWP, $F(1, 181) = 3.506$, $p = .063$. The percentage of correct solutions to the AWP did not differ between males and females, $F(1, 181) < .01$, $p = .958$.

Table 2. Descriptive statistics for measures included in the study.

	Mean	SD
AWP CON	3.30	.89
AWP INC	2.58	2.38
AWP I-C	.73	.80
AWP C-I	.51	.70
AWP total	5.87	2.79
Updating	10.47	2.39
Inhibition (time difference)	-84.07	23.06
Fluid intelligence	31.02	4.80
Reading comprehension	8.98	2.60

The effect of consistency on problem-solving performance

The effect of consistency between the linguistic terms used in the problems was examined. As predicted, differences emerged from the repeated-measures ANOVA between problems that were worded consistently (AWP CON: $M = 3.30$, $SD = 0.89$) and those that were not (AWP INC: $M = 1.34$; $SD = 1.25$, $F\ (1, 181) = 415.08$, $MSE = 348.18$, $p < .01$; $\eta_p^2 = .69$.)

When problems with both consistent and inconsistent requests were analyzed, the predicted effect of their order emerged ($F\ (1, 181) = 12.97$, $MSE = 4.18$, $p < .01$; $\eta_p^2 = .07$); performance was significantly worse in the case of AWP C-I ($M = 0.51$; $SD = 0.70$) than in the case of AWP I-C ($M = 0.73$; $SD = 0.80$). Nevertheless, this result has to be taken very cautiously, since the AWP C-I were presented always before the AWP I-C. However, what was more interesting was to verify whether the differences in performance in responding to AWP C-I and AWP I-C were due to a working memory overload. To analyse this issue, we ran a one-way repeated-measures ANOVA, but used the updating measure as a covariate. After controlling for participants' updating ability, the difference in the difficulty of the two types of problem disappeared ($F\ (1, 181) = 1.10$, $MSE = .36$, $p = .295$). This result gives additional support to the assumption that the efficiency of updating processes is strongly related to the ability to solve AWPs and to build a correct mental representation of the problem.

Correlations

Table 3 shows the correlations between the variables considered in the study. All measures were correlated with each other, but in the range from low to moderate.

Table 3. Correlations between measures of interest.

	1	2	3	4	5	6	7	8	9
1. AWP I-C	1								
2. AWP C-I	.44**	1							
3. AWP CON	.26**	.26**	1						
4. AWP INC	.84**	.75**	.31**	1					
5. AWP total	.80**	.72**	.59**	.95**	1				
6. Updating	.20**	.27**	.18**	.27**	.28**	1			
7. Inhibition	.20**	.13*	.18**	.20**	.23**	.15*	1		
8. Fluid intelligence	.25**	.23**	.31**	.32**	.37**	.26**	.10	1	
9. Reading comprehension	.27**	.30**	.30**	.36**	.40**	.18**	.02	.31**	1

** $p < .01$; * $p < .05$. one-tail

Predictors of AWP-solving ability

We computed a series of path analyses as regression equations to clarify the relationships between the variables and the magnitude of their ability to explain performance in AWP with one (consistent or inconsistent) operation or two (one consistent and the next inconsistent (C-I) or one inconsistent and the next consistent (I-C). These analyses included three basic cognitive measures (fluid intelligence, inhibition and updating), one mediated measure (reading comprehension), and the outcome variables that were analyzed separately: Total AWP, consistent AWP (AWP CON), inconsistent AWP (AWP INC), AWP C-I and AWP I-C.

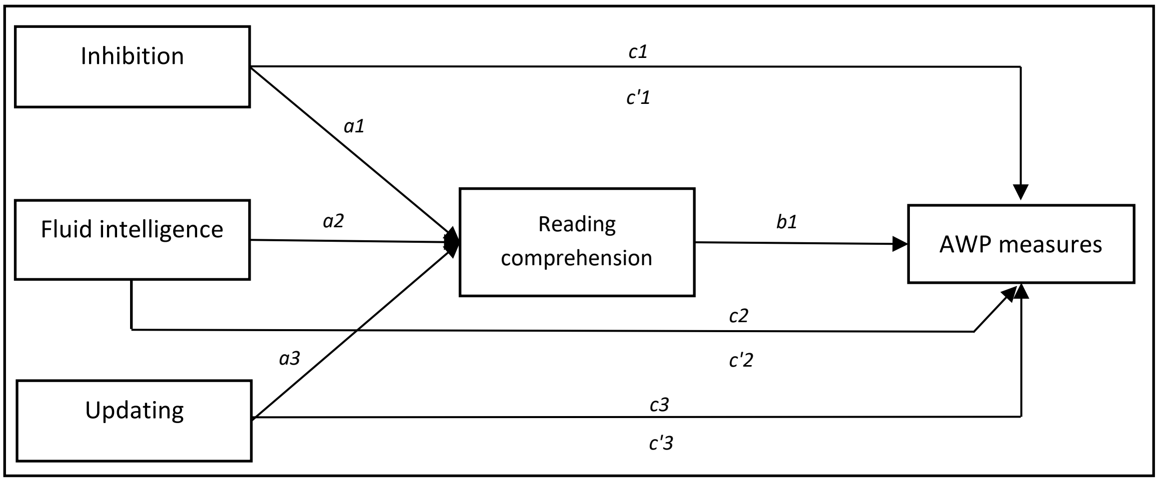
General equations for different outcomes were:

Reading Compreh. $= \beta 1 + \alpha 1$ *(Inhibition)* $+ e1 + \alpha 2$ *(Fluid intelligence)* $+ e2 + \alpha 3$ *(Updating)* $+ e3$

AWP Outcome $= \beta 2 + b1$ *(Reading)* $+ c'1$ *(Inhibition)* $+ c'2$ *(Fluid intelligence)* $+ c'3$ *(Updating)* $+ e4$

Every path model was computed using the same procedure, but changing the AWP outcome predicted (total AWP, AWP CON, AWP INC, AWP C-I or AWP I-C). Figure 1 depicts the path model tested.

Figure 1. Path model tested in the study.



The first regression equation estimated the effect that fluid intelligence, inhibition and WM updating had on reading comprehension, i.e. the mediator. The second regression equation estimated the direct effect (c’) that fluid intelligence, inhibition, updating and the mediator (reading comprehension) had on the outcome variables (i.e., AWP-performance). The indirect effect of inhibition (a1xb1), fluid intelligence (a2xb1) and updating (a3xb1) through reading comprehension tested the mediation effect of these variables on AWP-outcome measures. The direct effect (c’) of the cognitive measures (inhibition, fluid intelligence or updating) on AWP-outcome measures was independent of the indirect effects. Consequently, direct effect was equivalent to subtract the indirect effect from the total effect of cognitive variables on AWP-outcome measures. Total effect (c) represents instead the overall effects of direct and indirect relations on AWP-outcome measures.

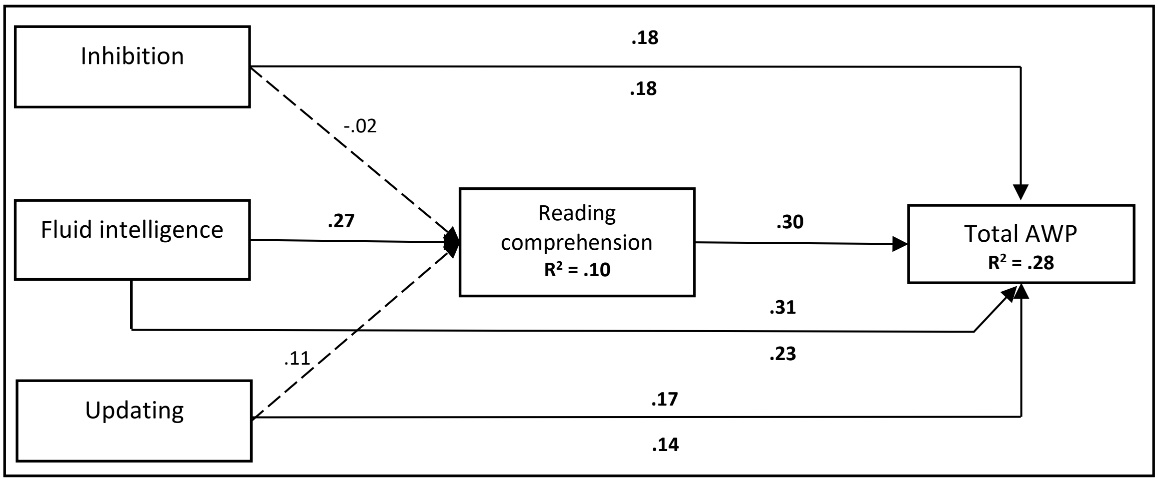
Regarding the total AWP-solving performance (total AWP), results are presented in figure 2 (and table 4). The total amount of variance explained by the whole model was R^2 = .28. The effects of executive functions and fluid intelligence on mediator (paths a) were only significant for fluid intelligence (β = .27). This result was always common in all models of AWP-outcome.

Table 4. Direct, indirect and total effects of inhibition, fluid intelligence, updating and reading comprehension on total AWP-solving performance.

AWP Total	Estimate	Std error	z-value	P(>|z|)	Std.all
Reading comprehension (b1)	0.328	0.072	4.588	.000	0.304
Inhibition (c'1)	0.022	0.008	2.878	.004	0.184
Fluid intelligence (c'2)	0.143	0.042	3.410	.001	0.230
Updating (c'3)	0.170	0.081	2.103	.036	0.140
Reading comprehension					
Inhibition (a1)	-0.002	0.008	-0.255	.799	-0.018
Fluid intelligence (a2)	0.155	0.042	3.715	.000	0.271
Updating (a3)	0.121	0.083	1.449	.147	0.107
Indirect effects					
Inhibition → Reading → AWP Total (a1xb1)	-0.001	0.003	-0.254	.799	-0.006
Intellig. → Reading → AWP Total (a2xb1)	0.051	0.018	2.887	.004	0.082
Updating → Reading → AWP Total (a3xb1)	0.040	0.029	1.382	.167	0.033
Total effects					
Inhibition → AWP Total (c1)	0.022	0.008	2.643	.008	0.178
Intelligence → AWP Total (c2)	0.194	0.043	4.545	.000	0.313
Updating → AWP Total (c3)	0.210	0.085	2.469	.014	0.172

The effect of the mediator, reading comprehension, on Total AWP (path b) was significant (β = .30). The total effects of executive functions on Total AWP (paths c) were significant for inhibition (β = .18), fluid intelligence (β = .31) and updating (β = .17). When direct effects on Total AWP (paths c') were considered, significant relations were found for inhibition (β = .18), fluid intelligence (β = .23) and updating (β = .14).

Figure 2. Path model for total AWP solving performance. Lines in bold refer to statistically significant relations, dashed lines to non-significant relations.

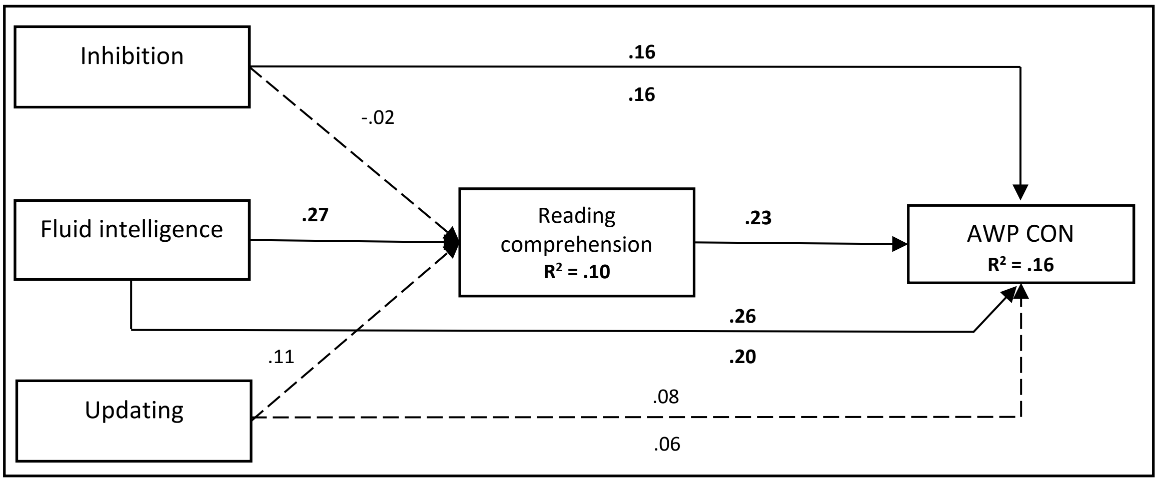


Regarding the Consistent AWP-solving performance (AWP-CON), results are presented in figure 3 (and table 5). The total amount of variance explained by the whole model was $R^2 = .16$. The effect of the mediator, reading comprehension, on AWP-CON (path b) was significant (β = .23). The total effects of executive functions and fluid intelligence on AWP-CON (paths c) were significant for inhibition (β = .16) and fluid intelligence (β = .26), but not for updating (β = .09, p = .24). When direct effects on Total AWP-CON (paths c') were considered, significant relations were found for fluid intelligence (β = .23), but not for inhibition (β = -.02; p = .80) nor updating (β = .11, p = .15).

Table 5. Direct, indirect and total effects of inhibition, fluid intelligence, updating and reading comprehension on consistent AWP-solving performance.

Consistent AWP	Estimate	Std error	z-value	P(>|z|)	Std.all
Reading comprehension (b1)	0.080	0.025	3.249	.001	0.232
Inhibition (c'1)	0.006	0.003	2.345	.019	0.161
Fluid intelligence (c'2)	0.039	0.014	2.722	.006	0.198
Updating (c'3)	0.024	0.028	0.845	.398	0.060
Reading comprehension					
Inhibition (a1)	-0.002	0.008	-0.255	.799	-0.018
Fluid intelligence (a2)	0.155	0.042	3.715	.000	0.271
Updating (a3)	0.121	0.083	1.449	.147	0.107
Indirect effects					
Inhibition → Reading → AWP-CON (a1xb1)	-0.000	0.001	-0.254	.800	-0.004
Intellig. → Reading → AWP-CON (a2xb1)	0.012	0.005	2.446	.014	0.063
Updating → Reading → AWP-CON (a3xb1)	0.010	0.007	1.324	.186	0.025
Total effects					
Inhibition → AWP-CON (c1)	0.006	0.003	2.220	.026	0.157
Intelligence → AWP-CON (c2)	0.052	0.014	3.615	.000	0.261
Updating → AWP-CON (c3)	0.033	0.028	1.166	.244	0.085

Figure 3. Path model for consistent AWP solving performance. Lines in bold refer to statistically significant relations, dashed lines to non-significant relations.

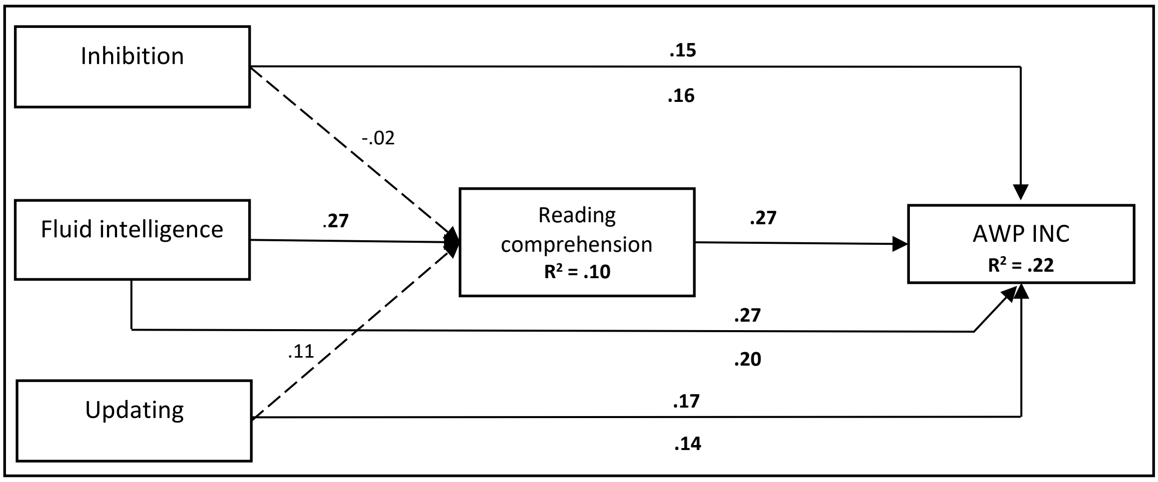


Regarding the Inconsistent AWP-solving performance (AWP-INC), results are presented in figure 4 (and table 6). The total amount of variance explained by the whole model was R^2 = .22. The effect of the mediator, reading comprehension, on AWP-INC (path b) was significant (β = .27). The total effects of executive functions on AWP INC (paths c) were significant for inhibition (β = .15), fluid intelligence (β = .27) and updating (β = .17). When direct effects on AWP-INC (paths c') were considered, significant relations were found for inhibition (β = .16), fluid intelligence (β = .20) and updating (β = .14).

Table 6. Direct, indirect and total effects of inhibition, fluid intelligence, updating and reading comprehension on inconsistent AWP-solving performance.

Inconsistent AWP (all)	Estimate	Std. error	z-value	P(>|z|)	Std.all
Reading comprehension (b1)	0.250	0.063	3.951	.000	0.272
Inhibition (c'1)	0.016	0.007	2.389	.017	0.158
Fluid intelligence (c'2)	0.105	0.037	2.825	.005	0.198
Updating (c'3)	0.145	0.072	2.016	.044	0.139
Reading comprehension					
Inhibition (a1)	-0.002	0.008	-0.255	.799	-0.018
Fluid intelligence (a2)	0.155	0.042	3.715	.000	0.271
Updating (a3)	0.121	0.083	1.449	.147	0.107
Indirect effects					
Inhibition → Reading → AWP-INC (a1xb1)	-0.001	0.002	-0.254	.799	-0.005
Intellig. → Reading → AWP-INC (a2xb1)	0.039	0.014	2.707	.007	0.074
Updating → Reading → AWP-INC (a3xb1)	0.030	0.022	1.361	.174	0.029
Total effects					
Inhibition → AWP-INC (a1xb1+c1)	0.016	0.007	2.221	.026	0.153
Intelligence → AWP-INC (a2xb1+c2)	0.144	0.037	3.856	.000	0.272
Updating → AWP-INC (a3xb1+c3)	0.175	0.074	2.353	.019	0.168

Figure 4. Path model for inconsistent AWP-solving performance. Lines in bold refer to statistically significant relations, dashed lines to non-significant relations.

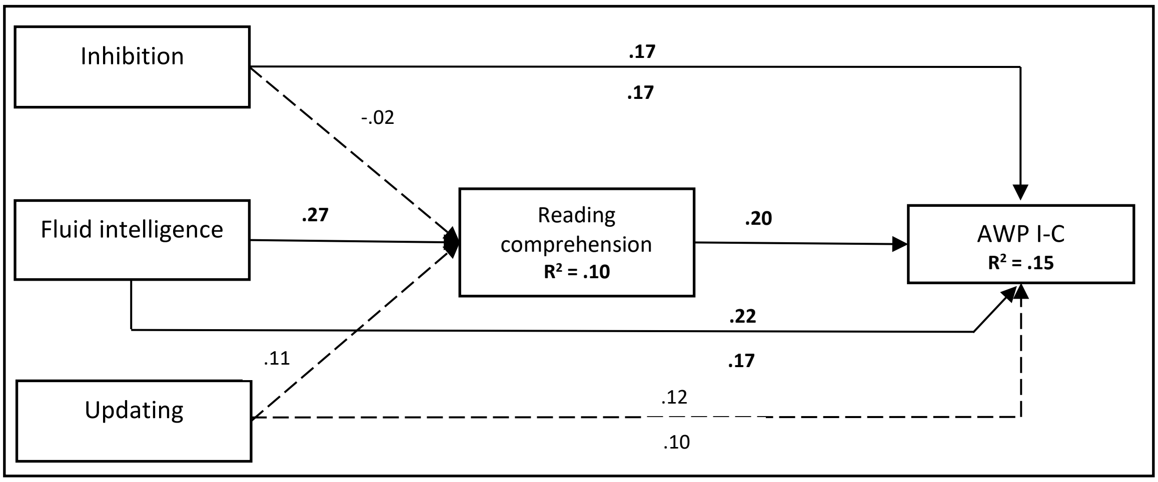


Regarding the two-operation, inconsistent-then-consistent AWP (AWP I-C), results are presented in figure 5 (and in table 7). The total amount of variance explained by the whole model was R^2 = .15. The effect of the mediator, reading comprehension, on AWP I-C (path b) was significant (*β* = .20). The total effects of executive functions on AWP I-C (paths c) were significant for inhibition (*β* = .17) and fluid intelligence (*β* = .22), but not for updating (*β* = .12, *p* = .09). When direct effects on Total AWP-I-C (paths c') were considered, significant relations were found for inhibition (*β* = .17) and for fluid intelligence (*β* = .17), but not for updating (*β* = .11, *p* = .15).

Table 7. Direct, indirect and total effects of inhibition, fluid intelligence, updating and reading comprehension on two-operation, inconsistent-then-consistent AWP-solving performance.

AWP I-C	Estimate	Std. error	z-value	P(>|z|)	Std.all
Reading comprehension (b1)	0.061	0.022	2.727	.006	0.197
Inhibition (c'1)	0.006	0.002	2.496	.013	0.173
Fluid intelligence (c'2)	0.030	0.013	2.303	.021	0.169
Updating (c'3)	0.035	0.025	1.384	.166	0.100
Reading comprehension					
Inhibition (a1)	-0.002	0.008	-0.255	.799	-0.018
Fluid intelligence (a2)	0.155	0.042	3.715	.000	0.271
Updating (a3)	0.121	0.083	1.449	.147	0.107
Indirect effects					
Inhibition → Reading → AWP I-C (a1xb1)	-0.000	0.000	-0.254	.800	-0.004
Intellig. → Reading → AWP I-C (a2xb1)	0.009	0.004	2.199	.028	0.053
Updating → Reading → AWP I-C (a3xb1)	0.007	0.006	1.280	.201	0.021
Total effects					
Inhibition → AWP I-C (a1xb1+c1)	0.006	0.002	2.396	.017	0.170
Intelligence → AWP I-C (a2xb1+c2)	0.040	0.013	3.078	.002	0.223
Updating → AWP I-C (a3xb1+c3)	0.042	0.026	1.651	.099	0.121

Figure 5. Path model for inconsistent-then-consistent AWP-solving performance. Lines in bold refer to statistically significant direct relations, dashed lines to non-significant direct relations.

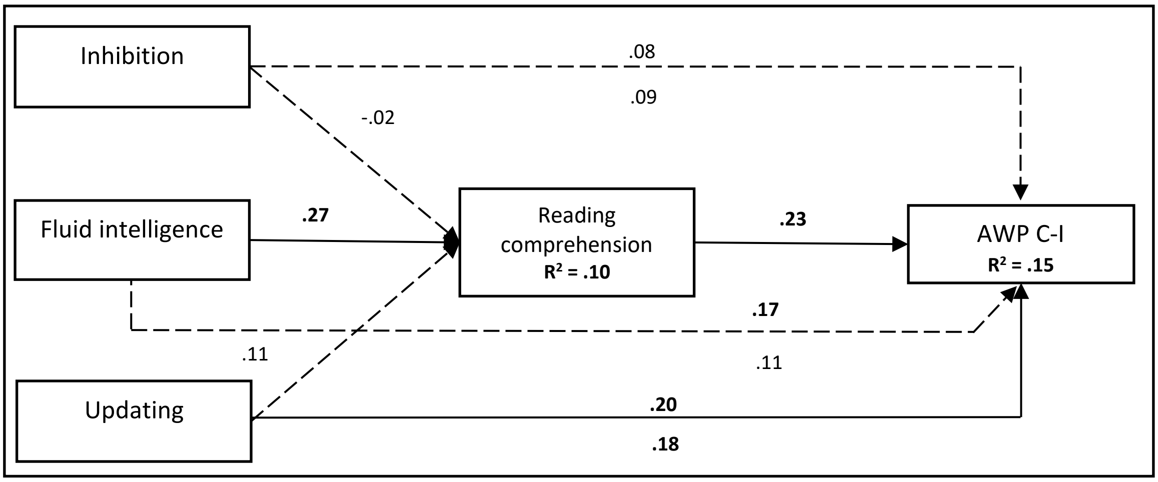


Finally, the two-operation, consistent-then-inconsistent AWP (AWP C-I), results are presented in figure 6 (and table 8). The total amount of variance explained by the whole model was R^2 = .15. The effect of the mediator, reading comprehension, on AWP I-C (path b) was significant (β = .23). The total effects of executive functions on AWP I-C (paths c) were significant for fluid intelligence (β = .17) and updating (β = .20), but not for inhibition (β = .08, p = .26). When direct effects on Total AWP-CI (paths c') were considered, significant relations were found for updating (β = .18), but not for inhibition (β = .09, p = .22), nor fluid intelligence (β = .11, p = .14).

Table 8. Direct, indirect and total effects of inhibition, fluid intelligence, updating and reading comprehension on two-operation, consistent-then-inconsistent AWP-solving performance.

AWP C-I	Estimate	Std. error	z-value	P(>|z|)	Std.all
Reading comprehension (b1)	0.063	0.020	3.236	.001	0.233
Direct effects					
Inhibition (c'1)	0.003	0.002	1.223	.221	0.085
Fluid intelligence (c'2)	0.017	0.011	1.477	.140	0.109
Updating (c'3)	0.054	0.022	2.452	.014	0.177
Reading comprehension					
Inhibition (a1)	-0.002	0.008	-0.255	.799	-0.018
Fluid intelligence (a2)	0.155	0.042	3.715	.000	0.271
Updating (a3)	0.121	0.083	1.449	.147	0.107
Indirect effects					
Inhibition → Reading → AWP C-I (a1xb1)	-0.000	0.001	-0.254	.800	-0.004
Intellig. → Reading → AWP C-I (a2xb1)	0.010	0.004	2.440	.015	0.063
Updating → Reading → AWP C-I (a3xb1)	0.008	0.006	1.323	.186	0.025
Total effects					
Inhibition → AWP C-I	0.002	0.002	1.130	.259	0.081
Intelligence → AWP C-I	0.027	0.011	2.357	.018	0.172
Updating → AWP C-I	0.062	0.023	2.736	.006	0.202

Figure 6. Path model for consistent-then-inconsistent AWP-solving performance. Lines in bold refer to statistically significant direct relations, dashed lines to non-significant direct relations.



3.2.4. *Discussion*

The present study sheds new light on the relationship between consistent and inconsistent AWP-solving performance and several cognitive variables (i.e., inhibition, updating and fluid intelligence), showing the mediating role of reading comprehension. First, as expected and in line with the literature (Hegarty et al., 1995; Lewis & Mayer, 1987; Pape, 2003), the results confirmed a *consistency effect* on children's AWP-solving performance (Hypothesis 1). Participants were more efficient at solving consistent than inconsistent AWPs. It should be noted that the sample considered in the present study: (a) obtained a very high percentage of correct answers for consistent AWPs; and (b) two-operation consistent AWPs were easier than either AWP I-C or AWP C-I, although the number of operations was a factor contributing to the problems' difficulty (Castro-Martínez & Frías-Zorilla, 2013; Quintero, 1983). The pattern of correlations identified supports the crucial role of comprehension-related abilities (Kintsch, 1998; Peng et al., 2019; Thevenot, 2010), which emerged as a stronger correlate of performance for inconsistent than for consistent AWPs. These results suggest that the differences observed in how accurately participants solved consistent vs inconsistent AWPs might not relate to their arithmetical skills *per se* (Hypothesis 2). Instead, it would be the semantic representation of inconsistent AWPs that would influence the difficulty of an AWP, linked to the construction of an integrated mental model of the problem (see Fuson, 1992; Hasanah et al., 2017; Thevenot & Barrouillet, 2015). We know that semantic elements of AWP influence their difficulty, and the strategies children use to solve them (e.g., Boonen et al., 2016; De Corte et al., 1985; Pape, 2003). As mentioned in the Introduction, a key aspect that differentiates inconsistent from consistent AWPs concerns the conflict between the relational term (e.g., "more than") and the arithmetical operation required (e.g., a subtraction) to solve an inconsistent problem. Instead of using a superficial, direct strategy or well-established association (e.g., "more" with "addition and increases in amount"; Schumacher & Fuchs, 2012), the child needs to use a problem-model strategy. This involves translating the statement of the inconsistent problem into a situational model representing the appropriate relations between the variables (de Koning et al., 2017; Hegarty et al., 1995; Thevenot, 2010), then planning and completing the required arithmetical operations.

Such non-routine thinking to solve inconsistent AWPs requires reflection and a controlled application of executive processes, such as inhibition and updating the information about the problem in WM (Hypothesis 3). Failure to inhibit the strong association between the key word in the literal statement (e.g., more) and the operation (e.g., "add") or to update the operation, switching from the wrong to the right one (e.g., "subtract") would tend to prompt a superficial, erroneous response. For an adequate situational representation of an inconsistent step, the problem solver needs to reason reflexively and inhibit any automatic inappropriate response (e.g., Jiang et al., 2020; Passolunghi & Siegel, 2004), or misleading overlearned strategy (Lubin et al., 2013, 2016).

A second main finding concerns the *consistency order effect* (Hypothesis 5) on children's AWP-solving performance. We focused specifically on whether success rates would differ for two-operation problems that differed in the order of presentation of a consistent and an inconsistent step embedded in the statements (i.e., the underlying problem structure could be consistent then inconsistent or inconsistent then consistent). As expected, our results indicated that AWP C-I were more difficult than AWP I-C. The correlational results showed that the relationship between updating and AWP C-I was stronger than it was between updating and AWP I-C. After controlling for participants updating abilities, moreover, the differences in accuracy between AWP C-I and I-C disappeared. However, this result has to be taken very cautiously, since the AWP C-I were presented always before the AWP I-C. Taken together, these results seem to indicate that the predicted consistency order effect is related to the efficiency of updating processes (Passolunghi & Pazzaglia, 2005; Iglesias-Sarmiento et al., 2015; Thevenot & Barrouillllet, 2015).

As concerns the specific contribution of updating, inhibition, reading and fluid intelligence to AWP solving, their role depended on the type of problem (Hypothesis 4). The model confirmed that both executive functions - updating and inhibition - had a distinct and significant effect on AWP-solving accuracy. We also found that fluid intelligence had both direct and indirect (mediated by reading comprehension) effects on the overall measure of AWP-solving performance. It contributed a unique part of the variance in AWP-solving accuracy, even after accounting for the effect of reading comprehension. The magnitude of the beta weights suggests that fluid intelligence had only a moderate effect, however, and that updating and inhibition had even lower effects on AWP-solving performance. Our results likewise confirmed that reading comprehension is the most important predictor of AWP-solving ability in the final grades of primary school (10- and 11-year-olds), by which time basic reading skills have become relatively automated. Reading comprehension had a mediating role in explaining AWP-solving accuracy. The model explained a moderate proportion of variance (30%) in total AWP accuracy, though its weight varied for different types of AWP, being lowest for AWP I-C, and highest for total AWPs.

To sum up, the findings presented in this paper provide new insight on the relationships between fluid cognition, executive functions and AWP-solving performance, in particular in the case of consistent and inconsistent compare problems. We showed that AWP solving relies not only on the acquisition of arithmetical skills, but also on reading comprehension and, importantly, on reasoning and the central executive processes involved in actively processing and updating relevant information, while inhibiting irrelevant information and inappropriate strategies. These domain-general factors are important, each in their own right, in explaining children's performance in consistent and inconsistent AWP solution.

The results of the present study should be interpreted in the light of some limitations. A first possible limitation arises from the fact that a measure of shifting is lacking. Some authors (e.g., Toll et al., 2011) have suggested, however, that the role of shifting in mathematics would relate to the complexity of the task in hand and the knowledge required. This being the case, then shifting might be minimal in such AWPs as those analysed in the present study. A second limitation concerns other factors not studied here that might influence and explain the variation in AWP-solving performance (see, e.g., Fung & Swanson, 2017). For instance, individual differences and other domain-specific factors (such as processing speed, attention ability, visuospatial ability) might contribute to performance in AWP solving (see, Boonen et al., 2013; Cragg et al., 2017; Passolunghi & Lanfranchi, 2012). It may be that the relations described in this paper would change if other variables were examined, and future research should aim to refine the proposed model.

3.2.5. Conclusions

Be that as it may, the findings of this study could contribute to highlight the importance to teach the students to integrate the textual information of the problem into an adequate mental representation, which is the basis for a solution strategy (Jiménez & Verschaffel, 2014; Thevenot

& Oakhill, 2005; Verschaffel et al., 2020). Indeed, solving an AWP is not a simple translation of problem sentences into arithmetic operations, but other factors like solution strategies and building up a correct mental model of the problem could also play an important role in the AWP performance. Educational intervention focusing on comprehension skills, controlled thinking, and working memory processes might be an effective way to improve AWP-solving skills (e.g., Boonen et al., 2016, Cornoldi et al., 2015; Fuchs et al., 2020). Understanding the role of these cognitive and linguistic factors is essential to promote tailored training program of word problem solving abilities.

3.3. ESTUDIO 3

Reflection and intuition in arithmetic word problems: cognitive and executive processes in secondary school students.

Arithmetic problem solving is an educational competence whose acquisition is not free from difficulties. The aim of this article was to compare secondary students' performance in two types of mathematical problems, classical arithmetic word problems (AWP) and cognitive reflection test (CRT), and to determine which cognitive and executive variables are related to their responses. Two experiments were carried out. In the first experiment, we analysed 5 AWPs, and the 3 basic CRT problems in order to check how they were interrelated and related to reading comprehension skills and fluid intelligence. In the second experiment, we explored the extent to which specific measures of executive functioning (updating, inhibition and planning) contributed to solving these problems, and the metacognitive experience of difficulty that participants had during the solving process. The associative ability of each type of problems with two measures of academic performance was also analysed. Both experiments demonstrated that AWP and CRT have different response patterns, in particular a higher difficulty to solve the CRT than the AWP, and a greater number of superficial responses in CRT than AWP. They also showed that all cognitive and executive measures were related to AWP and CRT. In exp. 2 results showed that participants underestimated the difficulty of the problems, especially in CRT; likewise, path analyses display the specific roles played by cognitive and executive measures on both type of problems. Results of second experiment also showed associative relations with the two measures of academic performance. (242 words)

Keywords: arithmetic word problems, executive functions, cognitive reflection, reasoning, academic achievement.

3.3.1. Introduction

One of the most widely used tools in education for learning mathematics has been arithmetic word problem solving. Solving arithmetic problems involves a complex cognitive process that requires a deep understanding of problem statement and the ordered applying of diverse arithmetic calculation and operations. The resolution of these apparently simple problems is not free of difficulties since it may require the activation of specific executive functions of the working memory, such as inhibiting automatic responses updating the representation of information and planning the resolution sequence.

The aim of this paper is to compare and study the role played by cognitive and executive processes in the resolution of two types of arithmetic problems: classical arithmetic word problems (AWP) and cognitive reflection problems (CRT). Most of studies about word problem solving has been carried out in primary education (Passolunghi & Mammarella, 2010; Passolunghi & Pazzaglia, 2004, 2005; Re et al., 2016; Riley et al., 1984; Sánchez-Pérez et al., 2015; Vicente et al., 2008) and many of them from the perspective of the strategies that solvers follow to solve them. In addition, a considerable number of works has been carried out with adults, most of them aimed at testing their expertise in solving these problems (Gros et al., 2019; Lubin et al., 2016; Thevenot et al., 2007). However, we hardly find previous studies with adolescents and from the point of view of the cognitive processes involved in word problems solving. In particular, how executive processes act in a range of age and schooling, secondary education, in which the basic skills of reading comprehension and mathematics have been recently acquired. Furthermore, in our knowledge, there are not prior studies comparing AWPs and CRT as well as the predictive capacity of both tasks (see Duque de Blas et al., 2021; Gómez-Veiga et al., 2018; Fuchs et al., 2006; Gómez-Chacón et al., 2014, 2011) on academic performance in Mathematics. Accurately determining these aspects could be relevant not only for educational purposes, but also to design intervention programs for the improvement of mathematical performance.

Arithmetic Word Problems (AWPs)

There is a wide variety of mathematical problems (see Riley et al., 1984), perhaps one of the most common being problems of comparison; those where the problem statement could describe two known quantities and require the difference between them to be determined (unknown difference problems), or determine an unknown quantity from the relationship described with the element of known magnitude through the applying of the basic arithmetic operations of addition, subtraction, multiplication and/or division. Compare problems can be categorized according to the combination of some criteria (see Duque de Blas et al., 2021). Two of the most manipulated criteria in experimental field to modulate the difficulty are: the number of operations required for the resolution; and the consistence/inconsistence between the verbal terms used in the statement. A greater number of propositions and significant elements in the problem's statement will lead to an increase in the number of operations needed and consequently an increase of difficulty. On the other hand, the calculation procedure will be determined on the basis of the type of relationships established between the elements described in the statement. These relations are made explicit through the use of adverbs of quantity and can give rise to two types of problems: consistent and inconsistent. We will say that a problem establishes consistent relations when the adverb of quantity ("more than" or "less than") coincides with the arithmetic operation required ("addition" or "subtraction", respectively). When the meaning of the adverb does not agree with the required operation to solve the problem, we will say that the relations it establishes are inconsistent. In that case, the applying of the correct arithmetic operation will require the inhibition of a heuristic calculation procedure ("more than" – addition; "less than" - subtraction; see Jiang et al., 2020; Khng & Lee, 2009; Shum & Chan, 2020) and the updating of the algorithmic sequence of resolution by means of the inverse operation to the one suggested by the adverb ("more than - subtraction" and "less than - addition"). The problems used in this study were one and two add/subtract operations, which can be consistent and/or inconsistent plus a final simple multiplication. An example of one inconsistent operation problem is:

In SportGood, NK brand sneakers cost 50 Euros. This is 3 Euros more than what you have to pay for the same shoes at Primark. If you need to buy 4 pairs of NK sneakers, how much will you have to pay at Primark?

The main difficulty in this problem consists on determining the reversal operation (subtraction) to that suggested by the verbal term ("more than"). The possible answers in this type of problems are correct, intuitive-superficial and incorrect responses. Correct response is the consequence of correctly computing a correct plan for problem solving (in the example, 188€). An intuitive erroneous response would appear when the resolver correctly executes a wrong resolution plan (212€), giving rise a superficial response due to a misunderstanding of its content. The third possibility, incorrect response, would occur when the solver makes mistakes in the calculation, whether his/her solving plan was correct or not. Previous studies with secondary students show that the most frequent response in global AWP will be correct answers. Likewise, since the computing of three operations is necessary in all cases, it is reasonable to expect that errors will be the most frequent wrong answer in AWPs (see Duque de Blas, et al., 2021.

Cognitive Reflection Test (CRT)

The cognitive reflection test (Frederick, 2005) is a brief and difficult task which presents 3 mathematical word problems designed to generate an intuitive, attractive and immediate, but erroneous, solution. The purpose of this test is to test the resolver's ability to overcome this response trend and to construct a more elaborate situational model that allows the correct response to be achieved (for a review, see Campitelli & Gerrans, 2014; for a meta-analysis, see Brañas-Garza et al., 2019). An example of the CRT task could be:

> *A bat and a ball cost $1.10. The bat costs $1.00 more than the ball. How much does the ball cost? ___ cents*

The task of cognitive reflection served to evidence the existence of two types or systems of reasoning processes, one intuitive fast and guided by data, expectations and preferences; and the other slow reflective, capable of inhibiting immediate prepotent responses (Evans & Stanovich, 2013; Frederick, 2005; Kahneman, 2011; Toplak et al., 2014). The willingness to reflect has traditionally been related to an attitude of greater individual patience (Frederick, 2005) and would probably imply the inclusion in the sequence of the problem-solving process of a higher order skill capable of evaluating and supervising the response in order to detect inconsistencies or processing biases committed during the integration phase. In this sense, CRT would constitute a task in which not only the subject's reflective tendency to inhibit automatic responses would come into play, but also a test in which impulsivity and speed of response would play a leading role (Szaszi et al., 2017). Thus, the detection of inconsistencies would activate a process of inhibition (Kahneman, 2003) and, later, of attentional change (Oldrati et al., 2016) that would allow directing the focus of the efforts towards the construction of a situational model that includes those characteristics that, by effect of the prepotent response or any other reason, were not integrated in the initial model. The popularity of CRT has increased over the last decade and it is relevant to this study because its measure has been associated to some arithmetic skills, especially those where numeracy plays a role in arriving at the correct response (Sinayev & Peters, 2015). In that sense, despite the different nature of the tests, CRT problems could have some common features to the inconsistent AWPs. In particular, in both types of problems, especially in CRT, superficial responses are frequent answers to be found. This is one of the hypotheses that will be considered in this study.

Studies conducted with the *bat and ball* problem indicate that participants, even university students, tend to erroneously answer that the ball costs 10 cents as a consequence of a simple subtraction of the two magnitudes expressed in the problem (Carriedo et al., 2020). However, a reflective and paused approach to the problem would allow a *cognitive decoupling* (Evans & Stanovich, 2013a) of this tendency towards the intuitive superficial response to conclude that the solution to the problem is 5 cents, since that amount is $1 *more than the* cost of the bat ($1.05). The reason why the solver decides to carry out this intuitive simple arithmetic calculation has been the source of numerous research works. Thompson et cols., (2013) identify it as a consequence of the performance of an heuristic called "perceptual fluency" that would provoke a facilitation of access to superficial response.

A main feature of CRT is that the erroneous intuitive responses to the problems is related to a tendency to underestimate their difficulty (Frederick, 2005). Thus, participants giving the erroneous intuitive response “10 cents” consider the problem easier than those participants giving the correct reflective response “5 cents”. Therefore, the results in this test can also be analysed from the point of view of the solver’s metacognition and the role that it can plays in the transition from intuitive to reflective responses (Thompson, 2010). In this line, Thompson et al. (2011) maintains that System 1 processes are associated to a metacognitive awareness, called the *feeling of rightness* (FOR), which can suggest the necessity of an additional System 2 analysis.

Reasoning skills in solving AWPs and CRT

Reasoning is the process by which we operate with a representation or mental model in order to infer non-explicit information (Johnson-Laird, 2006). Although mathematical thinking has been linked to various reasoning skills, there is some consensus that deduction is one of its main components. Deductive reasoning constitutes a criterion from which some differences can be established between expert and novice behaviour in a task (Sternberg & Ben-Zeev, 1996). Deduction allows propositions to be categorized as true or false regardless of their content, and therefore acts as a protective factor against the assumption of unrealistic, or even implausible, answers when faced with the task of solving AWP or CRT (Wyndhamn & Säljö, 1997).

The operational descriptions of the two systems of reasoning processes make possible to establish an explanatory framework on the strategies that individuals use when faced with the task of arithmetic problem solving. On the one hand, there would be an intuitive strategy, characteristic of the System-1 processes, focused on the quantitative aspects of the problem, in which the procedures considered appropriate for the task are applied on the basis of the experience previously acquired by the individual. This resolution strategy would lead to the mere selection of the magnitudes described in the problem in order to unreflectively apply a concrete calculation procedure based on the directionality indicated by the literality of the keywords contained in the statement. This direct strategy, also present on CRT problems, reveals a biased tendency of thinking aimed at economizing cognitive resources that may cause an inappropriate estimation of the difficulty of the task, something that has motivated some authors, such as Siegler & Jenkins (1989), to define it as a process in which action drives the thinking of the individual: "calculate first; think later".

Another strategy for solving AWPs is called the "problem model strategy". Unlike the previous one, it is an ordered and controlled sequence of stages in which the action of the individual is at the mercy of his ideas, and whose goal is the creation of an image or mental model that represents the qualitative aspects of the problem (Hegarty et al., 1995; Mayer & Hegarty, 1996; Thevenot, 2010).

The process of forming the situational representation of the problem can be explained with great precision by applying the construction-integration model (Kintsch, 1998; Kintsch & Greeno, 1985) that includes four stages for reading comprehension. The rationale of this analysis is that the result of final phase is the formation of a situational model that require the inhibition of the prior representations generated automatically by the reading process. This would be the case of inconsistent AWP, where a literal interpretation of the quantity adverbs described in the statement would lead to the application of the wrong arithmetic operation. But it is also key in CRT problems because, guided by the context of the task, participants could generate incomplete interpretations and models based on superficial characteristics of the statement.

After the consolidation and integration of the mental representation of the problem, it is necessary to start a phase in which the calculation procedures to be carried out to find the solution to the problem. The statement and context of the problem will allow to infer whether the correspondence "more than - addition" and "less than - subtraction" is valid for each arithmetic problem. Otherwise, as with inconsistent and CRT problems, the arithmetic operation applied during the final calculation process should be the inverse/different of that suggested by the literality of problem statement. The final step will consist in an arithmetic calculation from the magnitudes described in the problem. This operation will require the recovery of the algorithmic procedures involved in each of the arithmetic operations, as well as a process of supervision of the action aimed at minimizing calculation errors.

The role of Working Memory and executive functions

Mathematical thinking is an ability which involves to handle linguistic codes and numerical symbols with which one operates to achieve a numerical response (Sternberg & Ben-Zeev, 1996). This ability has been obviously related to fluid intelligence abilities and inductive, deductive and quantitative reasoning, generally measured by means of visuo-spatial reasoning tasks. Thus, multiple studies have shown the importance of intelligence as a predictor of mathematical problem solving (see, for instance, Bull & Lee, 2014; Fung & Swanson, 2017; McGrew & Hessler, 1995). More recently, Peng and collaborators (2019) carried out a metanalytic study that showed that the relations between fluid intelligence and mathematics increase with age, and are stronger with complex mathematical skills and problems than with initial and basic ones.

As we have underlined, solving mathematical problems is a cognitive complex process depending on the execution of various skills that operate in a mental space, the working memory (WM). In this mental space of processing and temporary storage, the incoming information is managed and related to the pre-existing information through the coordinated activity exercised by the central executive (CE) system. The rest of components that constitute the WM are the phonological loop and the visuospatial sketchpad in charge respectively of verbal and visuospatial information; and the episodic buffer that directly connect with CE (Baddeley, 1986, 1990).

CE activity can be described in terms of specific executive processes, such as: 1) focusing and maintaining the attention during the time required to solve the task; 2) shifting the focus of attention between the diverse subtasks and subroutines involved; 3) updating the contents of the information as the task progresses and connecting and integrating the new information with the knowledge already available in the long-term memory; 4) inhibiting information or responses that are not relevant, like prepotent superficial responses, or computing biases, as well as preventing from an overload of the WM system; (see Baddeley, 2000; García-Madruga et al., 2016; Miyake et al., 2000). Apart from these on-line basic WM executive processes, there are other higher order executive functions that are required in most complex thinking activities as writing and some problem solving and reasoning task: planning and revision (García-Madruga et al., 2016; García-Madruga et al., 2022). Given the simple mathematical calculations required by the AWP and CRT problems, an exhaustive revision process may not be as necessary as the mandatory planning of the sequence of steps for resolution.

Previous research has identified that limitations in the span of WM can negatively influence the solving mathematical problems (Cooney & Swanson, 1990; Friso-van den Bos et al., 2013; Passolunghi & Siegel, 2001; St Clair-Thompson & Gathercole, 2006). Some studies quantify this influence of the amplitude of WM in a quarter of the variance explained in the resolution of mathematical problems (see Lee et al., 2009). These results have been confirmed recently both in children and adults. Formoso et al. (2018) found that verbal and visuospatial WM scores were associated to solving mathematical problems in primary school students. In particular, visuospatial WM was associated to children that depend on visual aids or finger counting, and both visuospatial and verbal WM were associated to those who used mental calculation. Likewise, Cragg et al., (2017) compared the arithmetic performance between children and adults, concluding that arithmetic performance rely heavily on WM, and is independent even of the strategy they use to carry it out.

As for executive functions, the ability to update content is highlighted as a basic and essential function during the construction of the situational model of the problem (Hegarty et al., 1995). Along these lines, Passolunghi & Siegel (2001) demonstrated that students with difficulties in solving arithmetical problems failed in their efforts to reduce access to non-relevant information contained in the statement and, as a result, registered a greater number of intrusions in WM tasks that compromised their competence. These results were later confirmed (Passolunghi & Pazzaglia, 2004, 2005) in a sample of students with difficulties in solving arithmetic problems, those who committed a greater number of intrusions and showed a lesser capacity to recover information in tasks that had required updating the contents in the WM. Recent works points to the crucial role of working memory on the explanation of mathematical performance and the selection of the proper strategy for solving. In this line, Friso-van den Bos et al. (2013) demonstrated that verbal updating ability is the component of WM with stronger correlation with mathematical performance, above inhibition or switching. Hammerstein et al. (2019) found that children with more efficient updating ability showed greater flexibility in choosing a solving strategy, were also more adaptive, and showed better performance on two-digit additive computational estimation tasks. Also, Passolunghi et al. (2021) demonstrated in a sample of 182 primary school participants that there was a difference in performance on two-operation problems with an inconsistency. When the inconsistency was located in the second operation, the number of correct answers was significantly lower than when the inconsistency was located in the first operation. And these differences were no longer significant after controlling for participants' updating ability.

Updating ability necessarily requires the symbiotic action of the *inhibition*, capable of displacing previous contents so that the new information that is the object of the task can be incorporated into the mental space. St Clair-Thompson & Gathercole (2006) reported that, when the updating abilities of participants were controlled, inhibition skills was significantly associated with general academic learning rather than the acquisition skills and knowledge in specific domain. Lubin et al., (2013, 2016) points to the importance of the inhibition for cancelling some overlearned strategies of students when they are faced to resolve AWP. Such heuristics are referred, for example, to "*add* if more" or "*subtract* if less" that, in presence of inconsistences inserted on the problem statement, would drive to commit mistakes. In that sense, the task would cause that analytical and heuristic thinking processes to compete with each other, if the solver catches a discrepancy in the problem situation described by the statement. Inhibition would act by cancelling the heuristic response sequence, which is more accessible in memory. Inhibition, as well as switching, has been also associated to a better performance on CRT task (Toplak et al., 2011).

As for planning, considered by some authors as an offline executive function (García-Madruga, Gómez-Veiga, et al., 2016), was defined by Luria (1966) as the ability to organize behaviour in order to achieve a specific goal which can be broken down into a series of sub-goals or intermediate steps. As defined in Sorel & Pennequin (2008), this function can be approached as both a bottom-up accommodative activity (formulating plans) and a top-down assimilative activity (carrying out plans). One of the tasks traditionally used as a planning measure is the Tower of London (Shallice, 1982). This task have been validated as a measure of functioning of frontal brain regions (Shum et al., 2009; Unterrainer & Owen, 2006). This function would be involved in complex intellectual abilities, such as reasoning (Andrews et al., 2014; Handley et al., 2002) or problem solving (Eichmann et al., 2019; Unterrainer & Owen, 2006); and also related to other executive functions, such as the inhibition processes (D'Antuono et al., 2017; Humes et al., 1997) or working memory (Gilhooly et al., 2002; Injoque-Ricle & Burin, 2011). In particular, Sikora and cols. (2002) found that a group with arithmetic difficulties exhibited significantly greater impairment on the Tower of London (TOL) than either a group with reading difficulties or a group with no difficulties than a group without difficulties.

3.3.2. Experiment-1

The first experiment aimed to compare the two types of mathematical problems (AWP & CRT) in 4th grade of secondary school, and to explore their relationships with fluid intelligence and reading comprehension. The proportion of correct and superficial responses in both AWP and CRT was also analysed to check the accuracy of solving the problems and infer a pattern of responses. Our hypotheses were:

1. As in previous studies, the most frequent response in global AWPs will be correct answers, and the most frequent errors will be non-superficial erroneous responses. Likewise, the most frequent answer with CRT will be superficial responses. We also expect a higher difficulty and greater number of superficial responses in CRT than in AWP Inconsistent.

2. Regarding the two criteria AWP are divided, we expect to find a significant higher number of correct responses in consistent than inconsistent problems. We also expect differences in difficulty between one and two-operations problems.

3. Given the common necessity to construct a situational model of problems to get a correct solution, we expect to find positive intercorrelations between inconsistent AWP and CRT correct and superficial measures. We also expect to find a moderate relationship of the cognitive measures, fluid intelligence and reading comprehension, with AWP and CRT. Therefore, regression analyses report a significant power of association of the problems with the cognitive measures.

3.3.2.1. Method

3.3.2.1.1. Participants

A sample of 81 native Spanish speaking students (36 female) of 4th grade of compulsory secondary education of a public institute (Madrid) participated in this study. The average age was 15.28 (*SD* = 0.68).

3.3.2.1.2. Task and measures

Arithmetic Word Problems (AWP)

The same problems described in Duque de Blas, et al. (2021) were used to assess AWP resolution skills. Therefore, 5 AWPs for comparison (2 consistent and 3 inconsistent) that involve elementary arithmetic calculations consisting of addition or subtraction, as well as one final multiplication. The more consistent or inconsistent relationships described between the elements of the problems, the more operations involved. A detailed description of these problems can be seen in table 1. Cronbach's alpha for internal consistency in AWPs was 0.63.

Table 1. Description of problems in terms of number of consistent or inconsistent operations (working memory load), inconsistency, and need of inhibition of superficial responses.

	WM Load Number of Operations (Add/Subtract + Mult.)	Inconsistency	Superficial Responses
Problem-1	1 + 1	No	No
Problem-2	1 + 1	Yes	Yes
Problem-3	2 + 1	No–No	No
Problem-4	2 + 1	No–Yes	Yes
Problem-5	2 + 1	Yes–Yes	Yes

Cognitive Reflection Test (CRT; Frederick, 2005)

The CRT problems were based in Frederick's cognitive reflection task (2005). We used the three classical CRT problems, administered always in the following order: bat-ball, machines and lily pads. Cronbach's alpha for internal consistency in CRT was 0.60.

Kaufman brief intelligence test (KBIT)

The matrix subtest of the Kaufman Intelligence Brief Test (KBIT, Kaufman & Kaufman, 2000) was the measure used to assess the abstract and visuospatial reasoning, and fluid intelligence of participants. This test was designed to avoid the cultural influences and the ability of the participant to form verbal concepts. The overall score varied in a range between 0 and 48 correctly solved items. Cronbach's alpha for internal consistency was 0.73

Reading processes assessment battery (PROLEC-SE)

The Semantic processes task of PROLEC-SE (Ramos & Cuetos, 1999) was the measure used to assess reading comprehension of the participants. The participants completed the task number five of the battery, formed by two expositive texts. Participants must read the texts and answer 10 questions of each text by remembering literal aspects mentioned or making inferences without further text consulting. The overall score varied in a range between 0 and 20. Cronbach's alpha for internal consistency was 0.70.

3.3.2.1.3. Procedure

The evaluation had place at the beginning of the course. The participants had to solve the tasks in two sessions distributed over a week. The order of test presentation for the first session was PROLEC-SE and AWP; while in the second session they performed KBIT and CRT. Participants completed the tasks digitally via online forms in the computer room available in the centre during school hours. The protocols of this study were approved by the ethics committee of the Universidad Nacional de Educación a Distancia (UNED).

3.3.2.1.4. Data analyses

The analyses were carried out using the statistical package for social sciences, SPSS v24. A winsorization procedure (Field, 2013; Wilcox, 2011) on global and partial tasks of the AWP, as well as on the global scores of the tasks (CRT, KBIT and PROLEC), was performed in order to control the outliers. Descriptive results on the tasks, ANOVAs, bivariate Pearson correlation coefficient and regressions among cognitive measures, AWPs were run in order to test the power of association between AWP and CRT on Reading comprehension task.

3.3.2.2. Results

Descriptive statistics and comparison

The first analysis was to check for differences between males and females in each of the used tasks. There was not any significant difference related to this variable. Descriptive statistics (mean and standard deviation), range and percentage of correct responses on AWP and CRT tasks are depicted in table 2. Means (standard deviations) and percentages of correct responses in the two cognitive variables were, respectively: KBIT: 37.38 (3.92) and 78%; PROLEC-SE:12.35 (3.51) and 62%.

An ANOVA for repeated measures confirmed that correct responses to CRT were significantly lower than to AWP Inconsistent (F (1, 81) = 133.89, MSE = 8.45, $p < .001$, $\eta 2p$ = .63). Another ANOVA confirmed the number of superficial responses was significantly higher on CRT than AWP Inconsistent (F (1, 81) = 150.16, MSE = 114.17, $p < .001$, $\eta 2p$ = .65).

Table 2. Descriptive statistics (mean and standard deviations), range and percentage of correct responses on tasks.

N = 81	Range	Correct and superficial responses	
		Mean (SD)	Percentage
Arithmetic Word Problems (AWP)			
Consistent			
1. AWP-1	0-1	0.83 (.38)	-
2. AWP-3	0-1	0.60 (.49)	-
3. AWP-Consistent	0-2	1.43 (0.59)	71
Inconsistent			
4. AWP-2	0-1	0.48 (.50)	-
AWP-2-sup	0-1	0.16 (.37)	-
5. AWP-4	0-1	0.48 (.50)	-
AWP-4-sup	0-1	0.15 (.36)	-
6. AWP-5	0-1	0.35 (.48)	-
AWP-5-sup	0-1	0.17 (.38)	-
7. AWP-Inconsistent	0-3	1.31 (1.14)	44
AWP-Inconsistent-sup	0-3	0.48 (0.91)	16
Operations			
8. one-operation	0-2	1.31 (0.61)	65
9. two-operations	0-3	1.43 (0.60)	48
10. AWP-Global	0-5	2.74 (1.38)	55
Cognitive reflection test (CRT)			
11. CRT-1	0-1	0.06 (.24)	-
CRT-1-sup	0-1	0.75 (.43)	-
12. CRT-2	0-1	0.14 (.35)	-
CRT-2-sup	0-1	0.75 (.43)	-
13. CRT-3	0-1	0.09 (.28)	-
CRT-3-sup	0-1	0.65 (.48)	-
14. CRT-Global	0-3	0.28 (.65)	9
CRT-Global-sup	0-3	2.16 (.99)	72

The differences in the number of correct responses among the diverse AWP problems were significant, as shown by a mixed ANOVA for repeated measures. The five AWP were arranged as levels of the factor, "performance differences in AWP" (F (4, 81) = 14.29, MSE = 2.65, $p <$ .001, η^2_p = .15.) Post-hoc tests using the Bonferroni correction showed significant differences only between AWP-1 and all of them ($p <$ .050), and between AWP-3 and AWP-5 (p = .004). Also, there was a significant difference in difficulty between consistent and inconsistent problems (F (1, 81) = 20.86, MSE = 1.10, $p <$ .001, η^2_p = .21.) and between one and two-operations problems (F (1, 81) = 16.28, MSE = 1.27, $p <$ .001, η^2_p = .17).

Regarding CRT, an ANOVA for repeated measures was run for checking whether the different number of correct responses obtained on the problems were significant. The three CRT problems were arranged as levels of the factor "difficulty of CRT". Results showed non-significant differences in difficulty among the problems (F (1.77, 81) = 2.03, MSE = 0.13, p = .141).

Correlations

Regarding the third hypothesis, table 4 shows the intercorrelations among the measures. Pearson correlation between AWP-Global and CRT-Global was positive, but low and only marginally significant (r (79) = 17, p = .07). No AWP tasks (consistent, inconsistent, one or two-operations) correlated significantly with CRT-Global. Neither consistent or one-operation problems correlated with fluid intelligence nor reading comprehension. On the contrary, inconsistent problems and two-operation problems correlated significantly with the cognitive variables, more with reading comprehension than fluid intelligence. As for CRT, there were positive and significant correlations with the cognitive variables fluid intelligence and reading comprehension.

Table 3. Bivariate Pearson's correlations between AWP (consistent, inconsistent, 1-operation, 2-operations, global and superficial inconsistent responses), CRT, fluid intelligence (KBIT) and reading comprehension (PROLEC-SE).

	1	2	3	4	5	6	7	8	9	10
1. AWP-1op	1									
2. AWP-2op	.29**	1								
3. AWP-CON	.46**	.49**	1							
4. AWP-INC	.57**	.84**	.19*	1						
5. AWP-Sup	-.43**	-.31**	-.18	-.42**	1					
6. AWP-Global	.67**	.91**	.59**	.91**	-.43**	1				
7. CRT-Global	.09	.16	.10	.15	-.02	.17	1			
8. CRT-Sup	-.02	.13	.09	.07	.16	.10	-.57**	1		
9. Fluid Intelligence	-.07	.32**	-.03	.28**	-.00	.22*	.29**	-.18	1	
10. Reading Compreh.	.16	.38**	.10	.38**	-.02	.36**	.27**	.06	.34**	1

Sup: superficial responses; ** $p < .01$; * $p < .05$; one tail

Regressions

Finally, a linear regression analysis was carried out to check the contribution of both problems (AWP Inconsistent and CRT) on the explaining of fluid intelligence and reading comprehension results. Results confirmed the power of association of both types of problems on the cognitive abilities. In particular CRT was similarly associated to fluid intelligence ($R^2adj = .07$; $F(1, 79) = 7.40$; $MSE = 105.13$; $p = .008$; $\beta = .29$) than AWP-Inconsistent ($R^2adj = .07$; $F(1, 79) = 6.83$; $MSE = 97.66$; $p = .011$; $\beta = .28$), but AWP-Inconsistent was more associated to reading comprehension ($R^2adj = .14$; $F(1, 79) = 13.61$; $MSE = 144.96$; $p < .001$; $\beta = .38$) than CRT ($R^2adj = .06$; $F(1, 79) = 6.28$; $MSE = 72.67$; $p = .014$; $\beta = .27$).

3.3.2.3. Discussion

The main result of this experiment is the relevant difference found between the two types of arithmetic word problems used: CRT problems are clearly more difficult that Inconsistent AWP and show a greater tendency to intuitive superficial responses. However, both kind of problems share some commonalities, particularly, they show System 1 and System 2 processes are acting on the solution processes: the two kind of responses, reflective and intuitive-superficial are present.

In relation to AWPs, the results obtained should be analysed from the two criteria in which AWPs are classified, consistency/inconsistency of the relationships between the elements of the problem and number of operations. According to the first criterion and confirming the results of prior studies, the results indicate a significant difference in difficulty between consistent and inconsistent one-operation problems (Hypothesis 2). Similarly, there are differences in difficulty between consistent and inconsistent two-operation problems. However, no entirely significant differences have been found between one- and two-operation inconsistent AWPs. Instead, differences only existed when these problems were consistent. Therefore, we can say that, for this sample of 15-year-old students, the number of computational operations required in a problem is less relevant than the presence of inconsistencies in the problems. This result coincides with that found by Duque de Blas and cols. (2021), who carried out a similar experiment with students of 2^{nd} and 3^{rd} secondary grade.

Inconsistent problems are more complex since they require a deep understanding and the construction of a situational model that involve the inhibition of superficial information and representations. Although problems with two inconsistent operations are more difficult since they imply more arithmetic calculations and an increase of the updating work, none of these requirements seem to be as relevant as that required by inconsistent problems

Regarding CRT problems, our results seem to confirm that they are not typical arithmetic word problems: they require more System 2 work, that is, they measure the willingness to give a reflective answer and overcome intuitive responses driven by the problem statement. Results observed in CRT show a remarkable ground effect, not entirely unexpected given the high difficulty of these problems even with elite university students (Frederick, 2005). Our 15 years old students found very difficult CRT problems (9% correct responses). However, results found in diverse versions of CRT with science students one year older, who have more Mathematics in their syllabus, are clearly higher, about 30% of correct responses (Gómez-Chacón et al., 2014; Gómez-Veiga et al., 2018).

Besides the diverse difficulty, the difference between the two tasks become clearly apparent when we compare the percentages of superficial answers on CRT (72%) and AWP Inconsistent (16%). Another interesting result is the percentage of erroneous non-superficial responses in AWP (40%), that is hardly lower than correct responses (44%). Our participants seen to have unexpected difficulties with computing simple arithmetic operations in AWP task, as well as in understanding the problem statement and developing a situational model of the problem. Other kinds of motivational causes cannot be ruled out for the high number of non-superficial errors in the task. Our third hypothesis on the expected correlations between AWP and CRT are not entirely confirmed. We found positive correlations between correct and superficial in both tasks, but they don't reach a significant level. A possible explanation of this lack of significance is the high difficulty found by our participants with CRT problems, at this age level.

From a developmental perspective, solving AWPs involves some fully acquired skills and thus, even most difficult inconsistent problems, are solved by a majority of our participants. On the contrary, CRT problems involve a deeper and more demanding reflective work that seems not have been acquired by the participants.

Both type of problems obtained, as predicted, significant correlations with the cognitive measures, fluid intelligence and reading comprehension. Fluid intelligence correlates much higher, as expected, with inconsistent AWP than consistent ones. The interesting result is that inconsistent AWP and CRT problems correlates in a similar way with fluid intelligence, whereas inconsistent AWP correlates higher with reading comprehension than CRT problems. This power of associations was also confirmed, by means of regression analyses, similar for CRT and AWP with fluid intelligence, and higher for AWP with reading comprehension than CRT. Therefore, although both cognitive measures are related to the efficacy in solving these problems, it seems that AWPs are more related to the processes of generating a situational model of the problem given their narrative structure. In contrast, conclusions regarding the cognitive processes involved in solving CRT problems are too bold due to the remarkable ground effect obtained in this study. It will therefore be necessary to confirm these results and also obtain new measures of executive processes to explore how they are involved in solving these problems, which will be addressed in experiment 2.

3.3.3. Experiment-2

Experiment 2 has four main objectives. First, to check the results obtained in the previous experiment in relation to the response accuracy obtained in AWP and CRT in a sample of 3rd and 4th grade secondary school students. We wanted to explore the origin of high difficulty of CRT using a new sample of 3rd and 4th grade secondary students, from a private school. Second, to relate the response accuracy in the problems to participants' metacognitive experience of difficulty. To measure participants assessment of problems difficulty we will use a Likert scale, not the classical measure of metacognitive rating of difficulty used in CRT problems, based on expected percentages of correct responses given by participants mates (Frederick, 2005). In a pilot study a relevant percentage of secondary students had showed diverse problems to correctly understand the classical metacognitive measure. Third, and most importantly, to explore relationships of the response accuracy on these problems to new measures of executive functions (inhibition, updating and planning) as well the previous cognitive measures (fluid intelligence and reading comprehension). Finally, we were interested in checking out the capacity of AWP and CRT, together with the cognitive variables, to explain academic achievement. The hypotheses considered in this experiment were:

1. We expect to confirm main results obtained in experiment 1. The most frequent response in AWP will be correct ones. From the two variables affecting to AWPs' problems consistency/inconsistency will be more relevant than one/two operations. The percentage of correct responses will be higher on AWP than on CRT and the percentage of superficial responses will be higher on CRT than on AWP.

2. We expect to find significant differences on the rate of difficulty perceived by participants on the AWP tasks. Consistent problems will be rated easier than inconsistent problems. The difficulty of CRT problems will be underestimated, that is, they will be considered easier than they really are.

3. There will be positive correlations between cognitive and executive measures and AWP and CRT problems. This pattern of correlations will be different for the diverse kind problems. AWP inconsistent and CRT will be also intercorrelated.

4. We expect a relevant association of updating, inhibition, planning, fluid intelligence, and reading comprehension on AWP and CRT tasks performance. Path analyses will confirm these associations.

5. We expect a significant contribution of the two basic cognitive abilities (reading comprehension and fluid intelligence), as well as AWP and CRT on the explanation of the results in both academic achievement measures (Mathematics performance and overall academic scores).

3.3.3.1. Method

3.3.3.1.1. Participants

A sample of 82 students (41 female) of 3rd grade (*N* = 42) 4th grade of secondary education of a private high school in (Madrid, Spain) participated in this study. The average age for 3rd grade was 14.00 (*SD* = 0.44) and 15.10 (*SD* = 0.44) for 4th grade.

3.3.3.1.2. Task and measures

Participants had to solve the same AWP and CRT problems that were used in experiment 1. In addition to the measures taken in experiment-1 -KBIT for fluid intelligence and PROLEC-SE for reading comprehension-, the following measures were also taken.

Inhibition: A STROOP test was administered consisting of 76 practice stimuli plus 48 experimental stimuli of three different colors (red, blue and green), randomly distributed in equal proportion between congruent and incongruent assays. Both responses, precision and reaction time (ms), were electronically recorded. Cronbach's alpha for internal consistency was 0.65.

Updating: PANAL-SE; (Prueba de Memoria operativa de Analogías para secundaria, or WM Analogies task for secondary school; García-Madruga & Duque de Blas, 2018; adapted from Gutiérrez-Martínez et al., 2005). In this task, participants were requested to store information while simultaneously processing the same or other concurrent information. They must read aloud and solve a series of verbal analogies and simultaneously hold in the WM the words that solve them; at the end of each series, they must recall and say the solution-words in the same serial order as the analogies are presented on a computer screen. Verbal analogies are simple and easy to solve, e.g., "Car is to road as train is to: a) travel; b) railway". The task has several levels of difficulty, so that the difficulty increases as the number of analogies (between 2 and 6) that make up a series increases; each level is made up of three series of analogies. Cronbach's alpha for internal consistency was 0.62.

Planning: Tower of London-Drexler (TOL^{DX}; Culbertson & Zillmer, 1998). This test measures the ability to mentally project the sequence of actions that will be taken to solve a problem. In this classic test, a board with three pegs of different sizes and three wooden balls of different colors (red, green and blue) are presented to the participants. Each ball can be inserted into the pegs through the hole that runs through them. The longer the peg, the more balls can be inserted into it. Thus, 3 balls can be inserted in the longest peg, 2 in the middle peg and 1 in the shortest peg. Participants must replicate a model from a specific initial arrangement of balls. To do so, they must replicate the model in as few moves as possible, moving the balls one by one. It consists of 12 trials that must be completed in a maximum time of 2 minutes per trial. Cronbach's alpha for internal consistency was 0.46.

Metacognition: Participants had to assess in a Likert scale from 1 (very easy) to 7 (very difficult) the grade of difficulty experienced after responding each AWP and each CRT problem.

Academic Performance: two different measures were considered:

COMPEMAT: It measures the mathematical knowledge available to students in Secondary Education (ESO) independently of any other behavioural or motivational variables in classroom. It requires to complete different exercises such as arithmetic, geometry, interpretation of diagrams, sectorial graphics or Cartesian axes. This test is formed by the examples released by the Ministry of Education of the Government of Spain for the periodic assessment of the level of knowledge acquired in ESO (version 1.0). Cronbach's alpha for internal consistency was 0.63.

Overall academic scores: It is the mean of final teachers scores received by each student in the diverse subject matters (Mathematics, Natural and Social Sciences, Spanish and English) of each of the two grades of participants.

3.3.3.1.3. Procedure

The evaluation had place at the end of first quarter of the season. The participants had to solve the tasks in three sessions distributed over a week. The test presentation order for the first session was TOL and PANAL; in second session, the order was PROLEC-SE and STROOP. Third session, the order was AWPs and KBIT. Participants completed the tasks digitally via online forms in the computer room available in the centre during school hours. They were divided in two classrooms working parallelly. The protocols of this study were approved by the ethics committee of the Universidad Nacional de Educación a Distancia (UNED).

3.3.3.1.4. Data analyses

Descriptive statistics, ANOVAs and Pearson's for comparison between scholar grades and correlation analyses were carried out in SPSS v24. Path analysis and regressions for prediction on academic achievement were run by Lavaan R package in R studio environment (Rosseel, 2012). Previously, a winsorization procedure (Field, 2013; Wilcox, 2011) on global and partial scales of the AWP task, as well as on the global scores of the tasks, was performed in order to control the outliers.

3.3.3.2. Results

Descriptive statistics and comparison between school grades

Table 4 shows the descriptive results across grades and the whole sample. It includes a multivariate ANOVA that was conducted to test differences between 3rd and 4th grades on the administered tasks.

Results confirmed no differences between school grades in arithmetical problems (AWP and CRT). There was a reliable difference between two grades in the metacognitive assessment of difficulty of AWP but not in CRT. Regarding the cognitive measures, a multivariate ANOVA confirmed differences between school grades in fluid intelligence (KBIT), updating (PANAL-SE), inhibition (Stroop) and the academic performance in Mathematics (COMPEMAT).

Table 4. Descriptive results (mean of correct responses, standard deviation and percentage) for 3rd, 4th grade and both in arithmetic problems (AWP, CRT and their metacognitive assessments), fluid intelligence (KBIT), reading comprehension (PROLEC-SE), updating (PANAL), planning (TOL), inhibition (STROOP), Mathematical knowledge (COMPEMAT) and overall academic scores, and differences between scholar grades.

Tasks	Mean correct responses		Difference between 3rd and 4th grades			
			Correct responses			
	M (SD)	Percentage	3rd M (SD)	4th M (SD)	*F*	*p*
Word Arithmetic Problems & Cognitive Reflection						
AWP-Global	2.67 (1.41)	53	2.55 (1.35)	2.8 (1.47)	0.66	.420
AWP-meta	2,31 (0,97)		2,59 (1,06)	2,03 (0,79)	7.36	.008
CRT-Global	0.18 (0.47)	6	0.21 (0.47)	0.15 (0.48)	0.37	.543
CRT-meta	3,14 (1,27)		3,17 (1,39)	3,11 (1,15)	0.05	.815
Fluid Intelligence & Reading Comprehension						
KBIT	37.59 (4.79)	78	36.07 (4.86)	39.18 (4.22)	9.48	.003
PROLEC-SE	12.62 (3.62)	63	12.10 (3.59)	13.17 (3.62)	1.84	.179
Executive processes						
Updating	20.02 (9.95)	44	16.76 (9.26)	23.44 (9.59)	10.28	.002
Planning	32.73 (18.65)	--	36.64 (19.36)	28.63 (17.18)	3.92	.051
Inhibition	45.28 (4.53)	94	44.31 (5.92)	46.30 (1.92)	4.10	.046
Academic Performance						
COMPEMAT	11.52 (3.20)	52	10.61 (3.15)	12.49 (2.99)	7.67	.007
Overall academic score	6.80 (1.26)	68	6.88 (1.30)	6.71 (1.23)	0.40	.530

Hypothesis 1: Accuracy in Arithmetic Problems (AWP and CRT)

Table 5 shows the scores obtained by participants on the arithmetic problems. It includes the mean of correct responses (and standard deviations), rates of perceived difficulty labelled as "metacognition" (and standard deviations).

Table 5. Descriptive results (score of correct and metacognitive responses) in AWP (consistent and inconsistent problems, one and two-operations, and global score) and CRT.

	Correct responses (both)	
Task	Score M (SD)	Metacog. M (SD)
Consistent		
AWP-1	.95 (.22)	1.44 (1.01)
AWP-3	.63 (.49)	2.72 (1.38)
Global	.79 (.28)	2.08 (1.04)
Inconsistent		
AWP-2	.37 (.49)	1.63 (.99)
AWP-4	.34 (.48)	2.56 (1.33)
AWP-5	.38 (.49)	3.21 (1.59)
Global	.36 (.37)	2.47 (1.12)
Number of operations		
AWP-one-step	.66 (.28)	2.83 (1.22)
AWP-two-steps	.45 (.35)	2.83 (1.22)
AWP-Global	2.67 (1.41)	2.31 (.97)
CRT-1	.04 (.19)	2.55 (1.57)
CRT-2	.10 (.30)	2.72 (1.61)
CRT-3	.05 (.22)	4.16 (1.81)
CRT-Global	.18 (.47)	3.14 (1.27)

Overall results of correct responses in AWP (53%) and CRT (6%) were similar to the experiment-1. An ANOVA for repeated measures confirmed that correct responses to CRT were significantly lower than to AWP Inconsistent (F (1, 82) = 49.53, MSE = 3.71, $p < .001$, $\eta 2p$ = .38).

The percentage of superficial responses in CRT (65%) was similar to the experiment-1 (72%), but this percentage in AWP (31%) was higher than the obtained in experiment-1 (16%). An ANOVA confirmed the number of superficial responses was significantly higher on CRT than AWP Inconsistent (F (1, 82) = 8.29, MSE = 0.74, p = .005, $\eta 2p$ = .09).

Regarding the correct responses obtained in AWPs, an ANOVA for repeated measures was run. The five AWPs were englobed in a factor called “AWP accuracy”, whilst school grade was introduced as inter-subject factor. Results confirmed an overall significant difference in the accuracy of the problems (F (3.3, 82) = 38.24, MSE = 6.81, $p < .001$, $\eta 2p$ = .32).

A post-hoc analysis of the problems identified differences in accuracy among all the problems (Bonferroni $p < .05$) except when comparisons were between AWP-2 (one-inconsistent operation) and AWP-4 (consistent then inconsistent operations), AWP-2 and AWP-5 (two inconsistent operations) and AWP-4 and AWP-5; that is, there were no significant differences among inconsistent problems in terms of the number of operations. Taken together, we found that both the effect of problem “consistency/inconsistency” and the effect of the “number of operations” were significant. This effect was greater in the case of consistent/inconsistent problems (F (1, 82) = 98.24; MSE = 7.61; $p < .001$; $\eta 2$p = .55) than in the case of one/two operation problems (F (1, 82) = 32.82; MSE = 1.76; $p < .001$; $\eta 2p$ = .29).

Same procedure was carried out to compare the performance on CRT. A mixed ANOVA for repeated measures confirmed no differences between problems in CRT (F (1.8, 82) = 1.77, MSE = 0.10, p = .179).

Hypothesis 2: Effect of the perceived difficulty (metacognition) on accuracy response

The second hypothesis aimed to test the relationship between the accuracy of the response to the problems and the degree of difficulty perceived by the participants. After finishing each problem, participants had to assess the grade of difficulty of the problems in a Likert scale from 1 (very easy) to 7 (very difficult).

Overall, participants' ratings of perceived difficulty of each problem were positively biased (see table 5). Participants correctly rated significantly lower the difficulty of AWP (2.31 over 7) than CRT (3.14 over 7; F (1, 82) = 59.72, MSE = 28.25, p < .001, $\eta 2p$ = .42).

An ANOVA for repeated measures on participant's valuations on the five AWP were conducted. The results confirmed a global difference between the difficulty experienced by solvers during the AWP task (F (2.7, 82) = 53.14, MSE = 68.07, p < .001, $\eta 2p$ = .40). Regarding metacognition in CRT, an ANOVA for repeated measures was performed. Results confirmed a global difference in "experienced difficulty" factor among the three problems (F (1, 82) = 36.60, MSE = 64.11, p < .001, $\eta 2p$ = .31).

Regarding the relationship between the accuracy and difficulty perception, table 8 shows the Pearson's correlations obtained between the AWP or CRT problems and their metacognition valuations.

Table 6. Pearson's correlations between accuracy and perceived difficulty on AWP and CRT.

AWP task	Metacognitive pair
AWP-Consistent	-.07
AWP-Inconsistent	-.12
AWP-Global	-.14
CRT-Global	.15

Perceived difficulty and accuracy responses were not significantly correlated. One difference between the AWP and CRT correlations is the direction of the relationship: negative in AWP and positive in CRT.

6.4. Hypothesis 3: Correlations between cognitive measures and arithmetic problems

Table 7 shows the intercorrelations between all the measures in the whole sample. The intercorrelations between the mathematical problems were positive, but low. AWP global correlated significantly and in the direction expected with all the tasks (negative for Planning task). Highest correlations were with fluid intelligence and academic Mathematics performance (COMPEMAT).

Table 7. Bivariate Pearson's correlations between AWP (inconsistent and global), CRT, fluid intelligence (KBIT), reading comprehension (PROLEC-SE), Updating (PANAL), Inhibition (Stroop), Planning (Stroop), Math's performance (COMPEMAT) and overall academic score.

	1	2	3	4	5	6	7	8	9	10
1. AWP-Inconsistent	1									
2. AWP-Global	.92**	1								
3. CRT-Global	.13	.18*	1							
4. Fluid intelligence	.31**	.39**	.19*	1						
5. Reading Compreh.	.18	.24*	.15	.46**	1					
6. Updating	.20*	.20*	.09	.32**	.41**	1				
7. Inhibition	.21*	.31**	.11	.35**	.23*	.35**	1			
8. Planning	-.21*	-.30**	-.24*	-.19*	-.14	-.27**	-.25*	1		
9. Math's performance	.38**	.42**	.29**	.64**	.41**	.31**	.34**	-.24*	1	
10. Overall academic score	.28**	.33**	.24*	.29**	.31**	.26**	.29**	-.30**	.34**	1

** $p < 0.01$; * $p < 0.05$; one tail

CRT obtained very low profile of correlations with the measures, likely due to notable ground effect of the responses. It did not correlate significantly with updating task, nor reading comprehension (PROLEC-SE) or inhibition (Stroop). Highest correlations in fluid intelligence (KBIT) were with math performance (COMPEMAT) and reading comprehension (PROLEC-SE). As for the updating measures, they were associated differently with the tasks. Highest correlations in updating measure were with reading comprehension, inhibition and fluid intelligence. Finally, higher correlations in planning (TOL) were with AWP, updating (PANAL) and the global academic mean.

To test if differences between the two types of tasks (consistent and inconsistent problems) decrease or disappear when covariates effects are controlled, we ran a repeated measures ANOVA in which cognitive and performance measures were introduced individually as covariates. Table 8 shows the variations in the effect sizes found in each of the conducted analyses in the total sample.

Table 8. Accuracy differences between consistent and inconsistent AWP tasks in terms of "effect size" after controlling for cognitive measures as covariates, in the whole sample.

Repeated measures	Size effect (η^2p) Both (N=82)
Consistent vs Inconsistent	.55***
+ cov (PANAL-SE)	.26***
+ cov (STROOP)	.00
+ cov (TOL-extra-moves)	.25***
+ cov (KBIT)	.03
+ cov (PROLEC-SE)	.09*
+ cov (COMPEMAT)	.17***

Cov: covariable; *** $p < .001$; ** $p < .010$; * $p < .050$

Results confirmed that there were not significant differences in accuracy between consistent and inconsistent problems after controlling inhibition (Stroop; F (1, 82) = 0.20, MSE = 0.02; p = .656) or fluid intelligence (KBIT; F (1, 82) = 2.65, MSE = 0.21; p = .107).

Hypothesis 4: Path analyses for AWP and CRT

Path analyses were carried out as regression equations to assess the influence of each cognitive and domain dependant measure on the AWP-Global or CRT-Global accuracy.

Both models include three basics executive functions (updating, planning and inhibition) and two mediated general cognitive measures (reading comprehension and fluid intelligence). There were computed direct effects of cognitive skills on the AWP-Global or CRT-Global, as well as the indirect effect of them mediated by domain variables. General equations for different outcomes were:

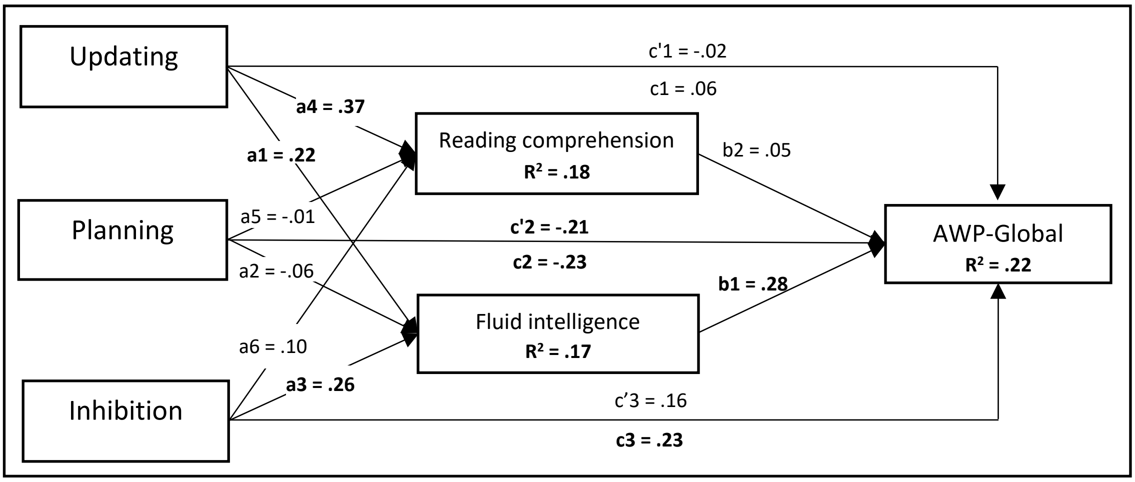
$$\textit{Fluid Intelligence} = \beta 1 + \alpha 1\ (\textit{Updating}) + e1 + \alpha 2\ (\textit{Planning}) + e2 + \alpha 3\ (\textit{Inhibition}) + e3,$$

$$\textit{R. Comprehension} = \beta 2 + \alpha 4\ (\textit{Updating}) + e4 + \alpha 5\ (\textit{Planning}) + e5 + \alpha 6\ (\textit{Inhibition}) + e6,$$

$$\textit{AWP or CRT Total} = \beta 3 + \beta 1\ (\textit{Fluid Intelligence}) + \beta 2\ (\textit{Reading Comprehension}) + c'1\ (\textit{Inhibition}) + c'2\ (\textit{Fluid intelligence}) + c'3\ (\textit{Updating}) + e7.$$

The first two regression equations estimated the effect that performance on the PANAL-SE (updating), TOL (planning) and Stroop (inhibition) tasks had on each mediator: fluid intelligence and reading comprehension. The third regression equation estimated the direct effect (c') that executive functions and mediators had on the outcomes variables (performance on AWP-Total or CRT-Total tasks). The specific indirect effect of updating on solving problems through fluid intelligence and reading comprehension was the product $a1 \times b1$ and $a4 \times b2$, respectively. The specific indirect effect of planning on solving problems through fluid intelligence and reading comprehension was the product *a2 x b1* and *a5 x b2*, respectively. Last specific indirect effect of inhibition on solving problems through fluid intelligence and reading comprehension was the product *a3 x b1* and *a6 x b2*. The direct effect (c') of the cognitive measures (updating, planning and inhibition) on the outcome was independent of the indirect effects. Consequently, direct effect was equivalent to subtract the indirect effect from the total effect of cognitive measures on AWP-Total or CRT total. Finally, direct effects were computed as the effect of the cognitive skills on the outcome after subtracting the effect of mediated variables.

Figure 1. Path model for total AWP-solving performance. Values in bold are statistically significant.



Regarding the overall measure of AWP-Global performance the total amount of variance explained by the model was R^2 = .22. The effects of the mediators (paths b) AWP-Global was explained significantly by the fluid intelligence (β = .28). Direct effects on AWP-Global were found significantly for planning (β = .21), whilst total effects were significant for planning (β = -.23) and inhibition (β = .23; see figure 1 and table 16).

Table 9. Direct, indirect and total effects of inhibition, planning, updating, reading comprehension and fluid intelligence on total AWP-solving accuracy.

AWP-Global	*Estimate*	*Std. Err*	*z-value*	*P(>|z|)*	*Std. all*
Fluid Intelligence (b1)	0.083	0.031	2.661	.008	0.284
R. Comprehension (b2)	0.020	0.041	0.482	.630	0.052
Updating (c'1)	-0.003	0.016	-0.165	.869	-0.019
Planning (c'2)	-0.016	0.008	-2.025	.043	-0.208
Inhibition (c'3)	2.311	1.623	1.424	.154	0.156
Fluid Intelligence					
Updating (a1)	0.104	0.053	1.971	.049	0.216
Planning (a2)	-0.017	0.027	-0.608	.544	-0.064
Inhibition (a3)	12.969	5.534	2.343	.019	0.255
R. Comprehension					
Updating (a4)	0.134	0.040	3.376	.001	0.369
Planning (a5)	-0.003	0.021	-0.123	.902	-0.013
Inhibition (a6)	3.904	4.168	0.937	.349	0.102
Indirect Effects					
Updating -> Intel. -> AWP	0.009	0.005	1.584	.113	0.061
Planning -> Intel. -> AWP	-0.001	0.002	-0.592	.554	-0.018
Inhibition -> Intel. -> AWP	1.076	0.612	1.759	.079	0.073
Updating -> Reading -> AWP	0.003	0.006	0.477	.634	0.019
Planning -> Reading -> AWP	-0.000	0.000	-0.119	.905	-0.001
Inhibition -> Reading -> AWP	0.078	0.182	0.428	.668	0.005
Total Effects					
Updating -> AWP (c1)	0.009	0.016	0.555	.579	0.061
Planning -> AWP (c2)	-0.017	0.008	-2.122	.034	-0.227
Inhibition -> AWP (c3)	3.466	1.631	2.124	.034	0.234

The second model includes the same equations, but with CRT-Global as the outcome. The total amount of variance explained by the model was R^2 = .08. The main feature of this path analysis with CRT-Global is that neither the predictors nor the mediators were statistically significant in explaining the results. Total Planning effect was the variable more associated with CRT-Global, it was nearly significant (β = -.22, p = .052; see figure 1 and table 9).

Figure 2. Path model for total CRT-solving performance. Values in bold are statistically significant.

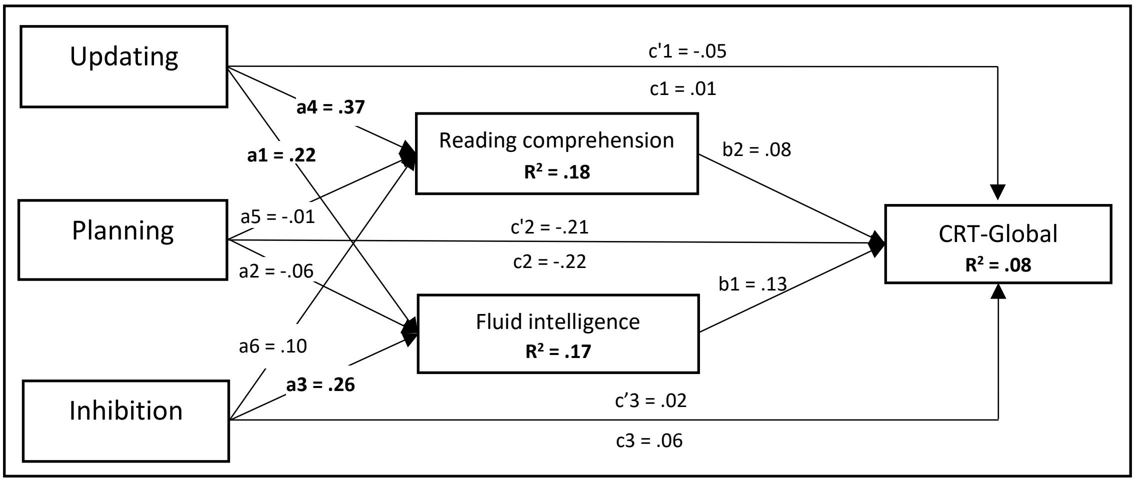


Table 10. Direct, indirect and total effects of inhibition, planning, updating, reading comprehension and fluid intelligence on total CRT-solving accuracy.

CRT-Global	*Estimate*	*Std. Err*	*z-value*	*P(>|z|)*	*Std. all*
Fluid Intelligence (b1)	0.012	0.012	1.086	.278	0.126
R. Comprehension (b2)	0.010	0.015	0.655	.513	0.077
Updating (c'1)	-0.002	0.006	-0.368	.713	-0.046
Planning (c'2)	-0.005	0.003	-1.878	.060	-0.211
Inhibition (c'3)	0.077	0.598	0.129	.897	0.015
Fluid Intelligence					
Updating (a1)	0.104	0.053	1.971	.049	0.216
Planning (a2)	-0.017	0.027	-0.608	.544	-0.064
Inhibition (a3)	12.969	5.534	2.343	.019	0.255
R. Comprehension					
Updating (a4)	0.134	0.040	3.376	.001	0.369
Planning (a5)	-0.003	0.021	-0.123	.902	-0.013
Inhibition (a6)	3.904	4.168	0.937	.349	0.102
Indirect effects					
Updating -> Intel. -> CRT	0.001	0.001	0.951	.342	0.027
Planning -> Intel. -> CRT	-0.000	0.000	-0.530	.596	-0.008
Inhibition -> Intel. -> CRT	0.162	0.164	0.985	.325	0.032
Updating -> Reading -> CRT	0.001	0.002	0.643	.520	0.028
Planning -> Reading -> CRT	-0.000	0.000	-0.121	.904	-0.001
Inhibition -> Reading -> CRT	0.039	0.073	0.537	.592	0.008
Total Effects					
Updating -> CRT (c1)	0.000	0.006	0.079	.937	0.009
Planning -> CRT (c2)	-0.006	0.003	-1.945	.052	-0.220
Inhibition -> CRT (c3)	0.278	0.582	0.478	.633	0.055

Hypothesis 5: Regression analyses for the two measures of academic achievement.

Finally, two regressions analyses were carried out in order to check the contribution of both fluid intelligence, reading comprehension and two arithmetic problems (AWP and CRT) on the explaining the academic performance in Mathematics (COMPEMAT) and Global Academic scores (see tables 11 and 12). First regression analysis showed that independent variables explained 45% of variance of Mathematics acquired abilities (COMPEMAT), being significant the effects of fluid intelligence and marginally AWP.

Table 11. Regression analysis of Fluid Intelligence, Reading Comprehension, AWP and CRT tasks on academic performance on Mathematics (COMPEMAT).

Dependent Variable	Predictors	*R2*	*R2adj*	*F*	*B*	*Beta*	*t*	*p*
COMPEMAT		.48	.45	17.848***				
	1. Fluid Intelligence				0.33	0.49	5.00	.000
	2. Reading Comp.				0.11	0.12	1.31	.196
	3. AWP-Global				1.99	0.18	1.94	.056
	4. CRT-Global				2.86	0.14	1.68	.097

*** p < .001

Second regression analysis showed that independent variables explained 15% of variance, being significant only the effect of AWP.

Table 12. Regression analysis of Fluid Intelligence, Reading Comprehension, AWP and CRT tasks on Global academic scores.

Dependent Variable	Predictors	*R2*	*R2adj*	*F*	*B*	*Beta*	*t*	*p*
Global Academic Score		.19	.15	4.62***				
	1. Fluid Intelligence				0.02	0.08	0.66	.514
	2. Reading Comp.				0.07	0.20	1.68	.097
	3. AWP-Global				1.01	0.23	2.00	.049
	4. CRT-Global				1.21	0.15	1.44	.153

*** p < .001

3.3.3.3. Discussion

First goal to be achieved in this study was to confirm the results obtained in experiment 1 on the accuracy response in both mathematical tasks, AWP and CRT. Results confirmed that AWP obtained a significant higher number of correct responses than CRT, as well as a significant higher number of superficial responses for CRT, similar to those obtained in Experiment 1. As in previous studies (Duque de Blas et al., 2021; Hegarty et al., 1992, 1995; Passolunghi et al., 2021) and in the experiment 1, consistent problems obtained a higher accuracy than inconsistent problems. We also found differences in terms of the number of operations that problems require. However, a more detailed analysis identified that these differences in accuracy occurred when the problems were consistent. In inconsistent problems, there were no significant differences due to the number of operations, since the main difficulty of these problems came from determining whether or not the arithmetic operation required to be solved was consistent with the operation(s) suggested by the key words inserted in the problem statement. In terms of accuracy, this experiment proved that solving the consistency or inconsistency of the keywords inserted in the problems is harder ($\eta^2 p$ = .51) than solving a larger number of operations ($\eta^2 p$ = .29).

On the other hand, the results of experiment 2 obtained in CRT confirmed the low-level effectiveness of the participants in solving these problems observed in experiment 1 and the high-level of superficial responses. Response accuracy was, again, lower than that obtained in other studies (Gómez-Chacón et al., 2014, 2011, Gómez-Veiga et al., 2018). Participants were unable to resist the prepotent response induced by the problems. This lack of cognitive disengagement may be due to several reasons, most importantly due to age and school level of our participants as well as their familiarity with mathematics.

Hypothesis 2 aimed to relate response accuracy in AWP and CRT to participants' awareness of the degree of task complexity. The results showed that there was no correlation between response accuracy and awareness of task difficulty. Secondary school students show a marked positive bias towards underestimating the difficulty of the kind of problems. However, the underestimation was higher in CRT than AWP considering the low proportion of correct responses collected. The AWP test was only completed correctly in just over 50% of the problems, which is very low considering that most of the students estimated the global task as "quite easy" (2.47 in the Likert scale). Therefore, there is no match between the participant's awareness of difficulty and accuracy on the task. However, we did find that the assessment of the task difficulty seems to increase with the age of the students. The higher school grade, the more aware the participants are of the difficulty of the problems, which would constitute a sign of metacognitive developmental. The case of CRT problems is different, they are extremely difficult, but participants consider them “easy” (3.41 in the Likert scale). This acute underestimation of problems difficulty confirms one of the main features of CRT.

Another difference between the problems has to do with the direction of the correlations between accuracy and perception of difficulty. Although the correlations are not reliable, the direction in AWP is negative (the easier AWPs are, the less difficulty is perceived by participants), whereas in CRT is positive (the easier CRT problems are, the more difficulty is perceived by participants). This result should be interpreted with caution due to the marked ground effect obtained in CRT. CRT is clearly more difficult task than AWP, and it is only partially solved by very few participants, so the positive direction between response accuracy and perceived difficulty should be interpreted as further evidence of the greater difficulty of their problems. However, the results obtained in CRT regarding perceived difficulty and the direction of the correlations between perceived difficulty and response accuracy invite us to wonder whether, along the lines supported by other authors (De Neys, 2012; De Neys et al., 2013), participants were somehow aware that their response to CRT problems was clearly erroneous given the deceptive nature of these problems. Perhaps their answers could be understood as an approximation with a semblance of certainty that arises as a consequence of not having an appropriate resolution model for the problem. We do not have data to support this hypothesis, so one line of future research could be to include a measure of how confident participants are that their answer is correct. This measure, also utilised by De Neys et al. (2013) might be more useful than the measure of perceived difficulty.

Hypothesis 3 aimed to correlate both problems with each other and, also, with the cognitive and performance measures. Inconsistent AWP obtained a similar marginal correlation with CRT than in experiment 1. But, in global terms, we could confirm positive correlations between AWP-Global and CRT. Again, this lack of significance has to be explained as a deleterious effect resulting from the poor performance on CRT.

As for the AWP, it showed a moderate correlation (r = .39) with the visuospatial reasoning-based measure of fluid intelligence. This result is similar to that found by Peng et al. (2019), with intelligence and word problems (between r = .40 and r = .45). On the other hand, AWP was significantly associated to the reading comprehension measure, showing the link between the generation of the situational model of the problem and the resolution, but somewhat lower than in experiment 1, or previous studies (Duque de Blas et al., 2021; Passolunghi et al., 2021).

Regarding the associations between AWP and EF, our results in experiment 2 also confirm the correlations of the taken measures on the AWP solving performance. In particular, AWP showed a moderate correlation with inhibition, confirming the crucial importance of the resistance to prepotent responses in the problem-solving process, previously reported by other researchers (Lubin et al., 2013, 2016; Passolunghi et al., 2021). As for planning, the results confirm the association between problem-solving sequence scheduling and the accuracy of response. AWP only showed a low correlation with the updating task.

An interesting question was to explain how much the different measures taken contribute to explaining the differences between consistent and inconsistent problems. Results confirmed that these differences between consistent and inconsistent problems were neutralised when controlling for participants' intelligence and inhibition. Consequently, inconsistent AWP would require a reflective process towards the inhibition of the heuristic strategy in favour of the inverse operation suggested by the verbal term (Jiang et al., 2020; Lubin et al., 2013, 2016; Shum & Chan, 2020). Both variables, inhibition and fluid intelligence has traditionally been linked to the mathematical performance (Friso-van den Bos et al., 2013; Passolunghi et al., 2021).

In CRT task, the results showed a low correlations profile with the measures. Only intelligence and planning showed a low and significant correlation. Consequently, the conclusions to be drawn from them are limited, since this ability to overcome the superficial response as a consequence of a reflective approach to the task does not seem to be present at these ages in most of participants of this sample, nor in the sample of experiment 1. In addition, it is noteworthy that the planning measure was the one most highly associated with CRT problems, which is consistent with the requirement for reorganisation in these problems to break the solver's functional fixedness (Mayer & Hegarty, 1996).

Hypothesis 4 aimed to figure out the direct and indirect effects that cognitive measures have on AWP and CRT. Complementary to correlation analyses, path analyses revealed that planning the sequence of solving in terms of goal and subgoals, the inhibition of partial or superficial responses and the fluid intelligence were the most significant abilities in explaining the performance in AWP solving. Solving problems requires skills that are not exclusively related to mathematics, but also another type of skills related to building a situational model of the problem that allow them to understand in depth the relationships between the problem's elements. In other words, solving AWP is a matter of understanding applied to a numeracy context, which requires not only reading comprehension or mathematical skills, but the joint action of some particular cognitive processes oriented towards a specific goal.

In contrast, the CRT path analysis was not very informative due to the low performance on this task. None of the measures had a significant effect on the explanation of the CRT scores, so the conclusion can only be that the ability to overcome the superficial response as a consequence of the absence of a reflective approach to the task does not seem to be present at these ages in most of participants of this sample.

Finally, hypothesis 5 checked the explanatory power of intelligence, reading comprehension, AWP and CRT on two academic performance measures. Regression results differed according to the type of measure predicted. With the measure of mathematical knowledge (COMPEMAT), intelligence was the most explanatory factor for performance, as other studies have shown (Fuchs et al., 2006; Peng et al., 2019). The contribution of AWP was marginally significant, confirming its validity as a measure of explanatory performance. CRT also contributed positively to the prediction of results, although in a low and marginally significant way due to the low performance obtained with this sample.

In contrast, with the overall mean score at the end of the course as dependent variable, only the AWP measure was significant. Reading Comprehension was not significant, probably because part of its effect was collected by AWP task. These data with two different measures of performance are of interest. When the factor to be predicted is only the student's mathematical knowledge, fluid intelligence stands out as the most relevant general factor in explaining these problems, followed by AWP. However, when it comes to predicting the overall grade given by the teacher, which is a joint measure of student knowledge, classroom behaviour, collaboration with peers and motivation, the effect of intelligence factor is moderated and is less explicative than the AWP task (see Roth et al., 2015). These results would be similar to those obtained by Duque de Blas et al. (2021) in which the AWP measure showed a higher capacity for association than intelligence when they had to explain an academic measure in Mathematics and History.

3.3.4. General conclusions

Solving arithmetic word problems is one of the curricular competencies that students must acquire during their basic and secondary education. Its mastery is relevant in academic and experimental settings, because they require the use a diverse set of cognitive skills and knowledge that, directed towards a goal, can carry out the resolution process.

The main goal of this study was to compare, in terms of accuracy and cognitive measures associated, two types of arithmetic problems, AWP and CRT. We started from the idea that inconsistent AWP are more closely related to CRT problems because both require the action of an analytical system of thought (Evans & Stanovich, 2013; Frederick, 2005; Kahneman, 2011; Toplak et al., 2014) capable of inhibiting the prepotent intuitive response and activate an alternative resolution sequence adjusted to the characteristics of the problem.

The results we obtained could not be conclusive due to the low CRT performance in both experiments. The relationship between both tasks was low, and only marginally significant with inconsistent AWP in both experiments. The solving of AWP inconsistent and CRT problems are different tasks despite some common features shared. For example, the low number of correct answers in CRT seems to confirm that the disposition towards reflective behaviour of the participants and to overcome their tendency towards cognitive miserliness as a coping mechanism towards the task, does not seem to be a fully acquired characteristic by most of them. It seems to be only a differential ability partially developed by a few of them. Participants persist in the idea that these problems can be solved on the basis of a previously acquired solving scheme, ignoring the fact that these problems probably require an incubation process that allows the *reorganisation* of the information, characteristic of a *productive thinking*, in order to generate an alternative response capable of address the participant towards a strategy far away from the mere calculation of magnitudes that suggests (Mayer, 1992). This belief that the problems are similar to those solved on previous occasions, would induce a "feeling of rightness" from which they would assess that the task does not require more cognitive effort than they make, disabling the possibility of the analytical thinking system to act (Thompson, 2010).

Analysing the most frequent type of response in the problems, we find another differential aspect between these two tasks: the most frequent response they generate is different. Even in the inconsistent AWPs, the most common response was the correct one, while CRT problems induce an overwhelmingly superficial response. This feature has to do with the fact that, unlike AWPs, CRT problems are intentionally misleading in their structure and content, and require the solver to decouple the direction to which he or she is forcibly led by the statement in order to resist the induced superficial response (Evans & Stanovich, 2013a). CRT requires the participant to consider a different organisation of the data and therefore an incubation period detached from the structure in which it is embedded that allows the response to emerge. This reorganisation might only be considered necessary if the solver believes that the intention of the problems was to induce an apparently correct, but superficial, answer. This tendency towards a superficial response, present in both problems, could be understood as a prelude to a later type of reflective, exhaustive and goal-directed behaviour.

In general terms, reasoning, inhibition and planning processes would be the most closely related abilities to AWP resolution. In particular, the differential effectiveness between consistent/inconsistent AWP seems to be related to an increase in the demand of reasoning (fluid intelligence) and inhibition resources. On the contrary, reasoning and planning seem to be associated to CRT solving.

Given the low performance on CRT problems, conclusions regarding the type of cognitive skills required are very limited and constitute the main limitation of this study. We do not consider, therefore, that this measure should be used for the prediction of academic performance in secondary school students, because even among expert adults this measure has traditionally been associated with a low level of success. On the contrary, AWP stands out as a task with a significant capacity for explaining academic performance in a variety of subjects and global measures where other measures were not efficient enough.

This study highlights the relevance of basic cognitive processes in explaining success in problem solving, which constitutes a starting point for the design of programmes aimed at reducing school failure by intervening in these processes. These programmes could help reduce the lack of mathematical competence that can occur in students and promote STEM careers.

4. DISCUSIÓN GENERAL Y CONCLUSIONES

El objetivo de esta tesis fue profundizar en el conocimiento sobre los procesos cognitivos y metacognitivos implicados en la resolución de problemas aritméticos verbales (AWP) en la adolescencia. A lo largo de este trabajo hemos expuesto las razones que nos llevan a pensar que los AWP pueden desempeñar un papel más extenso que el tradicionalmente asignado al ámbito de la educación y la adquisición de unas habilidades específicas en el ámbito de las Matemáticas. Los AWP pueden ser utilizados, además, como una herramienta relevante desde el punto de vista experimental en varios sentidos. Por un lado, porque permiten estudiar un conjunto extenso de habilidades generales y procesos ejecutivos de la memoria operativa subyacentes a la tarea mediante la manipulación del contenido del enunciado de los problemas. Por otro, porque dada la diversidad de habilidades intelectuales necesarias para resolver correctamente los problemas, la eficacia en la resolución en esta tarea puede representar un indicador predictivo del rendimiento académico general de los estudiantes.

La resolución de AWP no es un simple acto reproductivo consistente en la aplicación mecánica de unos procedimientos aritméticos concretos, tal y como algunos estudiantes consideran de modo reduccionista, sino que debe entenderse de manera más extensa como un problema de comprensión aplicado al ámbito de las Matemáticas. Este complejo proceso constituye una destreza esencial en cualquier ámbito de la vida para la cual se requiere la acción coordinada de distintas habilidades que, en conjunto, permiten al participante (1) entender el propósito e intención planteado en el problema; (2) determinar las relaciones significativas, y su valencia, que se establecen en el enunciado; (3) establecer una secuencia de tipo procedimental de naturaleza aritmética; (4) ejecutar el algoritmo de cálculo bajo una supervisión consciente de todo el proceso. La resolución de esta tarea requiere que la persona construya un modelo situacional del problema basado en sus aspectos semánticos nucleares. Este proceso le permitirá establecer de manera precisa las distintas relaciones entre los elementos del problema que se describen y, con ello, determinar las operaciones necesarias para hallar la respuesta solicitada. El modelo situacional del problema es construido en la mente de la persona, no como un "mero" ejercicio de extracción de conocimientos, sino como un paso intermedio orientado a un fin instrumental; esto es, la aplicación de una secuencia de cálculo que permita dar respuesta a la pregunta planteada. Esta actividad mental dirigida hacia un propósito concreto constituiría la definición básica de un proceso de resolución de problemas (Johnson-Laird, 1988).

Como hemos señalado, en el proceso de construcción del significado es necesario determinar la relevancia de la información que se presenta en el enunciado y el sentido de las relaciones que se establecen entre los sujetos y objetos del problema, ya que a partir de ellas se deducen las operaciones aritméticas necesarias que, posteriormente, tendrán que computarse. Este proceso de extracción de información no explícita en el enunciado del problema hace referencia a otra de las habilidades generales necesarias para llevar a cabo la tarea: las habilidades de razonamiento. Así, la persona que resuelve podría considerar que la resolución al problema puede alcanzarse a partir de un procedimiento de tipo heurístico consistente en identificar las palabras clave insertadas en el enunciado del problema y traducir su sentido literal como operadores aritméticos de la secuencia de cálculo. Estos adverbios de cantidad, "más" o "menos", sugieren una operación de cálculo específica de suma o resta, respectivamente. Llevar a cabo este tipo de *estrategia directa* de resolución resultará adecuado siempre que, por el contexto en que se encuentran insertadas las palabras clave, exista una correspondencia directa entre la operación aritmética que sugiere el término verbal y la necesaria para la resolución del problema (más-suma; menos-resta; Schumacher y Fuchs, 2012). En caso contrario, representará una de las principales causas por las que un alumno normotípico fracasa en el proceso de resolución de AWP.

Además del *tipo de relaciones* entre los elementos, otro de los aspectos relevantes que ha sido objeto de estudio en este trabajo sobre la resolución de AWP es el *número de relaciones* que se describen entre los elementos del problema. Esta característica resulta de especial relevancia, ya que es uno de los principales factores causales del incremento en la dificultad de resolución de los problemas. Ello es debido a que un número mayor de relaciones significativas descritas en el enunciado conlleva un mayor número de proposiciones que deben integrarse en el modelo situacional del problema, así como de operaciones necesarias que integran la secuencia de cálculo, lo cual tiene un efecto incremental de recursos cognitivos necesarios para la resolución.

Con esta perspectiva, diseñamos tres experimentos con varios propósitos. El primer estudio tenía por objetivo determinar la relación existente entre la eficacia en la resolución de AWP y la competencia de los estudiantes de 2º y 3º curso de Educación Secundaria Obligatoria (13-14 años) en dos habilidades cognitivas generales: comprensión lectora y razonamiento. Asimismo, comprobamos las diferencias en la eficacia con la que se resuelven los distintos tipos de problemas (de 1 o 2 operaciones consistentes y/o inconsistentes) y el tipo de respuesta más frecuente que se produce (correcta, incorrecta y superficial). Los resultados obtenidos fueron los siguientes.

En relación al número de operaciones de los problemas, los resultados indicaron que este factor influye significativamente en la eficacia de resolución de AWP (Castro-Martínez y Frías-Zorilla, 2013; Hegarty et al., 1992; Mayer et al., 1996). El incremento del número de pasos u operaciones requeridos para la resolución del problema implica un mayor número de proposiciones contenidas en la base del texto y un modelo situacional más complejo que, en conjunto, añaden dificultad al proceso. El proceso de comprensión de los problemas de dos operaciones es, por tanto, más complejo que los de una sola operación consistente, debido a una mayor demanda de recursos de memoria y los dedicados a la actualización e integración de los nuevos contenidos en la representación situacional del problema (Bull & Lee, 2014; Friso-van den Bos et al., 2013; Passolunghi & Pazzaglia, 2005). Sin embargo, este factor sólo es determinante en los problemas consistentes. En presencia de inconsistencias, las diferencias en eficacia entre los problemas de 1 y 2 operaciones desaparecen.

Este hallazgo resalta la idea de que la principal fuente de dificultad en los AWP recae en la forma en que los participantes interpretan las palabras clave que relacionan los distintos elementos del problema. Así, los problemas inconsistentes serán significativamente más difíciles de resolver que los problemas consistentes por la necesidad de inhibir la operación sugerida por la interpretación literal del adverbio de cantidad y sustituirla por la operación inversa, lo que requiere que el sujeto sea capaz de inhibir una estrategia sobreaprendida de respuesta que le conduciría a la toma de decisiones incorrectas como consecuencia de haber construido un modelo superficial del problema (Lubin et al., 2013, 2016). Además, los problemas inconsistentes probablemente requieren que la persona que resuelve lea repetidas veces el enunciado, lo que implica dedicar más tiempo y recursos mentales destinados a integrar las proposiciones que relacionan los elementos del texto e inhibir los modelos representacionales iniciales que surgen a partir de una lectura superficial de los enunciados. Estos resultados confirman la importancia de los procesos de actualización e inhibición de la respuesta intuitiva, automática y superficial que se genera por la interpretación literal de los enunciados, y que conduciría a la construcción de un modelo mental equivocado del problema (Evans & Stanovich, 2013; Fletcher & Carruthers, 2012; García-Madruga et al., 2007; Shum & Chan, 2020; Stanovich & Toplak, 2012).

Con relación a los errores producidos durante la resolución de AWP, pudo determinarse que los más comunes no son los errores superficiales, definidos como el resultado que se obtiene por la ejecución correcta de un plan equivocado de resolución, sino también otro tipo de respuestas erróneas no-superficiales. Estos errores constituyen un conjunto variado de elementos, entre los que se encontrarían los errores de cálculo, los causados por la dificultad para comprender el enunciado o el objetivo del problema planteado; los causados por un déficit a la hora de actualizar los contenidos en la base del texto; o aquellos que, por razones de tipo motivacional, hayan causado una baja adhesión de los participantes a la tarea. En relación a las respuestas superficiales obtenidas, la frecuencia fue baja y se mantuvo estable a lo largo de ambos cursos.

Este estudio permitió también confirmar la relación predicha entre las habilidades generales de los participantes y la eficacia en la resolución de estos problemas, llegando a explicar un 20%

de la varianza de resultados. Así, los problemas más difíciles de resolver (problemas de dos operaciones y problemas inconsistentes) obtuvieron una correlación más elevada con las habilidades de comprensión y razonamiento, lo que evidencia la relación entre la eficacia en la tarea y la habilidad para razonar reflexivamente durante la misma, un proceso característico del Sistema 2.

Por último, tal y como preveíamos, dada la diversidad de habilidades cognitivas implicadas en la resolución de AWP, logramos demostrar que estos problemas tenían potencial significativo para explicar no sólo las puntuaciones en rendimiento académico de los estudiantes en Matemáticas, sino también en un área de conocimiento muy diferente como es la Historia/Geografía.

En relación al segundo estudio, fue realizado con el fin de comprobar la eficacia en la resolución de AWP de una muestra de estudiantes de 4º y 5º curso de Educación Primaria de Italia (10-11 años) y su relación con cuatro habilidades cognitivas: comprensión lectora como variable mediadora de la respuesta, la inteligencia fluida, la inhibición y la actualización. Además, estudiamos como afectaba la localización de la inconsistencia en el enunciado del problema a la eficacia con la que los participantes resolvían los problemas de dos operaciones con un solo término verbal inconsistente, una aportación novedosa de este estudio.

Del mismo modo que se produjo en el estudio anterior, los participantes fueron más eficientes a la hora de resolver AWPs consistentes que inconsistentes. La presencia de términos verbales inconsistentes en los AWP incrementó su dificultad y cambió el tipo de estrategia necesaria para su resolución, corroborando los hallazgos obtenidos por otros estudios precedentes (por ejemplo, Boonen et al., 2016; De Corte et al., 1985; Pape, 2003). El número creciente de operaciones fue también un factor que contribuyó significativamente al incremento de la dificultad de los problemas.

En relación al patrón de correlaciones obtenido, los resultados evidenciaron el papel crucial de las habilidades relacionadas con la comprensión en la resolución de AWP (Kintsch, 1998; Peng et al., 2019; Thevenot, 2010), obteniendo unas correlaciones más elevadas entre la tarea de comprensión y los AWP inconsistentes, que con los consistentes. Así, estos resultados sugerirían que las diferencias observadas en la precisión de respuesta en AWP consistentes frente a los inconsistentes podrían no estar tan relacionadas con las habilidades aritméticas de los estudiantes, sino con sus habilidades para la construcción del modelo situacional del problema (véase Fuson, 1992; Hasanah et al., 2017; Thevenot y Barrouillet, 2015).

Respecto a las diferencias en la eficacia de resolución de AWP que se producen debidas al orden en que se encuentra la inconsistencia en el enunciado, los resultados indicaron que AWP C-I eran más difíciles de resolver que los AWP I-C. Por otro lado, se encontraron correlaciones más altas entre los problemas AWP C-I y la medida de actualización, que con los problemas AWP I-C, lo que nos permitió confirmar que la diferencia en la eficacia de resolución entre ambos tipos de problemas quedaba neutralizada cuando se controlaban las habilidades de actualización de los participantes. No obstante, estos resultados deben tomarse con cierta cautela, dado que el orden de presentación de los problemas fue siempre el mismo. Es por ello que, una cuestión a abordar en futuros estudios consistirá en determinar si las diferencias debidas a la posición en que se localiza la inconsistencia persisten con un orden de presentación distinto al utilizado en este estudio.

Los análisis de senderos permitieron identificar, además, varios aspectos relevantes relacionados con la resolución de AWP; en particular, con la significativa y moderada contribución que tuvo la inteligencia fluida en la explicación del rendimiento global obtenido en la tarea. Este resultado evidencia la contribución de unas habilidades relacionas con los procesos de razonamiento inductivo, deductivo y cuantitativo necesarios para llevar a cabo de manera exitosa esta tarea (Sternberg & Ben-Zeev, 1996). Las habilidades de inhibición y actualización contribuyeron a explicar los resultados globales obtenidos, aunque en menor grado que la inteligencia fluida. En conjunto, los análisis de senderos presentados en este estudio son otra aportación novedosa al campo de estudio en los que se destaca principalmente el papel crucial que juega la comprensión lectora en la resolución de AWP.

Finalmente, el tercer estudio llevado a cabo tenía por objetivo comparar y estudiar el papel que desempeñan las habilidades cognitivas generales y los procesos ejecutivos de la MO en la resolución de dos tipos de problemas aritméticos: los problemas aritméticos verbales (AWP) y los problemas de reflexión cognitiva (CRT; Frederick, 2005). Ambos tipos de problemas han sido relacionados con una capacidad predictiva del rendimiento académico (ver Gómez-Chacón et al., 2014, 2011). Asimismo, tanto los AWP inconsistentes como los CRT son dos tipos de problemas que guardan algunas semejanzas entre sí, destacando la necesidad de llevar a cabo un proceso controlado de razonamiento que permita elaborar un modelo situacional del problema descartando aquellas respuestas superficiales automáticamente generadas por la forma o contenido de los problemas. Para tal propósito, se llevaron a cabo dos experimentos. En el primer experimento, se comparó la eficacia en la resolución de ambos tipos de problemas matemáticos (AWP y CRT) en una muestra de 81 participantes de 4º de ESO (15 años), explorando la relación existente entre la correcta ejecución de la tarea y dos tipos de habilidades cognitivas: la inteligencia fluida y la comprensión lectora. Además, se analizó el tipo de respuesta predominante a cada tipo de problema (respuesta correcta, respuesta incorrecta y respuesta superficial) con el fin de establecer un patrón normativo de respuestas característico.

Los resultados obtenidos en este experimento indicaron que los participantes fueron notablemente más eficaces en resolver los problemas AWP inconsistentes que los problemas CRT. De modo similar al resto de estudios, los problemas AWP consistentes fueron más fáciles de resolver que los AWP inconsistentes. Asimismo, el número de operaciones también resultó un factor significativo en el incremento de la dificultad de los problemas, aunque solamente en los

AWP consistentes, tal y como ocurrió en el estudio 1. No se encontraron, por tanto, diferencias en la eficacia en la tarea en problemas AWP inconsistentes relativas al número de operaciones.

En relación al CRT, la eficacia en la resolución obtenida en este experimento fue muy limitada (9%), mostrando un notable efecto suelo. Estos resultados fueron, incluso, inferiores a los obtenidos en otros estudios con participantes de un curso superior y de la rama de ciencias (Gómez-Chacón et al., 2014; Gómez-Veiga, et al., 2018).

Respecto al perfil respuesta característico en ambos problemas, encontramos un predominio mayoritario de respuestas superficiales en CRT (72%) frente a AWP inconsistentes (16%), mientras que la respuesta mayoritaria registrada en AWP fue la correcta (44%). Estas diferencias limitaron la posibilidad de encontrar correlaciones significativas entre ambos tipos de problemas, una hipótesis contemplada que no pudo confirmarse debido, principalmente, al bajo rendimiento obtenido en CRT. Los resultados obtenidos parecen sugerir que ambas pruebas difieren en la cantidad de recursos o el uso extensivo de unas habilidades necesarias para llevar a cabo la tarea de manera eficaz. Mientras que en los problemas AWP inconsistentes pueden ser resueltos por una mayoría de participantes, los problemas CRT parecen requerir un esfuerzo reflexivo mayor, habilidad que parece no haber sido aun completamente adquirida por los participantes de esta muestra. Ambos tipos de problemas obtuvieron, como se preveía, correlaciones significativas con las medidas de inteligencia fluida y comprensión lectora. La inteligencia fluida correlacionó de manera más elevada y esperable con los AWP inconsistentes que con los consistentes. Asimismo, tanto los AWP inconsistentes como los CRT correlacionaron de manera similar con la inteligencia fluida. En cambio, los AWP inconsistentes y la comprensión lectora correlacionaron entre sí más que los problemas CRT con la medida de comprensión. Este aspecto parece denotar que la eficacia en la resolución de AWP está más relacionada con los procesos de generación de un modelo situacional del problema, dada su estructura narrativa, que con los CRT, aunque no es posible establecer conclusiones definitivas al respecto debido al bajo rendimiento obtenido en CRT.

El segundo experimento se llevó a cabo con el fin de comprobar 4 objetivos concretos. En primer lugar, confirmar los resultados obtenidos en el experimento anterior con relación a la precisión de las respuestas obtenidas en AWP y CRT en una muestra de 82 alumnos de 3º y 4º de secundaria. En segundo lugar, relacionar la precisión de las respuestas en los problemas con la experiencia metacognitiva de dificultad que experimentan los participantes durante el proceso de resolución. En tercer lugar, y más importante, explorar las relaciones entre la precisión de las respuestas dadas en estos problemas y dos tipos de medidas cognitivas: las habilidades generales anteriormente utilizadas (inteligencia fluida y comprensión) y algunas medidas de funcionamiento ejecutivo (inhibición, actualización y planificación). Por último, comprobar la capacidad de ambos tipos de problemas aritméticos, AWP y CRT, junto con las variables cognitivas, para explicar el rendimiento académico del estudiante de secundaria.

En relación al primer objetivo planteado, los resultados confirmaron las diferencias en la eficacia de resolución entre ambos problemas. El AWP obtuvo un número significativamente mayor de respuestas correctas que el CRT, mientras que el CRT presentó un perfil caracterizado por una elevada cantidad de respuestas superficiales, similar al obtenido en el experimento 1. Respecto al AWP, tanto el tipo de relaciones que se establecen (consistentes o inconsistentes) entre los elementos del enunciado de los problemas, como el número de operaciones que requiere la resolución, fueron factores significativos que incrementaron la dificultad de los problemas, reduciendo la eficacia con que fueron resueltos. La consistencia/inconsistencia de los problemas tuvo un efecto significativamente mayor en la reducción de esta eficacia que el incremento del número de operaciones necesarias de la secuencia de cálculo. Respecto al CRT, los resultados obtenidos confirmaron el bajo nivel de eficacia de los participantes en la resolución de estos problemas, quienes se mostraron incapaces de resistir la respuesta prepotente inducida por el enunciado del problema. La precisión de las respuestas fue, de nuevo, inferior a la obtenida en otros estudios precedentes (Gómez-Chacón et al., 2014, 2011, Gómez-Veiga et al., 2018), confirmando así los resultados obtenidos en el experimento 1. Esta falta de habilidad de los participantes para desacoplarse cognitivamente de la tendencia de respuesta superficial podría tener que ver no tanto con la edad de los participantes, sino especialmente con su nivel escolar y su familiaridad con las Matemáticas.

En relación al segundo objetivo de este experimento, los resultados confirmaron la inexistencia de correlaciones entre la eficacia de resolución de AWP o CRT, y la consciencia de dificultad que implicaba la tarea. Los participantes mostraron un marcado sesgo hacia la subestimación de la dificultad de ambas tareas, siendo mayor en CRT que en AWP considerando la baja eficacia de resolución alcanzada en la tarea. No obstante, la valoración de la dificultad de las tareas parece aumentar con la edad de los alumnos, de modo que los participantes son más conscientes de la dificultad de los problemas en 4º curso que en 3º. Este hecho constituiría un signo de desarrollo metacognitivo de los participantes. No obstante, las conclusiones que pueden extraerse a partir de la información proporcionada por esta medida metacognitiva de percepción de la dificultad son muy limitadas, particularmente en la prueba CRT, ya que no es posible establecer si los participantes fueron de algún modo conscientes de que su respuesta era errónea (ver De Neys, 2012; De Neys et al., 2013). Cabe la posibilidad de que los resultados obtenidos puedan ser debido a que, ante la incapacidad de encontrar un esquema de resolución apropiado para los problemas CRT, los participantes respondieran con aquella alternativa de respuesta que les proporcionaba una sensación de certeza mayor (Thompson, 2010); esto es, la respuesta superficial, por lo que sería necesario en futuros estudios utilizar otro tipo de medida más descriptiva de los procesos metacognitivos de los participantes.

En relación al tercer objetivo de este experimento, y de modo similar al experimento 1, no se pudieron encontrar correlaciones significativas entre AWP-inconsistentes y CRT, debido al bajo rendimiento obtenido, nuevamente, en la prueba CRT. Respecto a la resolución de AWP, tanto las correlaciones calculadas, como el análisis de senderos, indicaron que la planificación de la secuencia de resolución en términos de metas y submetas, la inhibición de respuestas parciales o superficiales y la inteligencia fluida fueron las habilidades más explicativas de la eficacia en la resolución de AWP. La resolución de problemas requiere habilidades que no están relacionadas exclusivamente con las Matemáticas, sino también con otro tipo de habilidades que posibilitan la construcción de modelo situacional del problema tras la comprensión de las interrelaciones entre los elementos del problema. En otras palabras, la resolución de un AWP constituiría un ejercicio de comprensión aplicado a un contexto matemático y orientado a un objetivo instrumental específico.

Otro aspecto reseñable fue la comprobación de que las diferencias en la eficacia con que se resuelven los AWP consistentes y AWP inconsistentes quedaban neutralizadas cuando se controlaba la inteligencia e inhibición de los participantes, medidas tradicionalmente relacionadas con el rendimiento matemático (Friso-van den Bos et al., 2013; Passolunghi et al., 2021). Este resultado confirmaría que la eficacia en la resolución de AWP inconsistentes estaría relacionada con la necesidad de que el participante reflexione en profundidad acerca de las relaciones que se establecen entre los distintos elementos del problema, incrementando así la posibilidad de inhibir la estrategia heurística de respuesta durante la fase de resolución (Jiang et al., 2020; Lubin et al., 2013, 2016; Shum & Chan, 2020).

Estos resultados nos llevan a considerar algunas cuestiones teóricas y evolutivas que aluden al desarrollo del comportamiento estratégico de los participantes, los esquemas de funcionamiento que activarían en su intento por resolver la tarea y sus habilidades cognitivas. En relación al comportamiento estratégico y los esquemas que se activarían durante la tarea, posiblemente una de las características que diferenciaría a una persona altamente competente en resolución de problemas frente a otra que no lo es tiene que ver, no sólo con el hecho de que posea un mayor número de estrategias de resolución, sino también con que sea capaz de adecuar el tipo de estrategia para cada tipo de problema específico al que se enfrente. Si la valoración de la adecuación del tipo de estrategia no tuviera lugar, el participante consideraría suficiente el esfuerzo cognitivo que realiza y adoptaría una estrategia particular de respuesta seleccionada a partir de criterios superficiales. Estos criterios de selección estratégica de respuesta estarían probablemente fundamentados en base a las experiencias previas adquiridas, por lo que se seleccionaría aquella estrategia cuya eficacia ha sido previamente comprobada en la resolución de un problema con apariencia similar; pero también podría seleccionarse como estrategia de resolución aquella que resulte más fácilmente accesible en la memoria, lo que conduciría hacia una respuesta probablemente ineficaz. Tales casos podrían suceder, por ejemplo, en la resolución de problemas de tipo inconsistente, en los que hay una tendencia a aplicar de manera irreflexiva un esquema distinto al requerido por la situación a la que se enfrenta. En caso contrario, esto es, que la persona a cargo de la resolución del problema valore que éste no se puede resolverse a partir de una estrategia particular disponible, se activaría un proceso analítico que llevaría al participante a la elaboración de un modelo situacional del problema en el que quedasen representadas de manera correcta las relaciones particulares que se establecen entre los elementos descritos en el enunciado. A partir de esa representación semántica del problema, el participante podría deducir las operaciones aritméticas necesarias para su resolución y aplicar el correspondiente cálculo de las magnitudes.

Este aspecto entronca directamente con los distintos modelos teóricos de resolución de problemas aritméticos verbales. Si bien las teorías basadas en los esquemas son apropiadas para explicar parcialmente algunos resultados que se obtienen durante el proceso de resolución de los problemas inconsistentes, no parecen suficientemente satisfactorias a la hora de explicar cuándo un participante sí es capaz de solucionar de manera adecuada un problema de tipo inconsistente, ya que ningún participante, por experto que sea, puede definir un esquema de respuesta característico para cada una de las indeterminadas situaciones en las que pudiere verse envuelto. La resolución a este de tipo de problemas tiene que ver, por tanto, con otro tipo de habilidades que conjuntamente dotan al individuo de la flexibilidad para adecuarse a las distintas condiciones problemáticas en las que se ve envuelto y que permitirían activar un proceso consciente de búsqueda de alternativas de respuesta mediante la comprensión de la estructura profunda del problema, una habilidad que también se relaciona con la eficacia en la resolución de los problemas de reflexión cognitiva (CRT; Campitelli & Labollita, 2010). De este modo, la teoría del modelo situacional complementa y precisa aquellas condiciones para las cuales los modelos basados en esquemas no son suficientemente explicativos, ni del error, ni tampoco del acierto en la resolución de problemas.

Por otro lado, tanto la evaluación de la adecuación de la estrategia de solución, como la búsqueda de una solución de respuesta alternativa adaptada a las condiciones del problema, son tareas del proceso de resolución para las que, como hemos visto, es necesario disponer -entre otras- de unas habilidades relacionadas con la cancelación de la respuesta automática y la cognición fluida. Estas habilidades han sido tradicionalmente ligadas al razonamiento y son características del pensamiento formal, que aparece y se desarrolla durante la adolescencia; y que fundamentan la existencia de dos sistemas de razonamiento, uno rápido, automático y basado en heurísticos de respuesta, y otro lento, analítico y soportado por el uso extensivo que realiza de los recursos cognitivos que disponen la memoria de trabajo.

Respecto a la prueba CRT, dado el bajo rendimiento obtenido en esta tarea, se obtuvieron correlaciones bajas con las medidas cognitivas, encontrándose únicamente significación con la inteligencia y la planificación. Estos resultados, aunque limitados, resaltan la importancia de los procesos de razonamiento necesarios para llevar a cabo de manera exitosa la tarea, así como la necesidad de que los participantes sean capaces de reorganizar desde un punto de vista alternativo estos problemas con el fin de romper la fijeza funcional que les conduce a la comisión de errores superficiales en la tarea (Mayer et al., 1996). En este sentido, algunos autores como Campitelli y Gerrans (2014) identifican la prueba CRT como una medida no sólo de las habilidades generales del individuo, como la cognición matemática, sino también de las habilidades de razonamiento y de una disposición del pensamiento denominada *apertura mental activa*. Estas últimas son habilidades de desarrollo tardío que se iniciarían durante la adolescencia y proseguirían durante la edad adulta, un aspecto que explicaría parte de los limitados resultados aquí encontrados.

El último de los objetivos relativos a este experimento tuvo que ver con la capacidad predictiva de ambos tipos de problemas, AWP y CRT, y las medidas cognitivas generales sobre el rendimiento académico de los estudiantes en dos medidas muy distintas: una medida de conocimientos matemáticos (COMPEMAT) y una medida de rendimiento global del estudiante a final del año académico proporcionada por el centro de estudios. Respecto a la primera medida de rendimiento, COMPEMAT, la variable con mayor capacidad explicativa del rendimiento fue la inteligencia fluida. Los AWP fueron marginalmente significativos, así como los CRT, aunque de forma mucho más escasa debido al bajo rendimiento obtenido en esta muestra. Sin embargo, sólo la medida AWP resultó significativa en la explicación del rendimiento global del alumno a final del curso. Estos resultados resultan interesantes debido, principalmente, al distinto comportamiento de la predicción en función de la naturaleza de la medida de rendimiento utilizada. Cuando el factor a explicar fue únicamente el conocimiento matemático disponible del alumno, la inteligencia fluida destacó como el factor general principal a la hora de explicar el rendimiento, seguido de la medida de AWP. Sin embargo, cuando se trataron de explicar las calificaciones generales otorgadas por el profesor a final del curso, esto es, una medida conjunta de los conocimientos del alumno, el comportamiento en el aula, la colaboración con los compañeros y la motivación del estudiante, el efecto del factor inteligencia se redujo y resultó menos explicativo que la medida de AWP (véase Roth et al., 2015).

A la vista de los resultados obtenidos, podemos concluir que existen notables diferencias entre ambos tipos de problemas aritméticos, AWP y CRT. Estas diferencias tienen que ver con una mayor dificultad de los problemas CRT debido a que generan una tendencia de respuesta superficial en los participantes mucho más pronunciada ante la cual son incapaces de resistirse. Esta tendencia hacia la respuesta superficial de los participantes, presente en ambos problemas, podría entenderse como el preludio de un tipo de conducta posterior más reflexiva, exhaustiva y dirigida hacia un objetivo concreto. Sin embargo, la disposición hacia el comportamiento reflexivo de los participantes de este estudio como mecanismo de afrontamiento a la tarea no parece ser una característica totalmente adquirida por la mayoría de ellos, sino una capacidad diferencial parcialmente desarrollada sólo por unos pocos. La tendencia de los alumnos que participaron en este estudio es la de persistir en la idea de que los problemas CRT pueden ser resueltos a partir de un esquema de resolución previamente adquirido, ignorando el hecho de que estos problemas requieren un proceso de incubación que permita reorganizar la información descrita en el enunciado de un modo distinto a partir del cual pudiera hallarse un procedimiento de resolución alternativo (Mayer, 1992). Los problemas CRT son, probablemente, engañosos en su estructura y contenido, lo que requiere que el participante se desacople cognitivamente de la dirección a la que es conducido por los datos y estructura del enunciado (Evans y Stanovich, 2013). Asimismo, es probable que los participantes consideren que los problemas son similares a otros resueltos en ocasiones previas, lo que induciría una "sensación de certeza" a partir de la cual valorarían que la tarea no merece mayor esfuerzo cognitivo que el ya realizado, desactivando así la posibilidad de que actúe el Sistema 2 de razonamiento (Thompson, 2010). No obstante, no debe obviarse el hecho de que estos problemas son particularmente difíciles, incluso, entre participantes universitarios y expertos, encontrándose una baja proporción de aciertos. Es por ello que, a diferencia de los AWP, la utilización de esta medida para la predicción del rendimiento académico no resulta conveniente, al menos, entre población adolescente debido a la baja eficacia mostrada en la resolución.

A diferencia de los anteriores, los resultados obtenidos en el estudio 1 y 3 permiten destacar la significativa capacidad de los AWP como medidas explicativas del rendimiento académico genera, y del rendimiento particular en diversas asignaturas. Un aspecto relevante y frecuentemente obviado tiene que ver con que el rendimiento de un alumno depende de un conjunto diverso de habilidades y procesos que no se reducen, ni explican suficientemente, en base a la información proporcionada por las tradicionales medidas de inteligencia. Por el contrario, los AWP destacan por ser una tarea válida en lo ecológico, en tanto que su resolución constituye una competencia curricular básica que debe adquirirse desde la enseñanza primaria, y que además ofrecen la posibilidad de ser utilizados en el ámbito experimental como una medida rápida y eficaz del funcionamiento cognitivo global del individuo.

Los estudios contenidos en esta tesis doctoral permiten, aunque de manera modesta y limitada, dibujar un perfil característico del funcionamiento ejecutivo que subyace a la resolución de problemas aritméticos verbales (AWP) en población normotípica adolescente a lo largo de dos niveles escolares distintos: final de la Educación Primaria (EP) y Educación Secundaria Obligatoria (ESO). Como característica compartida en las distintas muestras seleccionadas en los experimentos, encontramos que tanto la comprensión lectora, como la cognición fluida son habilidades cognitivas generales estrechamente relacionadas con la eficacia de la resolución de estos problemas. En los alumnos más jóvenes (estudio 2) encontramos, además, que las habilidades de actualización tienen un peso mayor que la inhibición en la explicación de los resultados encontrados. De hecho, la actualización de contenidos en la MO es capaz de explicar las diferencias encontradas en los problemas de dos operaciones en los que varía la posición de su única inconsistencia contenida.

Conforme incrementa la edad de los participantes y su nivel escolar, algunas habilidades parecen perder algo de protagonismo en adolescentes normotípicos en favor de otras en cuanto a la explicación de la precisión de la respuesta de resolución. En términos generales, la comprensión lectora resulta menos determinante que el razonamiento fluido, habilidad a partir de la cual pueden inferirse las relaciones comparativas entre los objetos y magnitudes descritos en el enunciado, así como las operaciones aritméticas necesarias para hallar la respuesta solicitada (estudio 1 y 3). De modo parecido, las habilidades de actualización no parecen determinantes en la explicación de los resultados globales en AWP, quizá debido a que los problemas no demandan una cantidad elevada de estos recursos de acuerdo a este tramo etario. Por el contrario, otros procesos cognitivos como la planificación de la secuencia de resolución y, especialmente, la inhibición de respuestas parecen ser variables con una mayor capacidad para explicar los resultados obtenidos por los participantes en la resolución problemas aritméticos verbales.

Tomados en conjunto, estos resultados permiten mostrar las grandes diferencias de rendimiento que pueden encontrarse en relación a una competencia específica, como la resolución de AWP, dentro de un grupo normotípico de estudiantes adolescentes. Dicha diversidad en competencia puede estar también determinada parcialmente por las diferencias en el desarrollo de las habilidades metacognitivas de los estudiantes. Autores como van der Stel y Veenman (2014) establecen que las distintas habilidades metacognitivas que facilitan el proceso de aprendizaje en adolescentes de 12 a 15 años no se desarrollarían de manera lineal, ni siquiera al mismo ritmo, por lo que gran parte de la variabilidad entre los alumnos de un mismo nivel escolar, como la aquí encontrada, podría estar justificada, también, en base a estas consideraciones.

Todos estos aspectos contenidos en esta tesis doctoral permiten, en conjunto, incrementar el grado de comprensión acerca de los procesos cognitivos subyacentes a la resolución de problemas aritméticos y proporcionar algunas claves relevantes acerca del desarrollo de la competencia matemática en la adolescencia y la predicción del rendimiento académico del estudiante con los que mejorar la educación. Este estudio resalta la necesidad de que la enseñanza de las matemáticas en la resolución de problemas se centre no sólo en los procedimientos computacionales necesarios para el cálculo de las operaciones. Es esencial que la acción del estudiante esté dirigida por sus ideas y su capacidad para reflexionar a partir de las condiciones que estable la situación problemática (Mayer et al., 1992). Todo ello contribuirá a la reducción de la brecha de conocimientos de los alumnos españoles en las pruebas externas (PISA) y a promocionar los estudios relacionados con el área de las ciencias (STEM).

5. LIMITACIONES Y FUTURAS DIRECCIONES

Los resultados obtenidos en este estudio han permitido profundizar y contribuir al conocimiento acerca de los procesos cognitivos implicados en la resolución de problemas aritméticos verbales (AWP). La relevancia del objetivo propuesto, así como los interesantes resultados mostrados a lo largo de este trabajo, no dispensa de la necesidad de establecer algunas de las limitaciones más importantes que tiene este trabajo y de indicar algunas direcciones futuras en el estudio de la competencia matemática.

Una de las críticas que pudieren realizarse a esta tesis doctoral y que afecta, principalmente, a los estudios 1 y 3 contenidos en este trabajo, tiene que ver con el reducido número de problemas que constituyen la tarea de resolución de AWP. La prueba de AWP fue diseñada como una tarea breve que permitiera obtener una medida rápida y eficaz de la competencia del estudiante en la resolución de los distintos tipos de problemas estudiados (de una o dos operaciones consistentes y/o inconsistentes). No nos resulta ajeno que este aspecto podría afectar a las conclusiones que se establecen respecto de la eficacia de respuesta de los participantes como consecuencia de una pérdida de robustez psicométrica de la tarea debido al escaso número de problemas utilizados. Sin embargo, a pesar del reducido número de problemas de la prueba, hemos encontrado a lo largo de los distintos estudios realizados unos coeficientes de fiabilidad alfa de Cronbach óptimos de la tarea, lo cual es especialmente significativo teniendo en cuenta que alcanzar una fiabilidad aceptable en un test psicométrico resulta más difícil cuanto menor es el número de ítems que componen dicha tarea. No obstante, en el estudio 2 se empleó una tarea AWP compuesta por un número mayor de problemas, encontrándose resultados similares en relación a la fiabilidad de la prueba y la eficacia diferencial de los participantes en los distintos tipos de problemas. Los resultados obtenidos en este trabajo son, en cualquier caso, consistentes con la literatura disponible hasta la fecha.

Otra de las limitaciones de este trabajo tiene que ver con la ausencia de resultados parciales en los AWP, así como el procedimiento de cálculo llevado a cabo por los participantes para la obtención de la respuesta. Este aspecto tiene importancia debido al tipo de respuesta que pueden dar los participantes durante el proceso de resolución. Como hemos visto, los problemas inconsistentes pueden generar tres tipos de respuestas: correctas, superficiales y errores no-superficiales. Así, dado el resultado concreto, podremos identificar el tipo de respuesta dada (correcta, superficial o errónea no-superficial) comparando la solución proporcionada por el participante con la colección de alternativas esperables. No obstante, aunque la identificación de la naturaleza de la respuesta correcta y superficial permite dibujar el proceso de resolución llevado a cabo por el participante, esta medida no permite establecer la secuencia de resolución llevada a cabo por el participante cuando da como solución al problema una respuesta errónea no-superficial, ya que no se dispone más que del resultado erróneo emitido como respuesta al problema. Dada la cantidad indeterminada de respuestas incorrectas que, como hemos visto, se producen en la resolución de AWP, sería conveniente contar con una medida de proceso que permitiera identificar el momento y la causa por la que el participante estableció como solución al problema una respuesta incorrecta. Este aspecto resulta relevante, ya que podría suceder que el participante emplease una estrategia heurística de respuesta y que, en el momento de calcular las operaciones, cometiera además un error en la secuencia de cálculo que condujera a clasificar su respuesta como una incorrecta del tipo no-superficial. Este doble error, superficial y procedimental, debe ser un aspecto a tener en cuenta en futuros trabajos y que, además, puede resultar de interés en caso de que quisiera desarrollarse una prueba basada en AWP para la identificación de dificultades matemáticas en los estudiantes.

Otro aspecto a tener en cuenta tiene que ver con la evaluación de la metacognición realizada en el estudio 3. La medida de dificultad no parece una medida suficiente para identificar los aspectos relativos a las distintas dificultades que experimentan los estudiantes frente al proceso de resolución de problemas, las estrategias que utilizan en cada una de las fases de resolución, o el grado de certidumbre que les proporciona su propia respuesta. Los resultados encontrados en este estudio parecen indicar una tendencia sesgada de los participantes a la hora de valorar la dificultad de los problemas. Sin embargo, no es posible identificar los motivos que fundamentan dicha valoración y que podrían justificar el bajo nivel de eficacia con la que resuelven los problemas, especialmente la tarea CRT. Por esta razón, deberá explorarse en próximos estudios de manera más exhaustiva y precisa los aspectos relativos a la metacognición de los estudiantes con medidas que informen acerca de los conocimientos y procedimientos, como la supervisión y revisión de respuestas, que ponen en marcha durante las distintas fases que constituyen el proceso de resolución de problemas.

En relación con las medidas de funcionamiento ejecutivo, tal y como se indicó en el estudio 2, no se han tomado medidas relativas al cambio atencional o shifting de los participantes, una habilidad que según algunos autores podría estar implicada en el rendimiento matemático al favorecer la alternancia entre estrategias y el manejo de soluciones intermedias en los problemas aritméticos de varios pasos (Toll et al., 2011). Asimismo, otras variables como la rotación espacial, la capacidad atencional de los participantes o la velocidad de procesamiento (véase, Allen et al., 2019; Boonen et al., 2013; Cragg et al., 2017; Passolunghi & Lanfranchi, 2012) podrían resultar medidas interesantes a tener en cuenta en futuros estudios con el fin de integrar un modelo teórico global sobre resolución de problemas.

Otra limitación de este estudio, aunque ciertamente relativa, hace referencia al hecho de que no han sido tomadas medidas de las habilidades aritméticas para establecer una línea base de los conocimientos mínimos que disponen los participantes para la ejecución de la tarea. Esta medida puede ser relevante entre los participantes de menor edad con objeto de asegurar que cuentan con los conocimientos mínimos necesarios para llevar a cabo el proceso de resolución de problemas. No obstante, dado el nivel escolar de los participantes y, por tanto, de los conocimientos, experiencias y formación que, consecuentemente, se les presupone, consideramos que esta medida podría no ser necesaria en adolescentes, dada la elementalidad de las operaciones aritméticas necesarias para resolver adecuadamente los problemas: suma, resta y multiplicación.

Una última limitación que debe señalarse y que tiene que ver con la naturaleza de este trabajo, basado en el funcionamiento ejecutivo subyacente a la resolución de problemas aritméticos verbales, es la ausencia de medidas de tipo emocional de los participantes, estrechamente vinculadas con las respuestas de ansiedad ante las Matemáticas que experimentan los alumnos y que son causantes, también, de un pobre rendimiento en la resolución de problemas (Justicia-Galiano et al., 2017). A este respecto, algunos autores como Pelegrina y colaboradores (2020) han podido identificar que las respuestas de ansiedad limitan el rendimiento de los procesos de actualización, dificultando la recuperación de la información de tipo numérico en la memoria operativa, lo que provoca que estos alumnos se demoren más en sus respuestas y cometan un mayor número de errores. Este aspecto debe también ser tenido en cuenta en futuros trabajos, ya que podría ser una de las causas por las que los participantes cometerían un elevado número de errores no-superficiales en los problemas. Del mismo modo, también las variables de tipo motivacional, tales como la motivación intrínseca o la autopercepción de competencia, han sido identificadas por algunos autores como variables que influyen en el rendimiento que los alumnos muestran en Matemáticas (Abín et al., 2020; Lipnevich et al., 2016).

Por todo ello, quisiéramos señalar como conclusión general a este trabajo la necesidad de que la acción del psicólogo-investigador esté dirigida hacia la determinación precisa de las magnitudes y funciones descritas por las fuerzas cognitivas y emocionales causantes de un fenómeno inteligente. Sólo así será posible establecer y cuantificar las causas de tipo psicológico responsables de un bajo rendimiento académico y establecer un punto de partida a partir del cual sea posible elaborar programas de intervención eficaces para la mejora y el desarrollo de la competencia matemática.

Madrid, septiembre de 2021.

6. REFERENCIAS

Referencias

Abdullah, A. H., Abidin, N. L. Z., & Ali, M. (2015). Analysis of students' errors in solving Higher Order Thinking Skills (HOTS) problems for the topic of fraction. *Asian Social Science*, *11*(21), 133. https://doi.org/10.5539/ass.v11n21p133

Abdullah, A. H., Rahman, S. N. S. A., & Hamzah, M. H. (2017). Metacognitive skills of malaysian students in non-routine mathematical problem solving. *Bolema: Boletim de Educação Matemática*, *31*(57), 310–322. https://doi.org/10.1590/1980-4415v31n57a15

Abín, A., Núñez, J. C., Rodríguez, C., Cueli, M., García, T., & Rosário, P. (2020). Predicting mathematics achievement in secondary education: The role of cognitive, motivational, and emotional variables. *Frontiers in Psychology*, *11*, 876. https://doi.org/10.3389/fpsyg.2020.00876

Agostino, A., Johnson, J., & Pascual-Leone, J. (2010). Executive functions underlying multiplicative reasoning: Problem type matters. *Journal of Experimental Child Psychology*, *105*(4), 286–305. https://doi.org/10.1016/j.jecp.2009.09.006

Agrillo, C. (2015). Numerical and arithmetic abilities in non-primate species. In R. Cohen Kadosh & A. Dowker (Eds.), *The Oxford handbook of numerical cognition.* (pp. 214–236). Oxford University Press.

Allen, J. D. (2005). Grades as valid measures of academic achievement of classroom learning. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *78*(5), 218–223. https://doi.org/10.3200/tchs.78.5.218-223

Allen, K., Higgins, S., & Adams, J. (2019). The Relationship between Visuospatial Working Memory and Mathematical Performance in School-Aged Children: A Systematic Review. *Educational Psychology Review*, *31*(3), 509–531. https://doi.org/10.1007/s10648-019-09470-8

Alonso-Tapia, J., Mateos, M., & Gutiérrez-Martínez, F. (1991). *Entrenamiento metacognitivo: Desarrollo de programas para la mejora de la comprensión lectora y los procesos de razonamiento*. Universidad Autónoma de Madrid (UAM). Retrieved October 10, 2021, from http://hdl.handle.net/11162/84512

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General*, *136*(4), 569–576. https://doi.org/10.1037/0096-3445.136.4.569

Alvarez, J. A., & Emory, E. (2006). Executive function and the frontal lobes: A meta-analytic review. *Neuropsychology Review*, *16*(1), 17–42. https://doi.org/10.1007/s11065-006-9002-x

Amso, D., & Johnson, S. P. (2008). Development of visual selection in 3- to 9-month-olds: Evidence from saccades to previously ignored locations. *Infancy*, *13*(6), 675–686. https://doi.org/10.1080/15250000802459060

Andersson, U. (2007). The contribution of working memory to children's mathematical word problem solving. *Applied Cognitive Psychology*, *21*(9), 1201–1216. https://doi.org/10.1002/acp.1317

Andrews, G., Halford, G. S., Chappell, M., Maujean, A., & Shum, D. H. K. (2014). Planning following stroke: A relational complexity approach using the tower of london. *Frontiers in Human Neuroscience*, *8*, 1032. https://doi.org/10.3389/fnhum.2014.01032

Aoki, T. (1977). On the counting process of patterned dots. *Tohoku Psychologica Folia*, *36*(1–4), 15–22.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *Psychology of Learning and Motivation* (Vol. 2, pp. 89–195). Academic Press. https://doi.org/10.1016/S0079-7421(08)60422-3

Baddeley, A. (1986). *Working memory*. Clarendon Press/Oxford University Press.

Baddeley, A. (1990). The development of the concept of working memory: Implications and contributions of neuropsychology. In G. Vallar & T. Shallice (Eds.), *Neuropsychological impairments of short-term memory.* (pp. 54–73). Cambridge University Press.

Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556. https://doi.org/10.1126/science.1736359

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

Baddeley, A. D., & Hitch, G. (1974). Working Memory. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 8, pp. 47–89). Academic Press. https://doi.org/10.1016/S0079-7421(08)60452-1

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychol Rev*, *105*(1), 158–173. PubMed. https://doi.org/10.1037/0033-295x.105.1.158

Banich, M. T. (2009). Executive function: The search for an integrated account. *Current Directions in Psychological Science*, *18*(2), 89–94. https://doi.org/10.1111/j.1467-8721.2009.01615.x

Barrouillet, P. (2011). Dual-process theories and cognitive development: Advances and challenges. *Developmental Review*, *31*(2–3), 79–85. https://doi.org/10.1016/j.dr.2011.07.002

Bassok, M., Chase, V., & Martin, S. (1998). Adding apples and oranges: Alignment of semantic and formal knowledge. *Cognitive Psychology*, *35*(2), 99–134. https://doi.org/10.1006/cogp.1998.0675

Bassok, M., Pedigo, S. F., & Oskarsson, A. T. (2008). Priming addition facts with semantic relations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(2), 343–352. https://doi.org/10.1037/0278-7393.34.2.343

Bassok, M., Wu, L. L., & Olseth, K. L. (1995). Judging a book by its cover: Interpretative effects of content on problem-solving transfer. *Memory & Cognition*, *23*(3), 354–367. https://doi.org/10.3758/bf03197236

Beatty, E. L., & Vartanian, O. (2015). The prospects of working memory training for improving deductive reasoning. *Frontiers in Human Neuroscience*, *9*, 56. https://doi.org/10.3389/fnhum.2015.00056

Bell, A., Fischbein, E., & Greer, B. (1984). Choice of operation in verbal arithmetic problems: The effects of number size, problem structure and context. *Educational Studies in Mathematics*, *15*(2), 129–147. https://doi.org/10.1007/BF00305893

Beran, M. J., Perdue, B. M., & Evans, T. A. (2015). Monkey mathematical abilities. In *The Oxford handbook of numerical cognition.* (pp. 237–257). Oxford University Press.

Blessing, S. B., & Ross, B. H. (1996). Content effects in problem categorization and problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(3), 792–810. https://doi.org/10.1037/0278-7393.22.3.792

Boonen, A. J. H., de Koning, B. B., Jolles, J., & van der Schoot, M. (2016). Word problem solving in contemporary math education: A plea for reading comprehension skills training. *Frontiers in Psychology*, *7*, 191. https://doi.org/10.3389/fpsyg.2016.00191

Boonen, A. J. H., & Jolles, J. (2015). Second grade elementary school students' differing performance on combine, change and compare word problems. *International Journal of School and Cognitive Psychology*, *2*(122), 1–6. https://doi.org/10.4172/2469-9837.1000122

Boonen, A. J. H., van der Schoot, M., van Wesel, F., de Vries, M. H., & Jolles, J. (2013). What underlies successful word problem solving? A path analysis in sixth grade students. *Contemporary Educational Psychology*, *38*(3), 271–279. https://doi.org/10.1016/j.cedpsych.2013.05.001

Boote, S. K., & Boote, D. N. (2018). ABC problem in elementary mathematics education: Arithmetic before comprehension. *Journal of Mathematics Teacher Education*, *21*(2), 99–122. https://doi.org/10.1007/s10857-016-9350-2

Borella, E., Carretti, B., & Pelegrina, S. (2010). The specific role of inhibition in reading comprehension in good and poor comprehenders. *Journal of Learning Disabilities*, *43*(6), 541–552. https://doi.org/10.1177/0022219410371676

Brañas-Garza, P., Kujal, P., & Lenkei, B. (2019). Cognitive reflection test: Whom, how, when. *Journal of Behavioral and Experimental Economics*, *82*, 101455. https://doi.org/10.1016/j.socec.2019.101455

Broadbent, D. E. (1958). *Perception and communication.* Pergamon Press.

Brown, A. L. (1987). Metacognition, executive control, self-regulation, and other more mysterious mechanisms. In F. E. Weinert & R. H. Kluwe (Eds.), *Metacognition, Motivation, and Understanding* (pp. 65–116). Lawrence Erlbaum Associates.

Bull, R., & Lee, K. (2014). Executive functioning and mathematics achievement. *Child Development Perspectives*, *8*(1), 36–41. https://doi.org/10.1111/cdep.12059

Bull, R., & Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: Inhibition, switching, and working memory. *Developmental Neuropsychology*, *19*(3), 273–293. https://doi.org/10.1207/S15326942DN1903_3

Cahoon, A., Cassidy, T., & Simms, V. (2017). Parents' views and experiences of the informal and formal home numeracy environment. *Learning, Culture and Social Interaction*, *15*, 69–79. https://doi.org/10.1016/j.lcsi.2017.08.002

Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology*, *96*(1), 31–42. https://doi.org/10.1037/0022-0663.96.1.31

Camos, V. (2003). Counting strategies from 5 years to adulthood: Adaptation to structural features. *European Journal of Psychology of Education*, *18*(3), 251. https://doi.org/10.1007/BF03173247

Camos, V., & Barrouillet, P. (2018). *Working memory in development*. Routledge.

Campbell, J. I. D. (2008). Subtraction by addition. *Memory & Cognition*, *36*(6), 1094–1102. https://doi.org/10.3758/MC.36.6.1094

Campbell, J. I. D., & Alberts, N. M. (2010). Inverse reference in adults-elementary arithmetic. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *64*(2), 77–85. https://doi.org/10.1037/a0015720

Campitelli, G., & Gerrans, P. (2014). Does the cognitive reflection test measure cognitive reflection? A mathematical modeling approach. *Memory & Cognition*, *42*(3), 434–447. https://doi.org/10.3758/s13421-013-0367-9

Campitelli, G., & Labollita, M. (2010). Correlations of cognitive reflection with judgments and choices. *Judgment and Decision Making*, *5*(3), 182–191.

Carretti, B., Borella, E., Cornoldi, C., & De Beni, R. (2009). Role of working memory in explaining the performance of individuals with specific reading comprehension difficulties: A meta-analysis. *Learning and Individual Differences*, *19*(2), 246–251. https://doi.org/10.1016/j.lindif.2008.10.002

Carretti, B., Caldarola, N., Tencati, C., & Cornoldi, C. (2014). Improving reading comprehension in reading and listening settings: The effect of two training programmes focusing on metacognition and working memory. *British Journal of Educational Psychology*, *84*(2), 194–210. https://doi.org/10.1111/bjep.12022

Carriedo, N., Corral, A., Montoro, P. R., & Herrero, L. (2020). A developmental study of the bat/ball problem of CRT: How to override the bias and its relation to executive functioning. *British Journal of Psychology*, *111*(2), 335–356. https://doi.org/10.1111/bjop.12400

Carriedo, N., Corral, A., Montoro, P. R., Herrero, L., & Rucián, M. (2016). Development of the updating executive function: From 7-year-olds to young adults. *Developmental Psychology*, *52*(4), 666–678. https://doi.org/10.1037/dev0000091

Carriedo-López, N. (1992). *Enseñar a comprender: Diseño y y valoración de un programa de instrucccción para formar a los profesores en la enseñanza de estrategias de comprensión de las ideas principales en el aula.* UAM.

Case, R. (1974). Structures and strictures: Some functional limitations on the course of cognitive growth. *Cognitive Psychology*, *6*(4), 544–574. https://doi.org/10.1016/0010-0285(74)90025-5

Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology*, *33*(3), 386–404. https://doi.org/10.1016/0022-0965(82)90054-6

Caspi, A., Wright, B. R. E., Moffitt, T. E., & Silva, P. A. (1998). Early Failure in the Labor Market: Childhood and Adolescent Predictors of Unemployment in the Transition to Adulthood. *American Sociological Review*, *63*(3), 424. https://doi.org/10.2307/2657557

Castro-Martínez, E., & Frías-Zorilla, A. (2013). Two-step arithmetic word problems. *The Mathematics Enthusiast*, *10*(1), 379–406.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, *54*(1), 1–22. https://doi.org/10.1037/h0046743

Cattell, R. B., & Cattell, A. K. S. (1963). *Culture fair intelligence test: Scale 3*. Institute for Personality and Ability Testing, Inc.

Chai, W. J., Abd Hamid, A. I., & Abdullah, J. M. (2018). Working memory from the psychological and neurosciences perspectives: A review. *Frontiers in Psychology*, *9*, 401. https://doi.org/10.3389/fpsyg.2018.00401

Cheng, Y. L., & Mix, K. S. (2014). Spatial training improves children's mathematics ability. *Journal of Cognition and Development*, *15*(1), 2–11. https://doi.org/10.1080/15248372.2012.725186

Chi, M. T. H., Glaser, R., & Farr, M. J. (1988). *The nature of expertise*. Psychology Press. https://doi.org/10.4324/9781315799681

Clark, S. V., Semmel, E. S., Aleksonis, H. A., Steinberg, S. N., & King, T. Z. (2021). Cerebellar-subcortical-cortical systems as modulators of cognitive functions. *Neuropsychology Review*. https://doi.org/10.1007/s11065-020-09465-1

Constantinidis, C., & Luna, B. (2019). Neural substrates of inhibitory control maturation in adolescence. *Trends in Neurosciences*, *42*(9), 604–616. https://doi.org/10.1016/j.tins.2019.07.004

Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences*, *7*(12), 547–552. https://doi.org/10.1016/j.tics.2003.10.005

Cooney, J. B., & Swanson, H. L. (1990). Individual differences in memory for mathematical story problems: Memory span and problem perception. *Journal of Educational Psychology*, *82*(3), 570–577. https://doi.org/10.1037/0022-0663.82.3.570

Cornoldi, C., Carretti, B., & Colpo, C. (2017). *Prove MT-Kit Scuola. Dalla valutazione degli apprendimenti di lettura e comprensione al potenziamento*. Giunti EDU.

Cornoldi, C., Carretti, B., Drusi, S., & Tencati, C. (2015). Improving problem solving in primary school students: The effect of a training programme focusing on metacognition and working memory. *British Journal of Educational Psychology*, *85*(3), 424–439. https://doi.org/10.1111/bjep.12083

Cragg, L., Richardson, S., Hubber, P. J., Keeble, S., & Gilmore, C. (2017). When is working memory important for arithmetic? The impact of strategy and age. *PLOS ONE*, *12*(12), e0188693. https://doi.org/10.1371/journal.pone.0188693

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Cuetos, D., Rodríguez, B., Ruano, E., & Arribas, D. (2004). *Batería de evaluación de los procesos lectores.* TEA.

Culbertson, W. C., & Zillmer, E. A. (1998). The Tower of London (DX): A standardized approach to assessing executive functioning in children. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *13*(3), 285–301.

Cummins, D. D., Kintsch, W., Reusser, K., & Weimer, R. (1988). The role of understanding in solving word problems. *Cognitive Psychology*, *20*(4), 405–438. https://doi.org/10.1016/0010-0285(88)90011-4

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450–466. https://doi.org/10.1016/S0022-5371(80)90312-6

D'Antuono, G., La Torre, F. R., Marin, D., Antonucci, G., Piccardi, L., & Guariglia, C. (2017). Role of working memory, inhibition, and fluid intelligence in the performance of the Tower of London task. *Applied Neuropsychology: Adult*, *24*(6), 548–558. https://doi.org/10.1080/23279095.2016.1225071

Daroczy, G., Meurers, D., Heller, J., Wolska, M., & Nürk, H. C. (2020). The interaction of linguistic and arithmetic factors affects adult performance on arithmetic word problems. *Cognitive Processing*, *21*(1), 105–125. https://doi.org/10.1007/s10339-019-00948-5

Daroczy, G., Wolska, M., Meurers, W. D., & Nuerk, H. C. (2015). Word problems: A review of linguistic and numerical factors contributing to their difficulty. *Frontiers in Psychology*, *6*, 348. https://doi.org/10.3389/fpsyg.2015.00348 de Boer, H., Donker, A. S., Kostons, D. D. N. M., & van der Werf, G. P. C. (2018). Long-term effects of metacognitive strategy instruction on student academic performance: A meta-analysis. *Educational Research Review*, *24*, 98–115. https://doi.org/10.1016/j.edurev.2018.03.002

De Corte, E., Verschaffel, L., & de Win, L. (1985). Influence of rewording verbal problems on children's problem representations and solutions. *Journal of Educational Psychology*, *77*(4), 460–470. https://doi.org/10.1037/0022-0663.77.4.460

De Cruz, H. (2006). Towards a Darwinian Approach to Mathematics. *Foundations of Science*, *11*(1), 157–196. https://doi.org/10.1007/s10699-004-5916-z

de Koning, B. B., Boonen, A. J. H., & van der Schoot, M. (2017). The consistency effect in word problem solving is effectively reduced through verbal instruction. *Contemporary Educational Psychology*, *49*, 121–129. https://doi.org/10.1016/j.cedpsych.2017.01.006

De Neys, W. (2012). Bias and Conflict: A Case for Logical Intuitions. *Perspectives on Psychological Science*, *7*(1), 28–38. https://doi.org/10.1177/1745691611429354

De Neys, W., & Franssens, S. (2009). Belief inhibition during thinking: Not always winning but at least taking part. *Cognition*, *113*(1), 45–61. https://doi.org/10.1016/j.cognition.2009.07.009

De Neys, W., & Pennycook, G. (2019). Logic, Fast and Slow: Advances in Dual-Process Theorizing. *Current Directions in Psychological Science*, *28*(5), 503–509. https://doi.org/10.1177/0963721419855658

De Neys, W., Rossi, S., & Houdé, O. (2013). Bats, balls, and substitution sensitivity: Cognitive misers are no happy fools. *Psychonomic Bulletin & Review*, *20*(2), 269–273. https://doi.org/10.3758/s13423-013-0384-5

Dehaene, S. (1992). Varieties of numerical abilities. *Numerical Cognition*, *44*(1), 1–42. https://doi.org/10.1016/0010-0277(92)90049-N

Dehaene, S., & Cohen, L. (1995). Towards an anatomical and functional model of number processing. *Mathematical Cognition*, *1*(1), 83–120.

Dempster, F. N., & Corkill, A. J. (1999). Interference and inhibition in cognition and behavior: Unifying themes for educational psychology. *Educational Psychology Review*, *11*(1), 1–88. https://doi.org/10.1023/A:1021992632168

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. https://doi.org/10.1007/s11409-008-9026-0

Desoete, A., Roeyers, H., & Buysse, A. (2001). Metacognition and mathematical problem solving in grade 3. *Journal of Learning Disabilities*, *34*(5), 435–447. https://doi.org/10.1177/002221940103400505

Desoete, A., Roeyers, H., & De Clercq, A. (2003). Can offline metacognition enhance mathematical problem solving? *Journal of Educational Psychology*, *95*(1), 188–200. https://doi.org/10.1037/0022-0663.95.1.188

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*(1), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Duncan, J., Parr, A., Woolgar, A., Thompson, R., Bright, P., Cox, S., Bishop, S., & Nimmo-Smith, I. (2008). Goal neglect and Spearman's g: Competing parts of a complex task. *Journal of Experimental Psychology: General*, *137*(1), 131–148. https://doi.org/10.1037/0096-3445.137.1.131

Duncker, K. (1945). On problem-solving. *Psychological Monographs*, *58*(5), i–113. https://doi.org/10.1037/h0093599

Duque de Blas, G., & Gómez-Veiga, I. (2018). The spelling test for secondary school. *Unpublished Work*.

Duque de Blas, G., Gómez-Veiga, I., & García-Madruga, J. A. (2021). Arithmetic Word Problems Revisited: Cognitive Processes and Academic Performance in Secondary School. *Education Sciences*, *11*(4). https://doi.org/10.3390/educsci11040155

Eccles, J. S., Wigfield, A., & Byrnes, J. (2003). Cognitive development in adolescence. In *Handbook of psychology: Developmental psychology, Vol. 6.* (pp. 325–350). John Wiley & Sons, Inc. https://doi.org/10.1002/0471264385.wei0613

Ecker, U. K. H., Lewandowsky, S., Oberauer, K., & Chee, A. E. H. (2010). The components of working memory updating: An experimental decomposition and individual differences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(1), 170–189. https://doi.org/10.1037/a0017891

Eichmann, B., Goldhammer, F., Greiff, S., Pucite, L., & Naumann, J. (2019). The role of planning in complex problem solving. *Computers & Education*, *128*, 1–12. https://doi.org/10.1016/j.compedu.2018.08.004

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working Memory, Short-Term Memory, and General Fluid Intelligence: A Latent-Variable Approach. *Journal of Experimental Psychology: General*, *128*(3), 309–331. https://doi.org/10.1037/0096-3445.128.3.309

Evans, J. St. B. T. (1984). Heuristic and analytic processes in reasoning. *British Journal of Psychology*, *75*(4), 451–468. https://doi.org/10.1111/j.2044-8295.1984.tb01915.x

Evans, J. St. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences*, *7*(10), 454–459. https://doi.org/10.1016/j.tics.2003.08.012

Evans, J. St. B. T. (2007). *Hypothetical thinking*. Psychology Press.

Evans, J. St. B. T. (2008). Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. *Annual Review of Psychology*, *59*(1), 255–278. https://doi.org/10.1146/annurev.psych.59.103006.093629

Evans, J. St. B. T., & Over, D. E. (1996). Rationality in the selection task: Epistemic utility versus uncertainty reduction. *Psychological Review*, *103*(2), 356–363. https://doi.org/10.1037/0033-295x.103.2.356

Evans, J. St. B. T., & Stanovich, K. E. (2013a). Dual-process theories of higher cognition. *Perspectives on Psychological Science*, *8*(3), 223–241. https://doi.org/10.1177/1745691612460685

Evans, J. St. B. T., & Stanovich, K. E. (2013b). Theory and metatheory in the study of dual processing: Reply to comments. *Perspectives on Psychological Science*, *8*(3), 263–271. https://doi.org/10.1177/1745691613483774

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*(7), 307–314. https://doi.org/10.1016/j.tics.2004.05.002

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. SAGE Publications.

Fischbein, E. (1989). Tacit models and mathematical reasoning. *For the Learning of Mathematics*, *9*(2), 9–14. JSTOR.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Fletcher, L., & Carruthers, P. (2012). Metacognition and reasoning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1594), 1366–1378. https://doi.org/10.1098/rstb.2011.0413

Follmer, D. J. (2018). Executive function and reading comprehension: A meta-analytic review. *Educational Psychologist*, *53*(1), 42–60. https://doi.org/10.1080/00461520.2017.1309295

Formoso, J., Jacubovich, S., Injoque-Ricle, I., & Barreyro, J. P. (2018). Resolution of arithmetic problems, processing speed and working memory in children. *Trends in Psychology*, *26*, 1249–1266.

Frank, M. C., Vul, E., & Johnson, S. P. (2009). Development of infants' attention to faces during the first year. *Cognition*, *110*(2), 160–170. https://doi.org/10.1016/j.cognition.2008.11.010

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42. https://doi.org/10.1257/089533005775196732

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology. General*, *133*(1), 101–135. https://doi.org/10.1037/0096-3445.133.1.101

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science*, *17*(2), 172–179. https://doi.org/10.1111/j.1467-9280.2006.01681.x

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General*, *137*(2), 201–225. https://doi.org/10.1037/0096-3445.137.2.201

Friso-van den Bos, I., Kroesbergen, E. H., Van Luit, J. E. H., Xenidou-Dervou, I., Jonkman, L. M., Van der Schoot, M., & Van Lieshout, E. C. D. M. (2015). Longitudinal development of number line estimation and mathematics performance in primary school children. *Journal of Experimental Child Psychology*, *134*, 12–29. https://doi.org/10.1016/j.jecp.2015.02.002

Friso-van den Bos, I., van der Ven, S. H. G., Kroesbergen, E. H., & van Luit, J. E. H. (2013). Working memory and mathematics in primary school children: A meta-analysis. *Educational Research Review*, *10*, 29–44. https://doi.org/10.1016/j.edurev.2013.05.003

Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M., Schatschneider, C., & Fletcher, J. M. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *Journal of Educational Psychology*, *98*(1), 29–43. https://doi.org/10.1037/0022-0663.98.1.29

Fuchs, L. S., Fuchs, D., Seethaler, P. M., & Barnes, M. A. (2020). Addressing the role of working memory in mathematical word-problem solving when designing intervention for struggling learners. *ZDM*, *52*(1), 87–96. https://doi.org/10.1007/s11858-019-01070-8

Fuchs, L. S., Geary, D. C., Compton, D. L., Fuchs, D., Hamlett, C. L., Seethaler, P. M., Bryant, J. D., & Schatschneider, C. (2010). Do different types of school mathematics development depend on different constellations of numerical versus general cognitive abilities? *Developmental Psychology*, *46*(6), 1731–1746. https://doi.org/10.1037/a0020662

Fuchs, L. S., Gilbert, J. K., Fuchs, D., Seethaler, P. M., & Martin, B. L. N. (2018). Text comprehension and oral language as predictors of word-problem solving: Insights into word-problem solving as a form of text comprehension. *Scientific Studies of Reading*, *22*(2), 152–166. https://doi.org/10.1080/10888438.2017.1398259

Fuentes, P. (1998). Reading comprehension in mathematics. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *72*(2), 81–88. https://doi.org/10.1080/00098659809599602

Fung, W., & Swanson, H. L. (2017). Working memory components that predict word problem solving: Is it merely a function of reading, calculation, and fluid intelligence? *Memory & Cognition*, *45*(5), 804–823. https://doi.org/10.3758/s13421-017-0697-0

Fürst, A. J., & Hitch, G. J. (2000). Separate roles for executive and phonological components of working memory in mental arithmetic. *Memory & Cognition*, *28*(5), 774–782. https://doi.org/10.3758/BF03198412

Fuson, K. C. (1992). Research on whole number addition and subtraction. In *Handbook of research on mathematics teaching and learning: A project of the National Council of Teachers of Mathematics.* (pp. 243–275). Macmillan Publishing Co, Inc.

Fyfe, E. R., Rittle-Johnson, B., & DeCaro, M. S. (2012). The effects of feedback during exploratory mathematics problem solving: Prior knowledge matters. *Journal of Educational Psychology*, *104*(4), 1094–1108. https://doi.org/10.1037/a0028389

Garcia-Barrera, M. A. (2019). Unity and diversity of dysexecutive syndromes. In A. Ardila, S. Fatima, & M. Rosselli (Eds.), *Dysexecutive Syndromes: Clinical and Experimental Perspectives* (pp. 3–27). Springer International Publishing. https://doi.org/10.1007/978-3-030-25077-5_1

García-Madruga, J. A. (2006). *Lectura y conocimiento*. Paidós.

García-Madruga, J. A., & Duque de Blas, G. (2018). WM Analogies task for secondary school. *Unpublished Work*.

García-Madruga, J. A., Elosúa, M. R., Gil, L., Gómez-Veiga, I., Vila, J. O., Orjales, I., Contreras, A., Rodríguez, R., Melero, M. A., & Duque, G. (2013). Reading Comprehension and Working Memory's Executive Processes: An Intervention Study in Primary School Students. *Reading Research Quarterly*, *48*(2), 155–174. https://doi.org/10.1002/rrq.44

García-Madruga, J. A., Gómez-Veiga, I., & Vila, J. O. (2016). Executive functions and the improvement of thinking abilities: The intervention in reading comprehension. *Frontiers in Psychology*, *7*, 58. https://doi.org/10.3389/fpsyg.2016.00058

García-Madruga, J. A., Gutiérrez, F., Carriedo, N., Luzón, J. M., & Vila, J. O. (2007). Mental models in propositional reasoning and working memory's central executive. *Thinking & Reasoning*, *13*(4), 370–393. https://doi.org/10.1080/13546780701203813

García-Madruga, J. A., Orenes, I., Vila-Chaves, J. O., & Gómez-Veiga, I. (2022). Executive functions and thinking: An intervention program to improve deductive reasoning abilities. *Manuscript Submitted for Publication*.

García-Madruga, J. A., Santamaría, C., & Moreno-Ríos, S. (2016). Pensamiento y razonamiento. In M. T. Bajo, L. J. Fuentes, J. Lupiañez, & M. Rueda (Eds.), *Mente y cerebro: De la psicología experimental a la neurociencia cognitiva: Pío Tudela, una trayectoria científica* (pp. 409–442). Alianza Editorial.

García-Madruga, J. A., Santamaría, C., Moreno-Ríos, S., Vila-Chaves, J. O., Gómez-Veiga, I., & Orenes, I. (2014). A test of deductive reasoning. *Unpublished Work*.

García-Madruga, J. A., & Vila, J. O. (2013). Dificultades de aprendizaje de las Matemáticas. In J. O. Vila & F. Gutiérrez-Martínez (Eds.), *Manual básico de dificultades de aprendizaje. Concepto, evaluación e intervención.* (pp. 331–364). Sanz y Torres.

García-Madruga, J. A., Vila, J. O., Gómez-Veiga, I., Duque, G., & Elosúa, M. R. (2014). Executive processes, reading comprehension and academic achievement in 3th grade primary students. *Learning and Individual Differences*, *35*, 41–48. https://doi.org/10.1016/j.lindif.2014.07.013

Garofalo, J., & Lester, F. K. (1985). Metacognition, cognitive monitoring, and mathematical performance. *Journal for Research in Mathematics Education JRME*, *16*(3), 163–176. https://doi.org/10.5951/jresematheduc.16.3.0163

Gathercole, S. E., & Alloway, T. P. (2006). Practitioner review: Short-term and working memory impairments in neurodevelopmental disorders: Diagnosis and remedial support. *Journal of Child Psychology and Psychiatry*, *47*(1), 4–15. https://doi.org/10.1111/j.1469-7610.2005.01446.x

Gathercole, S. E., Pickering, S. J., Knight, C., & Stegmann, Z. (2004). Working memory skills and educational attainment: Evidence from national curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology*, *18*(1), 1–16. https://doi.org/10.1002/acp.934

Gathmann, B., Brand, M., & Schiebener, J. (2017). One executive function never comes alone: Monitoring and its relation to working memory, reasoning, and different executive functions. *Cognitive Processing*, *18*(1), 13–29. https://doi.org/10.1007/s10339-016-0773-6

Geary, D. C., & Hoard, M. K. (2003). Learning disabilities in basic mathematics: Deficits in memory and cognition. In J. M. Royer (Ed.), *Mathematical Cognition* (pp. 93–115). Greenwich, CT.

Gelman, R., & Gallistel, C. R. (1978). *The child's understanding of number.* (pp. xv, 260). Harvard University Press.

Gick, M. L. (1986). Problem-solving strategies. *Educational Psychologist*, *21*(1–2), 99–120. https://doi.org/10.1080/00461520.1986.9653026

Gilhooly, K. J., Wynn, V., Phillips, L. H., Logie, R. H., & Sala, S. D. (2002). Visuo-spatial and verbal working memory in the five-disc Tower of London task: An individual differences approach. *Thinking & Reasoning*, *8*(3), 165–178. https://doi.org/10.1080/13546780244000006

Ginsburg, N. (1978). Perceived numerosity, item arrangement, and expectancy. *The American Journal of Psychology*, *91*(2), 267–273. JSTOR. https://doi.org/10.2307/1421536

Goldstein, S., Naglieri, J. A., Princiotta, D., & Otero, T. M. (2014). Introduction: A history of executive functioning as a theoretical and clinical construct. In S. Goldstein & J. A. Naglieri (Eds.), *Handbook of executive functioning.* (pp. 3–12). Springer.

Gómez-Chacón, I., García-Madruga, J. A., Rodríguez, R., Vila, J. O., & Elosúa, M. R. (2011). *Mathematical beliefs and cognitive reflection: Do they predict academic achievement*. 64–73. Retrieved October 10, 2021, from https://www.researchgate.net/profile/Maria-Elosua/publication/265235541_Mathematical_beliefs_and_cognitive_reflection_Do_the y_predict_academic_achievement/links/55a643c708aec374938f9049/Mathematical-beliefs-and-cognitive-reflection-Do-they-predict-academic-achievement.pdf

Gómez-Chacón, I., García-Madruga, J. A., Vila, J. O., Elosúa, M. R., & Rodríguez, R. (2014). The dual processes hypothesis in mathematics performance: Beliefs, cognitive reflection, working memory and reasoning. *Learning and Individual Differences*, *29*, 67–73. https://doi.org/10.1016/j.lindif.2013.10.001

Gómez-Veiga, I., Vila Chaves, J. O., Duque de Blas, G., & García Madruga, J. A. (2018). A new look to a classic issue: Reasoning and academic achievement at secondary school. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.00400

Gómez-Veiga, I., Vila Chaves, J. O., Duque, G., & García Madruga, J. A. (2018). A New Look to a Classic Issue: Reasoning and Academic Achievement at Secondary School. *Frontiers in Psychology*, *9*, 400. https://doi.org/10.3389/fpsyg.2018.00400

González-Labra, M. J. (2019). *Psicología del Pensamiento*. Sanz y Torres.

Gros, H., Sander, E., & Thibaut, J. P. (2019). When masters of abstraction run into a concrete wall: Experts failing arithmetic word problems. *Psychonomic Bulletin & Review*, *26*(5), 1738–1746. https://doi.org/10.3758/s13423-019-01628-3

Gros, H., Thibaut, J. P., & Sander, E. (2020). Semantic congruence in arithmetic: A new conceptual model for word problem solving. *Educational Psychologist*, *55*(2), 1–19. https://doi.org/10.1080/00461520.2019.1691004

Gross, J., Hudson, C., & Price, D. (2009). *The long term costs of numeracy difficulties. Every child a chance trust and KPMG*. National Numeracy.

Guthormsen, A. M., Fisher, K. J., Bassok, M., Osterhout, L., DeWolf, M., & Holyoak, K. J. (2016). Conceptual integration of arithmetic operations with real-world knowledge: Evidence from event-related potentials. *Cognitive Science*, *40*(3), 723–757. https://doi.org/10.1111/cogs.12238

Gutiérrez-Martínez, F., García-Madruga, J. A., Carriedo, N., Vila, J. O., & Luzón, J. M. (2005). Dos pruebas de amplitud de memoria operativa para el razonamiento. [Two working memory measures for reasoning.]. *Cognitiva*, *17*(2), 183–207. https://doi.org/10.1174/0214355054739255

Gutiérrez-Martínez, F., Ramos-Ortega, M., & Vila-Chaves, J. O. (2017). Eficacia ejecutiva en tareas de interferencia tipo Stroop. Estudio de validación de una versión numérica y manual (CANUM). *Anales de Psicología / Annals of Psychology*, *34*(1), 184–196. https://doi.org/10.6018/analesps.34.1.263431

Gutiérrez-Martínez, F., & Vila, J. O. (2013). Dificultades de las estrategias de aprendizaje. In J. O. Vila & F. Gutiérrez-Martínez (Eds.), *Manual básico de dificultades de aprendizaje. Concepto, evaluación e intervención.* (pp. 205–243). Sanz y Torres.

Hammerstein, S., Poloczek, S., Lösche, P., Lemaire, P., & Büttner, G. (2019). Effects of working memory updating on children's arithmetic performance and strategy use: A study in computational estimation. *Journal of Experimental Child Psychology*, *184*, 174–191. https://doi.org/10.1016/j.jecp.2019.04.003

Handley, S. J., Capon, A., Copp, C., & Harper, C. (2002). Conditional reasoning and the Tower of Hanoi: The role of spatial and verbal working memory. *British Journal of Psychology*, *93*(4), 501–518. https://doi.org/10.1348/000712602761381376

Hannula, M. S., Di Martino, P., Pantziara, M., Zhang, Q., Morselli, F., Heyd-Metzuyanim, E., Lutovac, S., Kaasila, R., Middleton, J. A., Jansen, A., & Goldin, G. A. (2016). Attitudes, beliefs, motivation, and identity in mathematics education. In M. S. Hannula, P. Di Martino, M. Pantziara, Q. Zhang, F. Morselli, E. Heyd-Metzuyanim, S. Lutovac, R. Kaasila, J. A. Middleton, A. Jansen, & G. A. Goldin (Eds.), *Attitudes, Beliefs, Motivation and Identity in Mathematics Education: An Overview of the Field and Future Directions* (pp. 1–35). Springer International Publishing. https://doi.org/10.1007/978-3-319-32811-9_1

Harel, G., Behr, M., Post, T. R., & Lesh, R. (1994). The impact of number type on the solution of multiplication and division problems: Further considerations. In G. Harel & J. Confrey (Eds.), *The development of multiplicative reasoning in the learning of mathematics* (pp. 365–388). SUNY Press.

Harnishfeger, K. K. (1995). The development of cognitive inhibition: Theories, definitions, and research evidence. In F. N. Dempster, C. J. Brainerd, & C. J. Brainerd (Eds.), *Interference and Inhibition in Cognition* (pp. 175–204). Academic Press. https://doi.org/10.1016/B978-012208930-5/50007-6

Hasanah, N., Hayashi, Y., & Hirashima, T. (2017). An analysis of learner outputs in problem posing as sentence-integration in arithmetic word problems. *Research and Practice in Technology Enhanced Learning*, *12*(1), 9. https://doi.org/10.1186/s41039-017-0049-5

Hegarty, M., Mayer, R. E., & Green, C. E. (1992). Comprehension of arithmetic word problems: Evidence from students' eye fixations. *Journal of Educational Psychology*, *84*(1), 76–84. https://doi.org/10.1037/0022-0663.84.1.76

Hegarty, M., Mayer, R. E., & Monk, C. A. (1995). Comprehension of arithmetic word problems: A comparison of successful and unsuccessful problem solvers. *Journal of Educational Psychology*, *87*(1), 18–32. https://doi.org/10.1037/0022-0663.87.1.18

Hernández-Armenteros, J., Pérez-García, J. A., Furió-Párraga, B., Hernández-Chica, J., & Salinas-González, L. (2018). *La Universidad Española en cifras 2017/2018*. Retrieved October 10, 2021, from https://www.crue.org/wp-content/uploads/2020/02/UEC-1718_FINAL_DIGITAL.pdf

Heyder, K., Suchan, B., & Daum, I. (2004). Cortico-subcortical contributions to executive control. *Executive Control of Human Action*, *115*(2), 271–289. https://doi.org/10.1016/j.actpsy.2003.12.010

Hoff, K. A., Briley, D. A., Wee, C. J. M., & Rounds, J. (2018). Normative changes in interests from adolescence to adulthood: A meta-analysis of longitudinal studies. *Psychological Bulletin*, *144*(4), 426–451. https://doi.org/10.1037/bul0000140

Holmes, J., Adams, J. W., & Hamilton, C. J. (2008). The relationship between visuospatial sketchpad capacity and children's mathematical skills. *European Journal of Cognitive Psychology*, *20*(2), 272–289. https://doi.org/10.1080/09541440701612702

Horn, J. L., & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica*, *26*(2), 107–129. https://doi.org/10.1016/0001-6918(67)90011-x

Humes, G. E., Welsh, M. C., Retzlaff, P., & Cookson, N. (1997). Towers of Hanoi and London: Reliability and Validity of Two Executive Function Tasks. *Assessment*, *4*(3), 249–257. https://doi.org/10.1177/107319119700400305

Hyde, D. C., & Spelke, E. S. (2011). Neural signatures of number processing in human infants: Evidence for two core systems underlying numerical cognition. *Developmental Science*, *14*(2), 360–371. https://doi.org/10.1111/j.1467-7687.2010.00987.x

Iglesias-Sarmiento, V., Carriedo-López, N., & Rodríguez-Rodríguez, J. L. (2014). Updating executive function and performance in reading comprehension and problem solving. *Anales de Psicología*, *31*(1), 298–309. https://doi.org/10.6018/analesps.31.1.158111

Inhelder, B., & Piaget, J. (1955). *De la logique de l'enfant à la logique de l'adolescent.* (p. 314). Presses Universitaires de France.

Injoque-Ricle, I., & Burin, D. I. (2011). Memoria de trabajo y planificación en niños: Validación de la prueba Torre de Londres. *Neuropsicologia Latinoamericana*, *3*, 31–38.

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences*, *105*(19), 6829–6833. https://doi.org/10.1073/pnas.0801268105

Jarosz, A. F., & Jaeger, A. J. (2019). Inconsistent operations: A weapon of math disruption. *Applied Cognitive Psychology*, *33*(1), 124–138. https://doi.org/10.1002/acp.3471

Jiang, R., Li, X., Xu, P., & Lei, Y. (2020). Do teachers need to inhibit heuristic Bias in mathematics problem-solving? Evidence from a negative-priming study. *Current Psychology*. https://doi.org/10.1007/s12144-020-01209-x

Jiménez, L., & Verschaffel, L. (2014). Development of children's solutions of non-standard arithmetic word problem solving. *Revista de Psicodidáctica*, *19*(1), 93–123. https://doi.org/10.1387/RevPsicodidact.7865

Jitendra, A. (2002). Teaching Students Math Problem-Solving through Graphic Representations. *TEACHING Exceptional Children*, *34*(4), 34–38. https://doi.org/10.1177/004005990203400405

Johann, V., Könen, T., & Karbach, J. (2020). The unique contribution of working memory, inhibition, cognitive flexibility, and intelligence to reading comprehension and reading speed. *Child Neuropsychology*, *26*(3), 324–344. https://doi.org/10.1080/09297049.2019.1649381

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Harvard University Press.

Johnson-Laird, P. N. (1988). A taxonomy of thinking. In R. J. Sternberg & E. E. Smith (Eds.), *The psychology of human thought.* (pp. 429–457). Cambridge University Press.

Johnson-Laird, P. N. (2006). *How we reason.* Oxford University Press.

Justicia-Galiano, M. J., Martín-Puga, M. E., Linares, R., & Pelegrina, S. (2017). Math anxiety and math performance in children: The mediating roles of working memory and math self-concept. *British Journal of Educational Psychology*, *87*(4), 573–589. https://doi.org/10.1111/bjep.12165

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, *58*(9), 697–720. https://doi.org/10.1037/0003-066X.58.9.697

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux. https://doi.org/10.31440/dftb.17324

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment BT - Heuristics and biases: The psychology of intuitive judgment. In *Heuristics and biases: The psychology of intuitive judgment*. Cambridge University Press.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, *3*(3), 430–454. https://doi.org/10.1016/0010-0285(72)90016-3

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review*, *9*(4), 637–671. https://doi.org/10.3758/BF03196323

Karr, J. E., Areshenkoff, C. N., Rast, P., Hofer, S. M., Iverson, G. L., & Garcia-Barrera, M. A. (2018). The unity and diversity of executive functions: A systematic review and re-analysis of latent variable studies. *Psychological Bulletin*, *144*(11), 1147–1185. https://doi.org/10.1037/bul0000160

Katz, I. R., Bennett, R. E., & Berger, A. E. (2000). Effects of response format on difficulty of sat-mathematics items: It's not the strategy. *Journal of Educational Measurement*, *37*(1), 39–57. https://doi.org/10.1111/j.1745-3984.2000.tb01075.x

Kaufman, A. S., & Kaufman, N. L. (2013). *K-BIT, Test Breve de Inteligencia de KAUFMAN*. Pearson Clinical.

Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkmann, J. (1949). The discrimination of visual number. *The American Journal of Psychology*, *62*(4), 498–525. JSTOR. https://doi.org/10.2307/1418556

Khng, K. H., & Lee, K. (2009). Inhibiting interference from prior knowledge: Arithmetic intrusions in algebra word problem solving. *Learning and Individual Differences*, *19*(2), 262–268. https://doi.org/10.1016/j.lindif.2009.01.004

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge university press.

Kintsch, W., & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. *Psychological Review*, *92*(1), 109–129. https://doi.org/10.1037/0033-295x.92.1.109

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*(5), 363–394. https://doi.org/10.1037/0033-295X.85.5.363

Koffka, K. (1922). Perception: An introduction to the Gestalt-Theorie. *Psychological Bulletin*, *19*(10), 531–585. https://doi.org/10.1037/h0072422

Köhler, W. (1967). Gestalt psychology. *Psychologische Forschung*, *31*(1), XVIII–XXX. https://doi.org/10.1007/BF00422382

Kotsopoulos, D., & Lee, J. (2012). A naturalistic study of executive function and mathematical problem-solving. *The Journal of Mathematical Behavior*, *31*(2), 196–208. https://doi.org/10.1016/j.jmathb.2011.12.005

Kuhn, J. T., & Holling, H. (2009). Gender, reasoning ability, and scholastic achievement: A multilevel mediation analysis. *Learning and Individual Differences*, *19*(2), 229–233. https://doi.org/10.1016/j.lindif.2008.11.007

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity? *Intelligence*, *14*(4), 389–433. https://doi.org/10.1016/S0160-2896(05)80012-1

Laski, E. V., & Siegler, R. S. (2007). Is 27 a big number? Correlational and causal connections among numerical categorization, number line estimation, and numerical magnitude comparison. *Child Development*, *78*(6), 1723–1743. https://doi.org/10.1111/j.1467-8624.2007.01087.x

Lee, K., NG, E. L., & NG, S. F. (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. FRANCIS.

Lee, K., Ng, S. F., & Bull, R. (2018). Learning and solving algebra word problems. *The Roles of Relational Skills, Arithmetic, and Executive Functioning*, *54*(9), 1758–1772. https://doi.org/10.1037/dev0000561

Lee, K., Ng, S. F., Ng, E. L., & Lim, Z. Y. (2004). Working memory and literacy as predictors of performance on algebraic word problems. *Journal of Experimental Child Psychology*, *89*(2), 140–158. https://doi.org/10.1016/j.jecp.2004.07.001

LeFevre, J. A., Skwarchuk, S. L., Smith-Chant, B. L., Fast, L., Kamawar, D., & Bisanz, J. (2009). Home numeracy experiences and children's math performance in the early school years. *Canadian Journal of Behavioural Science / Revue Canadienne Des Sciences Du Comportement*, *41*(2), 55–66. https://doi.org/10.1037/a0014532

Legg, A. M., & Locker, L. (2009). Math performance and its relationship to math anxiety and metacognition. *North American Journal of Psychology*, *11*(3).

Lemaire, P. (2010). Cognitive strategy variations during aging. *Current Directions in Psychological Science*, *19*(6), 363–369. https://doi.org/10.1177/0963721410390354

Lemaire, P., & Lecacheur, M. (2011). Age-related changes in children's executive functions and strategy selection: A study in computational estimation. *Cognitive Development*, *26*(3), 282–294. https://doi.org/10.1016/j.cogdev.2011.01.002

Lemaire, P., & Reder, L. (1999). What affects strategy selection in arithmetic? The example of parity and five effects on product verification. *Memory & Cognition*, *27*(2), 364–382. https://doi.org/10.3758/BF03211420

Lemaire, P., & Siegler, R. S. (1995). Four aspects of strategic change: Contributions to children's learning of multiplication. *Journal of Experimental Psychology: General*, *124*(1), 83–97. https://doi.org/10.1037/0096-3445.124.1.83

Lerkkanen, M. K., Rasku-Puttonen, H., Aunola, K., & Nurmi, J. K. (2005). Mathematical performance predicts progress in reading comprehension among 7-year olds. *European*

Journal of Psychology of Education, *20*(2), 121–137. https://doi.org/10.1007/BF03173503

Lewis, A. B., & Mayer, R. E. (1987). Students' miscomprehension of relational statements in arithmetic word problems. *Journal of Educational Psychology*, *79*(4). https://doi.org/10.1037/0022-0663.79.4.363

Li, S., Ren, X., Schweizer, K., Brinthaupt, T. M., & Wang, T. (2021). Executive functions as predictors of critical thinking: Behavioral and neural evidence. *Learning and Instruction*, *71*, 101376. https://doi.org/10.1016/j.learninstruc.2020.101376

Lin, X. (2021). Investigating the Unique Predictors of Word-Problem Solving Using Meta-Analytic Structural Equation Modeling. *Educational Psychology Review*, *33*(3), 1097–1124. https://doi.org/10.1007/s10648-020-09554-w

Linares, R., Bajo, M. T., & Pelegrina, S. (2016). Age-related differences in working memory updating components. *Journal of Experimental Child Psychology*, *147*, 39–52. https://doi.org/10.1016/j.jecp.2016.02.009

Linderholm, T., & van den Broek, P. (2002). The effects of reading purpose and working memory capacity on the processing of expository text. *Journal of Educational Psychology*, *94*(4), 778–784. https://doi.org/10.1037/0022-0663.94.4.778

Lipnevich, A. A., Preckel, F., & Krumm, S. (2016). Mathematics attitudes and their unique contribution to achievement: Going over and above cognitive ability and personality. *Learning and Individual Differences*, *47*, 70–79. https://doi.org/10.1016/j.lindif.2015.12.027

López-Higes, R., Mayoral, J. A., & Villoria, C. (2002). *Batería de evaluación de la lectura (BEL)*. Psymtec.

Lubin, A., Rossi, S., Lanoë, C., Vidal, J., Houdé, O., & Borst, G. (2016). Expertise, inhibitory control and arithmetic word problems: A negative priming study in mathematics experts. *Learning and Instruction*, *45*, 40–48. https://doi.org/10.1016/j.learninstruc.2016.06.004

Lubin, A., Vidal, J., Lanoë, C., Houdé, O., & Borst, G. (2013). Inhibitory control is needed for the resolution of arithmetic word problems: A developmental negative priming study. *Journal of Educational Psychology*, *105*(3), 701–708. https://doi.org/10.1037/a0032625

Lucangeli, D., & Cornoldi, C. (1997). Mathematics and metacognition: What is the nature of the relationship? *Mathematical Cognition*, *3*(2), 121–139. https://doi.org/10.1080/135467997387443

Luria, A. R. (1966). *Higher cortical functions in man*. Basic Books.

Mañá, A., Vidal-Abarca, E., Domínguez, C., Gil, L., & Cerdán, R. (2009). The role of metacognitive processes in a question-answering task with written texts. *Journal for the Study of Education and Development*, *32*(4), 553–565. https://doi.org/10.1174/021037009789610412

Marjoribanks, K. (2005). Family environments and children's outcomes. *Educational Psychology*, *25:6*, 647–657. https://doi.org/10.1080/01443410500344704

Marzocchi, G. M., Lucangeli, D., De Meo, T., Fini, F., & Cornoldi, C. (2002). The disturbing effect of irrelevant information on arithmetic problem solving in inattentive children. *Developmental Neuropsychology*, *21*(1), 73–92. https://doi.org/10.1207/S15326942DN2101_4

Mateos, M. M. (1991). A program for teaching comprehension monitoring strategies. *Journal for the Study of Education and Development*, *14*(56), 61–76. https://doi.org/10.1080/02103702.1991.10822315

Mayer, R. E. (1992). *Thinking, problem solving, cognition* (2nd ed.). W. H. Freeman/Times Books/ Henry Holt & Co.

Mayer, R. E. (1998). Cognitive, metacognitive and motivational aspects of problem solving. *Instructional Science*, *26*(1), 49–63. https://doi.org/10.1023/A:1003088013286

Mayer, R. E., & Hegarty, M. (1996). The process of understanding mathematical problems. In R. J. Sternberg & T. Ben-Zeev (Eds.), *The Nature of Mathematical Thinking* (Lawrence Erlbaum Associates).

Mayer, R. E., Hegarty, M., & Lewis, A. B. (1992). Mathematical misunderstandings: Qualitative reasoning about quantitative problems. *Advances in Psychology*, *91*(C). https://doi.org/10.1016/S0166-4115(08)60886-9

McGrew, K. S., & Hessler, G. L. (1995). The Relationship between the WJ-R Gf-Gc Cognitive Clusters and Mathematics Achievement across the Life-Span. *Journal of Psychoeducational Assessment*, *13*(1), 21–38. https://doi.org/10.1177/073428299501300102

McLaughlin, A. C., & McGill, A. E. (2017). Explicitly teaching critical thinking skills in a history course. *Science and Education*, *26*(1–2). https://doi.org/10.1007/s11191-017-9878-2

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, *21*(1), 8–14. https://doi.org/10.1177/0963721411429458

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Morales, R. V., Shute, V. J., & Pellegrino, J. W. (1985). Developmental Differences in Understanding and Solving Simple Mathematics Word Problems. *Cognition and Instruction*, *2*(1), 41–57. https://doi.org/10.1207/s1532690xci0201_2

Mori, K., & Okamoto, M. (2017). The role of the updating function in solving arithmetic word problems. *Journal of Educational Psychology*, *109*(2), 245–256. https://doi.org/10.1037/edu0000134

Morris, N., & Jones, D. M. (1990). Memory updating in working memory: The role of the central executive. *British Journal of Psychology*, *81*(2), 111–121. https://doi.org/10.1111/j.2044-8295.1990.tb02349.x

Moshman, D. (2004). Adolescent rationality and development. In *Adolescent Rationality and Development*. Taylor & Francis. https://doi.org/10.4324/9781410611710

Munakata, Y., Herd, S. A., Chatham, C. H., Depue, B. E., Banich, M. T., & O'Reilly, R. C. (2011). A unified framework for inhibitory control. *Trends in Cognitive Sciences*, *15*(10), 453–459. https://doi.org/10.1016/j.tics.2011.07.011

Muncer, G., Higham, P., Gosling, C., Cortese, S., Wood-Downie, H., & Hadwin, J. A. (2021). A meta-analysis investigating the association between metacognition and math performance in adolescence. *Educational Psychology Review*, 84. https://doi.org/10.1007/s10648-021-09620-x

Murnane, R. J., Willett, J. B., Braatz, M. J., & Duhaldeborde, Y. (2001). Do different dimensions of male high school students' skills predict labor market success a decade later? Evidence from the NLSY. *Economics of Education Review*, *20*(4), 311–320. https://doi.org/10.1016/S0272-7757(00)00056-X

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups*. National Institute of Child Health and Human Development. Retrieved October 10, 2021, from https://www1.nichd.nih.gov/publications/pubs/nrp/Documents/report.pdf

Nauta, M. M., Kahn, J. H., Angell, J. W., & Cantarelli, E. A. (2002). Identifying the antecedent in the relation between career interests and self-efficacy: Is it one, the other, or both? *Journal of Counseling Psychology*, *49*(3), 290–301. https://doi.org/10.1037/0022-0167.49.3.290

Nelson, G., & Powell, S. R. (2018). A systematic review of longitudinal studies of mathematics difficulty. *Journal of Learning Disabilities*, *51*(6), 523–539. https://doi.org/10.1177/0022219417714773

Newell, A., & Simon, H. A. (1972). *Human problem solving.* (pp. xiv, 920). Prentice-Hall.

Nickerson, R. S. (2010). *Mathematical reasoning: Patterns, problems, conjectures, and proofs.* (pp. xi, 583). Psychology Press.

Nigg, J. T. (2000). On inhibition/disinhibition in developmental psychopathology: Views from cognitive and personality psychology and a working inhibition taxonomy. *Psychological Bulletin*, *126*(2), 220–246. https://doi.org/10.1037/0033-2909.126.2.220

Norman, D., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. Davidson, R. Schwartz, & D. Shapiro (Eds.), *Consciousness and Self-Regulation: Advances in Research and Theory IV*. Plenum Press.

Oakhill, J., Hartt, J., & Samols, D. (2005). Levels of comprehension monitoring and working memory in good and poor comprehenders. *Reading and Writing: An Interdisciplinary Journal*, *18*(7–9), 657–686. https://doi.org/10.1007/s11145-005-3355-z

Ober, T. M., Brooks, P. J., Plass, J. L., & Homer, B. D. (2019). Distinguishing direct and indirect effects of executive functions on reading comprehension in adolescents. *Reading Psychology*, *40*(6), 551–581. https://doi.org/10.1080/02702711.2019.1635239

Oberauer, K., Süß, H. M., Wilhelm, O., & Sander, N. (2007). Individual differences in working memory capacity and reasoning ability. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory.* (pp. 49–75). Oxford University Press.

OECD. (2019). *PISA 2018 results: What students know and can do. 1*. https://doi.org/10.1787/a89c90e1-es.

Oldrati, V., Patricelli, J., Colombo, B., & Antonietti, A. (2016). The role of dorsolateral prefrontal cortex in inhibition mechanism: A study on cognitive reflection test and similar tasks through neuromodulation. *Neuropsychologia*, *91*, 499–508. https://doi.org/10.1016/j.neuropsychologia.2016.09.010

Orrantia, J. (2003). The role of conceptual knowledge in solving addition and subtraction word problems. *Journal for the Study of Education and Development*, *26*(4), 451–468. https://doi.org/10.1174/021037003322553842

Orrantia, J. (2006). Dificultades en el aprendizaje de las matemáticas: Una perspectiva evolutiva. *Revista Psicopedagogía*, *23*(71), 158–180.

Orrantia, J., Rodríguez, L., Múñez, D., & Vicente, S. (2012). Inverse reference in subtraction performance: An analysis from arithmetic word problems. *Quarterly Journal of Experimental Psychology*, *65*(4), 725–738. https://doi.org/10.1080/17470218.2011.636824

Pape, S. J. (2003). Compare word problems: Consistency hypothesis revisited. *Contemporary Educational Psychology*, *28*(3), 396–421. https://doi.org/10.1016/S0361-476X(02)00046-2

Pape, S. J. (2004). Middle school children’s problem-solving behavior: A cognitive analysis from a reading comprehension perspective. *Journal for Research in Mathematics Education JRME*, *35*(3), 187–219. https://doi.org/10.2307/30034912

Pascual-Leone, J. (1970). A mathematical model for the transition rule in Piaget’s developmental stages. *Acta Psychologica*, *32*, 301–345. https://doi.org/10.1016/0001-6918(70)90108-3

Pascual-Leone, J. (1987). Organismic processes for neo-piagetian theories: A dialectical causal account of cognitive development. *International Journal of Psychology*, *22*(5–6), 531–570. https://doi.org/10.1080/00207598708246795

Pascual-Leone, J. (2000a). Mental attention, consciousness, and the progressive emergence of wisdom. *Journal of Adult Development*, *7*(4), 241–254. https://doi.org/10.1023/A:1009563428260

Pascual-Leone, J. (2000b). Reflections on working memory: Are the two models complementary? *Journal of Experimental Child Psychology*, *77*(2), 138–154. https://doi.org/10.1006/jecp.2000.2593

Pascual-Leone, J., & Johnson, J. (2011). A developmental theory of mental attention: Its application to measurement and task analysis. In P. Barrouillet & V. Gaillard (Eds.), *Cognitive development and working memory: A dialogue between neo-Piagetian theories and cognitive approaches.* (pp. 13–46). Psychology Press.

Passolunghi, M. C., & Costa, H. M. (2019). Working memory and mathematical learning. In A. Fritz, V. G. Haase, & P. Räsänen (Eds.), *International Handbook of Mathematical Learning Difficulties: From the Laboratory to the Classroom* (pp. 407–421). Springer International Publishing. https://doi.org/10.1007/978-3-319-97148-3_25

Passolunghi, M. C., Duque de Blas, G., Carretti, B., Gómez-Veiga, I., & García-Madruga, J. A. (2021). The role of working memory updating, inhibition, fluid intelligence and reading comprehension in explaining differences between consistent and inconsistent arithmetic word problem solving performance. *Manuscript Submitted for Publication*.

Passolunghi, M. C., & Lanfranchi, S. (2012). Domain-specific and domain-general precursors of mathematical achievement: A longitudinal study from kindergarten to first grade. *British Journal of Educational Psychology*, *82*(1), 42–63. https://doi.org/10.1111/j.2044-8279.2011.02039.x

Passolunghi, M. C., & Mammarella, I. C. (2010). Spatial and visual working memory ability in children with difficulties in arithmetic word problem solving. *European Journal of Cognitive Psychology*, *22*(6). https://doi.org/10.1080/09541440903091127

Passolunghi, M. C., & Pazzaglia, F. (2004). Individual differences in memory updating in relation to arithmetic problem solving. *Learning and Individual Differences*, *14*(4). https://doi.org/10.1016/j.lindif.2004.03.001

Passolunghi, M. C., & Pazzaglia, F. (2005). A comparison of updating processes in children good or poor in arithmetic word problem-solving. *Learning and Individual Differences*, *15*(4). https://doi.org/10.1016/j.lindif.2005.03.001

Passolunghi, M. C., & Siegel, L. C. (2001). Short-Term Memory, Working Memory, and Inhibitory Control in Children with Difficulties in Arithmetic Problem Solving. *Journal of Experimental Child Psychology*, *80*(1). https://doi.org/10.1006/jecp.2000.2626

Passolunghi, M. C., & Siegel, L. S. (2004). Working memory and access to numerical information in children with disability in mathematics. *Journal of Experimental Child Psychology*, *88*(4), 348–367. https://doi.org/10.1016/j.jecp.2004.04.002

Passolunghi, M. C., Vercelloni, B., & Schadee, H. (2007). The precursors of mathematics learning: Working memory, phonological ability and numerical competence. *Cognitive Development*, *22*(2), 165–184. https://doi.org/10.1016/j.cogdev.2006.09.001

Pelegrina, S., Justicia-Galiano, M. J., Martín-Puga, M. E., & Linares, R. (2020). Math anxiety and working memory updating: Difficulties in retrieving numerical information from working memory. *Frontiers in Psychology*, *11*, 669. https://doi.org/10.3389/fpsyg.2020.00669

Peng, P., Wang, T., Wang, C., & Lin, X. (2019). A meta-analysis on the relation between fluid intelligence and reading/mathematics: Effects of tasks, age, and social economics status. *Psychological Bulletin*, *145*(2), 189—236. https://doi.org/10.1037/bul0000182

Piaget, J. (1952). *The Child's Conception of Number*. Routledge & Paul.

Piaget, J., & Inhelder, B. (1973). *Memory and intelligence*. Routledge and Kegan Paul.

Pillow, B. H. (2002). Children's and adults' evaluation of the certainty of deductive inferences, inductive inferences, and guesses. *Child Development*, *73(3)*. https://doi.org/10.1111/1467-8624.00438

Pólya, G. (1945). *How to solve it; a new aspect of mathematical method.* (pp. xv, 204). Princeton University Press.

Pólya, G. (1981). *Mathematical discovery (combined version)*. Wiley.

Pongsakdi, N., Kajamies, A., Veermans, K., Lertola, K., Vauras, M., & Lehtinen, E. (2020). What makes mathematical word problem solving challenging? Exploring the roles of word problem characteristics, text comprehension, and arithmetic skills. *ZDM*, *52*(1), 33–44. https://doi.org/10.1007/s11858-019-01118-9

Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. Solso (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 55–85). Lawrence Erlbaum.

Powell, S. R. (2011). Solving Word Problems Using Schemas: A Review of the Literature. *Learning Disabilities Research & Practice*, *26*(2), 94–108. https://doi.org/10.1111/j.1540-5826.2011.00329.x

Quintero, A. H. (1983). Conceptual understanding in solving two-step word problems with a ratio. *Journal for Research in Mathematics Education (JRME)*, *14*(2), 102–112. https://doi.org/10.5951/jresematheduc.14.2.0102

Ramos, J. L., & Cuetos, F. (1999). *Batería de evaluación de los Procesos Lectores en los alumnos del tercer ciclo de Educación Primaria y Educación Secundaria Obligatoria, PROLEC-SE*. TEA Ediciones.

Re, A. M., Lovero, F., Cornoldi, C., & Passolunghi, M. C. (2016). Difficulties of children with ADHD symptoms in solving mathematical problems when information must be updated. *Research in Developmental Disabilities*, *59*. https://doi.org/10.1016/j.ridd.2016.09.001

Reed, J., & Kromrey, J. (2001). Teaching critical thinking in a community college history course: Empirical evidence from infusing Paul's model. *College Student Journal*, *35*(2), 201.

Resnick, L. B. (1989). Developing mathematical knowledge. *American Psychologist*, *44*(2), 162–169. https://doi.org/10.1037/0003-066X.44.2.162

Resnick, L. B. (1992). From protoquantities to operators: Building mathematical competence on a foundation of everyday knowledge. In G. Leinhardt, R. Putnam, & R. A. Hattrup (Eds.), *Analysis of arithmetic for mathematics teaching.* (pp. 373–429). Lawrence Erlbaum Associates, Inc.

Reusser, K. (1989). *Textual and situational factors in solving mathematical word problems*. University of Bern.

Ridderinkhof, K. R., van den Wildenberg, W. P. M., Segalowitz, S. J., & Carter, C. S. (2004). Neurocognitive mechanisms of cognitive control: The role of prefrontal cortex in action selection, response inhibition, performance monitoring, and reward-based learning. *Brain and Cognition*, *56*(2), 129–140. https://doi.org/10.1016/j.bandc.2004.09.016

Riley, M. S., & Greeno, J. G. (1988). Developmental Analysis of Understanding Language About Quantities and of Solving Problems. *Cognition and Instruction*, *5*(1), 49–101. https://doi.org/10.1207/s1532690xci0501_2

Riley, M. S., Greeno, J. G., & Heller, J. L. (1984). Development of children's problem-solving ability in arithmetic. In H. P. Ginsburg (Ed.), *The development of mathematical thinking* (pp. 153–196). Academic press.

Rogers, R. D., & Monsell, S. (1995). Costs of a predictible switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, *124*(2), 207–231. https://doi.org/10.1037/0096-3445.124.2.207

Rohmah, M., & Sutiarso, S. (2018). Analysis problem solving in mathematical using theory Newman. *EURASIA Journal of Mathematics, Science and Technology Education*, *14*(2), 671–681. https://doi.org/10.12973/ejmste/80630

Rojas-Barahona, C. A., Moreno-Ríos, S., García-Madruga, J. A., & Förster, C. E. (2021). Content of relationship, number of alternatives and working memory capacity in conditional inferences. *Current Psychology*. https://doi.org/10.1007/s12144-021-01966-3

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software, Articles*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Roth, B., Becker, N., Romeyke, S., Schäfer, S., Domnick, F., & Spinath, F. M. (2015). Intelligence and school grades: A meta-analysis. *Intelligence*, *53*. https://doi.org/10.1016/j.intell.2015.09.002

Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), *Theoretical Issues in Reading Comprehension: Perspectives from Cognitive Psychology, Linguistics, Artificial Intelligence and Education* (1st ed.). Lawrence Erlbaum Associates. https://doi.org/10.4324/9781315107493-4

Salthouse, T. A. (2005). Relations between cognitive abilities and measures of executive functioning. *Neuropsychology*, *19*(4), 532–545. https://doi.org/10.1037/0894-4105.19.4.532

Sánchez-Pérez, N., Fuentes, L. J., Pina, V., López-López, J. A., & González-Salinas, C. (2015). How do different components of Effortful Control contribute to children’s mathematics achievement? *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01383

Santamaría, C., Tse, P. P., Moreno-Ríos, S., & García-Madruga, J. A. (2013). Deductive reasoning and metalogical knowledge in preadolescence: A mental model appraisal. *Journal of Cognitive Psychology*, *25*(2). https://doi.org/10.1080/20445911.2012.743988

Schneider, W., & Löffler, E. (2016). The development of metacognitive knowledge in children and adolescents. In J. Dunlosky & S. K. Tauber (Eds.), *The Oxford handbook of metamemory.* (pp. 491–518). Oxford University Press.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*(1), 1–66. https://doi.org/10.1037/0033-295X.84.1.1

Schumacher, R. F., & Fuchs, L. S. (2012). Does understanding relational terminology mediate effects of intervention on compare word problems? *Journal of Experimental Child Psychology*, *111*(4), 607–628. https://doi.org/10.1016/j.jecp.2011.12.001

Schwaighofer, M., Bühner, M., & Fischer, F. (2016). Executive functions as moderators of the worked example effect: When shifting is more important than working memory capacity. *Journal of Educational Psychology*, *108*(7), 982–1000. https://doi.org/10.1037/edu0000115

Sebastián, N., Cuetos, F., Martí, M. A., & Carreiras, M. F. (2000). *LEXESP: Léxico informatizado del español*. Edicions de la Universitat de Barcelona.

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *298*(1089), 199–209. https://doi.org/10.1098/rstb.1982.0082

Shallice, T. (2002). Fractionation of the supervisory system. In D. T. Stuss & R. T. Knight (Eds.), *Principles of frontal lobe function.* (pp. 261–277). Oxford University Press.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, *84*(2), 127–190. https://doi.org/10.1037/0033-295X.84.2.127

Shum, D., Gill, H., Banks, M., Maujean, A., Griffin, J., & Ward, H. (2009). Planning ability following moderate to severe traumatic brain injury: Performance on a 4-disk version of the tower of london. *Brain Impairment*, *10*(3), 320–324. Cambridge Core. https://doi.org/10.1375/brim.10.3.320

Shum, H. Y., & Chan, W. W. L. (2020). Young children’s inhibition of keyword heuristic in solving arithmetic word problems. *Human Behaviour and Brain*, *1*(2), 43. https://doi.org/10.37716/HBAB.2020010202

Siegler, R., & Jenkins, E. A. (1989). *How children discover new strategies*. Lawrence Erlbaum Associates.

Sikora, D. M., Haley, P., Edwards, J., & Butler, R. W. (2002). Tower of London test performance in children with poor arithmetic skills. *Developmental Neuropsychology*, *21*(3), 243–254. https://doi.org/10.1207/S15326942DN2103_2

Simms, N. K., Frausel, R. R., & Richland, L. E. (2018). Working memory predicts children's analogical reasoning. *Journal of Experimental Child Psychology*, *166*, 160–177. https://doi.org/10.1016/j.jecp.2017.08.005

Sinayev, A., & Peters, E. (2015). Cognitive reflection vs. Calculation in decision making. *Frontiers in Psychology*, *6*, 532. https://doi.org/10.3389/fpsyg.2015.00532

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*(1). https://doi.org/10.1037/0033-2909.119.1.3

Snow, C. E., & Sweet, A. P. (2003). Reading for comprehension. In A. P. Sweet & C. E. Snow (Eds.), *Rethinking reading comprehension* (pp. 1–10). Guilford Press.

Sorel, O., & Pennequin, V. (2008). Aging of the planning process: The role of executive functioning. *Brain and Cognition*, *66*(2), 196–201. https://doi.org/10.1016/j.bandc.2007.07.006

Spencer, J. P. (2020). The development of working memory. *Current Directions in Psychological Science*, *29*(6), 545–553. https://doi.org/10.1177/0963721420959835

St Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *Quarterly Journal of Experimental Psychology*, *59(4)*, 745–759. https://doi.org/10.1080/17470210500162854

Stacey, K., & MacGregor, M. (1999). Learning the algebraic method of solving problems. *The Journal of Mathematical Behavior*, *18*(2), 149–167. https://doi.org/10.1016/S0732-3123(99)00026-7

Stanovich, K. E. (2006). Fluid intelligence as cognitive decoupling. *Behavioral and Brain Sciences*, *29*(2), 139–140. Cambridge Core. https://doi.org/10.1017/S0140525X06359031

Stanovich, K. E., & Toplak, M. E. (2012). Defining features versus incidental correlates of Type 1 and Type 2 processing. *Mind and Society*, *11*(1). https://doi.org/10.1007/s11299-011-0093-6

Stanovich, K. E., & West, R. F. (2000a). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, *23*(5), 645–665. https://doi.org/10.1017/s0140525x00003435

Stanovich, K. E., & West, R. F. (2000b). Individual differences in reasoning: Implications for the rationality debate? In *Behavioral and Brain Sciences* (Vol. 23). https://doi.org/10.1017/S0140525X00003435

Starkey, P., & Cooper, R. G. (1980). Perception of numbers by human infants. *Science*, *210*(4473), 1033. https://doi.org/10.1126/science.7434014

Starkey, P., Spelke, E. S., & Gelman, R. (1983). Detection of intermodal numerical correspondences by human infants. *Science*, *222*(4620), 179. https://doi.org/10.1126/science.6623069

Starkey, P., Spelke, E. S., & Gelman, R. (1990). Numerical abstraction by human infants. *Cognition*, *36*(2), 97–127. https://doi.org/10.1016/0010-0277(90)90001-Z

Stern, E. (1993). What makes certain arithmetic word problems involving the comparison of sets so difficult for children? *Journal of Educational Psychology*, *85*(1), 7. https://doi.org/10.1037/0022-0663.85.1.7

Stern, E., & Lehrndorfer, A. (1992). The role of situational context in solving word problems. *Cognitive Development*, *7*(2), 259–268. https://doi.org/10.1016/0885-2014(92)90014-I

Sternberg, R. J. (1996). What is mathematical thinking? In R. J. Sternberg & T. Ben-Zeev (Eds.), *The Nature of Mathematical Thinking* (pp. 303–318). Lawrence Erlbaum Associates.

Sternberg, R. J., & Ben-Zeev, T. (1996). *The nature of mathematical thinking*. Lawrence Erlbaum Associates. https://doi.org/10.4324/9780203053270

Sternberg, R. J., & Frensch, P. A. (1992). On being an expert: A cost-benefit analysis. In R. Hoffman (Ed.), *The psychology of expertise* (pp. 191–203). Springer.

Süß, H. M., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working-memory capacity explains reasoning ability—And a little bit more. *Intelligence*, *30*(3), 261–288. https://doi.org/10.1016/S0160-2896(01)00100-3

Swanson, H. L. (1990). Influence of metacognitive knowledge and aptitude on problem solving. *Journal of Educational Psychology*, *82*(2), 306–314. https://doi.org/10.1037/0022-0663.82.2.306

Swanson, H. L. (1994). Short-term memory and Working Memory: Do both contribute to our understanding of academic achievement in children and adults with learning disabilities? *Journal of Learning Disabilities*, *27*(1), 34–50. https://doi.org/10.1177/002221949402700107

Swanson, H. L. (2006). Cross-sectional and incremental changes in working memory and mathematical problem solving. *Journal of Educational Psychology*, *98*(2), 265–281. https://doi.org/10.1037/0022-0663.98.2.265

Swanson, H. L., Cochran, K. F., & Ewers, C. A. (1990). Can learning disabilities be determined from working memory performance? *Journal of Learning Disabilities*, *23*(1), 59–67. https://doi.org/10.1177/002221949002300113

Swanson, H. L., & Fung, W. (2016). Working memory components and problem-solving accuracy: Are there multiple pathways? *Journal of Educational Psychology*, *108*(8), 1153–1177. https://doi.org/10.1037/edu0000116

Swanson, H. L., & Sachse-Lee, C. (2001). Mathematical problem solving and working memory in children with learning disabilities: Both executive and phonological processes are important. *Journal of Experimental Child Psychology*, *79*(3), 294–321. https://doi.org/10.1006/jecp.2000.2587

Szaszi, B., Szollosi, A., Palfi, B., & Aczel, B. (2017). The cognitive reflection test revisited: Exploring the ways individuals solve the test. *Thinking & Reasoning*, *23*(3), 207–234. https://doi.org/10.1080/13546783.2017.1292954

Thevenot, C. (2010). Arithmetic word problem solving: Evidence for the construction of a mental model. *Acta Psychologica*, *133*(1). https://doi.org/10.1016/j.actpsy.2009.10.004

Thevenot, C., & Barrouillet, P. (2015). Arithmetic word problem solving and mental representations. In R. Cohen Kadosh & A. Dowker (Eds.), *The Oxford handbook of numerical cognition* (pp. 58–179). Oxford University Press.

Thevenot, C., Fanget, M., & Fayol, M. (2007). Retrieval or nonretrieval strategies in mental arithmetic? An operand recognition paradigm. *Memory & Cognition*, *35*(6), 1344–1352. https://doi.org/10.3758/bf03193606

Thevenot, C., & Oakhill, J. (2005). The strategic use of alternative representations in arithmetic word problem solving. *The Quarterly Journal of Experimental Psychology Section A*, *58*(7), 1311–1323. https://doi.org/10.1080/02724980443000593

Thompson, J. M., Nuerk, H. C., Moeller, K., & Cohen Kadosh, R. (2013). The link between mental rotation ability and basic numerical representations. *Acta Psychologica*, *144*(2), 324–331. https://doi.org/10.1016/j.actpsy.2013.05.009

Thompson, V. A. (2010). Towards a metacognitive dual process theory of conditional reasoning. In M. Oaksford & N. Chater (Eds.), *Cognition and conditionals: Probability and logic in human thinking.* (pp. 335–354). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199233298.003.0018

Thompson, V. A., Evans, J. St. B. T., & Campbell, J. I. D. (2013). Matching bias on the selection task: It's fast and feels good. *Thinking & Reasoning*, *19*(3–4), 431–452. https://doi.org/10.1080/13546783.2013.820220

Thompson, V. A., Prowse Turner, J. A., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology*, *63*(3), 107–140. https://doi.org/10.1016/j.cogpsych.2011.06.001

Tian, Y., Fang, Y., & Li, J. (2018). The Effect of Metacognitive Knowledge on Mathematics Performance in Self-Regulated Learning Framework—Multiple Mediation of Self-Efficacy and Motivation. *Frontiers in Psychology*, *9*, 2518. https://doi.org/10.3389/fpsyg.2018.02518

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

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, *39*(7), 1275. https://doi.org/10.3758/s13421-011-0104-1

Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning*, *20*(2), 147–168. https://doi.org/10.1080/13546783.2013.844729

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207–232. https://doi.org/10.1016/0010-0285(73)90033-9

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, *185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Uittenhove, K., & Lemaire, P. (2015). Numerical cognition during cognitive aging. In R. C. Kadosh & A. Dowker (Eds.), *The Oxford handbook of numerical cognition.* (pp. 345–364). Oxford University Press.

Unterrainer, J. M., & Owen, A. M. (2006). Planning and problem solving: From neuropsychology to functional neuroimaging. *Brain Imaging in Neurosciences - An*

Interdisciplinary Approach, *99*(4), 308–317. https://doi.org/10.1016/j.jphysparis.2006.03.014

van der Stel, M., & Veenman, M. V. J. (2014). Metacognitive skills and intellectual ability of young adolescents: A longitudinal study from a developmental perspective. *European Journal of Psychology of Education*, *29*(1), 117–137. https://doi.org/10.1007/s10212-013-0190-5

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. Academic Press.

Veenman, M. V. J., Van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning*, *1*(1), 3–14. https://doi.org/10.1007/s11409-006-6893-0

Verschaffel, L. (1994). Using retelling data to study elementary school children’s representations and solutions of compare problems. *Journal for Research in Mathematics Education*, *25*(2). https://doi.org/10.2307/749506

Verschaffel, L., De Corte, E., & Pauwels, A. (1992). Solving compare problems: An eye movement test of Lewis and Mayer’s consistency hypothesis. *Journal of Educational Psychology*, *84*(1), 85–94. https://doi.org/10.1037/0022-0663.84.1.85

Verschaffel, L., Depaepe, F., & Van Dooren, W. (2020). Word Problems in Mathematics Education. In S. Lerman (Ed.), *Encyclopedia of Mathematics Education* (pp. 908–911). Springer International Publishing. https://doi.org/10.1007/978-3-030-15789-0_163

Vicente, S., Orrantia, J., & Verschaffel, L. (2008). Influence of mathematical and situational knowledge on arithmetic word problem solving: Textual and graphical aids. *Journal for the Study of Education and Development*, *31*(4), 463–483. https://doi.org/10.1174/021037008786140959

Vidal-Abarca, E., Mañá, A., & Gil, L. (2010). Individual differences for self-regulating task-oriented reading activities. *Journal of Educational Psychology*, *102*(4), 817–826. https://doi.org/10.1037/a0020062

Vilenius-Tuohimaa, P. M., Aunola, K., & Nurmi, J. E. (2008). The association between mathematical word problems and reading comprehension. *Educational Psychology*, *28*(4), 409–426. https://doi.org/10.1080/01443410701708228

Viterbori, P., Traverso, L., & Usai, M. C. (2017). The role of executive function in arithmetic problem-solving processes: A study of third graders. *Journal of Cognition and Development*, *18*(5), 595–616. https://doi.org/10.1080/15248372.2017.1392307

Vula, E., Avdyli, R., Berisha, V., Saqipi, B., & Elezi, S. (2017). The impact of metacognitive strategies and self-regulating processes of solving math word problems. *International Electronic Journal of Elementary Education*, *10*(1), 49–59. https://doi.org/10.26822/iejee.2017131886

Wallas, G. (1926). *The art of thought*. Harcourt Brace Jovanovich.

Wang, T., Li, C., Wei, W., & Schweizer, K. (2020). An investigation on how inhibition in cognitive processing contributes to fluid reasoning. *Advances in Cognitive Psychology*, *16*(3), 176–185. PubMed. https://doi.org/10.5709/acp-0295-7

Watt, H. M. G., Shapka, J. D., Morris, Z. A., Durik, A. M., Keating, D. P., & Eccles, J. S. (2012). Gendered motivational processes affecting high school mathematics participation, educational aspirations, and career plans: A comparison of samples from Australia, Canada, and the United States. *Developmental Psychology*, *48*(6). https://doi.org/10.1037/a0027838

Weil, L. G., Fleming, S. M., Dumontheil, I., Kilford, E. J., Weil, R. S., Rees, G., Dolan, R. J., & Blakemore, S. J. (2013). The development of metacognitive ability in adolescence. *Consciousness and Cognition*, *22*(1), 264–271. https://doi.org/10.1016/j.concog.2013.01.004

Wertheimer, M. (1938). Gestalt theory. In W. D. Ellis (Ed.), *A source book of Gestalt psychology.* (pp. 1–11). Kegan Paul, Trench, Trubner & Company. https://doi.org/10.1037/11496-001

Widodo-Yulianto, F. (2020). Students' metacognitive skills in solving word problem. *Journal of Physics: Conference Series*, *1470*, 012090. https://doi.org/10.1088/1742-6596/1470/1/012090

Wilcox, R. R. (2011). *Introduction to robust estimation and hypothesis testing*. Academic press.

William, S. K., & Maat, S. M. (2020). Understanding Students' Metacognition in Mathematics Problem Solving: A Systematic Review. *International Journal of Academic Research in Progressive Education and Development*, *9*(3), 115–127. https://doi.org/10.6007/IJARPED/v9-i3/7847

Williams, R. L., & Worth, S. L. (2001). The relationship of critical thinking to success in college. *Inquiry: Critical Thinking across the Disciplines*, *21*(1), 5–16. https://doi.org/10.5840/inquiryctnews200121123

Wyndhamn, J., & Säljö, R. (1997). Word problems and mathematical reasoning—A study of children's mastery of reference and meaning in textual realities. *Learning and Instruction*, *7*(4), 361–382. https://doi.org/10.1016/S0959-4752(97)00009-1

Wynn, K. (1995). Origins of numerical knowledge. *Mathematical Cognition*, *1*(1), 35–60.

Yang, S. C. (2007). E-critical/thematic doing history project: Integrating the critical thinking approach with computer-mediated history learning. *Computers in Human Behavior*, *23*(5). https://doi.org/10.1016/j.chb.2006.02.012

Yeniad, N., Malda, M., Mesman, J., van IJzendoorn, M. H., & Pieper, S. (2013). Shifting ability predicts math and reading performance in children: A meta-analytical study. *Learning and Individual Differences*, *23*, 1–9. https://doi.org/10.1016/j.lindif.2012.10.004

Yin, L., Shi, Z., Liao, Z., Tang, T., Xie, Y., & Peng, S. (2020). The Effects of Working Memory and Probability Format on Bayesian Reasoning. *Frontiers in Psychology*, *11*, 863. https://doi.org/10.3389/fpsyg.2020.00863

Yuan, P., & Raz, N. (2014). Prefrontal cortex and executive functions in healthy adults: A meta-analysis of structural neuroimaging studies. *Neuroscience & Biobehavioral Reviews*, *42*, 180–192. https://doi.org/10.1016/j.neubiorev.2014.02.005

Zamary, A., Rawson, K. A., & Was, C. A. (2019). Do complex span and content-embedded working memory tasks predict unique variance in inductive reasoning? *Behavior Research Methods*, *51*(6), 2546–2558. https://doi.org/10.3758/s13428-018-1104-x

7. ANEXO

Tabla 1. Descripción de los problemas aritméticos verbales (AWP) utilizados en el Estudio 1 y 3: enunciado del problema, operaciones aritméticas para resolverlo, número de operaciones, inconsistencias y posibles respuestas superficiales.

#	Problem	Arithmetic	# Ops	Inconsistences	Superficial Responses
1	In Mercadona, a bottle of Rioja wine costs 20 Euros. In Ahorramás, the same bottle of wine costs 2 Euros less than in Mercadona. If you need to buy 4 bottles of this Rioja wine, how much will you have to pay in Ahorramás?	(20 - 2) x 4 = 72	2	No	--
2	In SportGood, NK brand sneakers cost 50 Euros. This is 3 Euros more than what you have to pay for the same shoes at Primark. If you need to buy 4 pairs of NK sneakers, how much will you have to pay at Primark?	(50 - 3) x 4 = 188	2	Yes	(50 + 3) x 4 = 212
3	In Media-Markt, a Samsung digital tablet costs 70 Euros and a mini SG tablet costs 6 Euros less. In Mundo-Electrónico, the same digital tablet costs 4 Euros less than in Media-Markt. If in an educational institute they need to buy 5 mini SG tablets, how much will they have to pay at Mundo-Electrónico?	(70 - 6 - 4) x 5 = 300	3	No – No	--
4	In Fnac, a box of coloured pencils costs 18 Euros. This is 2 Euros more than what you have to pay for a box of coloured pencils at the Folder bookstore. A box of crayons costs 4 Euros less than a box of coloured pencils in the Folder bookshop. If you need to buy 3 boxes of crayons, how much will you have to pay at the Folder bookstore?	(18 - 2 - 4) x 3 = 36	3	Yes – No	(18 + 2 - 4) x 3 = 48
5	In Toys-r-Us, a plastic skateboard costs 30 Euros. This is 5 Euros more than what you have to pay for a plastic skateboard at Decathlon. In Decathlon, a plastic skateboard costs 7 Euros less than a carbon fibre skateboard. If you need to buy 2 carbon fibre skateboards, how much will you have to pay in Decathlon?	(30 - 5 + 7) x 2 = 64	3	Yes – Yes	Sup1: (30 + 5 + 7) x 2 = 84 Sup2: (30 - 5 - 7) x 2 = 54 Sup3: (30 + 5 - 7) x 2 = 56

Tabla 2. Descripción de los problemas aritméticos verbales (AWP), *versión menos*, utilizados en el Estudio 2: enunciado del problema, operaciones aritméticas para resolverlo, número de operaciones, inconsistencias y posibles respuestas superficiales

#	Problem	Arithmetic	# Ops	Inconsistences	Superficial Responses
1	Da *Conad*, una bottiglia grande di Coca-Cola costa 2,20 euro. La stessa bottiglia di Coca-Cola costa 20 centesimi di meno da *Carrefour*. Quanto costa al *Carrefour* una bottiglia di Coca-Cola?	2.20 – 0.20 = 2.00	1	No	--
2	Da *Intersport*, una palla di calcio costa 8 euro. Questo prezzo è di 3 euro in meno rispetto a quello della stessa palla di calcio da *Decathlon*. Quanto costa da *Decathlon* una palla di calcio?	8 + 3 = 11	1	Yes	8 - 3 = 5
3	Da *Mediaworld*, un tablet normale costa 46 euro e un tablet mini costa 4 euro meno. Lo stesso tablet mini costa da *Unieuro* 1 euro meno che da *Mediaworld*. Quanto costa da *Unieuro* un tablet mini?	(46 - 4) - 1 = 41	2	No	--
4	Da *Buffetti* una scatola di pennarelli normali costa 15 euro. Questo prezzo è di 5 euro inferiore rispetto a quello che pagheresti per la stessa scatola di pennarelli normali presso *Cartorama*. Da *Cartorama*, una scatola di pennarelli magnetici costa 2 euro in più rispetto alla scatola dei pennarelli normali. Quanto costa una scatola di pennarelli magnetici da *Cartorama*?	(15 + 5) + 2 = 22	2	Yes - No	(15 - 5) + 2 = 12
5	Nel negozio *GameStop*, un joystick wireless per una console per videogiochi costa 30 euro. Lo stesso joystick wireless costa 6 euro in meno da *Expert*. Da *Expert*, il joystick wireless costa 5 euro in più rispetto alle cuffie con microfono. Quanto costano le cuffie con microfono da *Expert*?	(30 - 6) - 5 = 19	2	No - Yes	(30 - 6) + 5 = 29
6	Da *Decathlon* un normale orologio per il nuoto costa 25 euro. Questo prezzo è di 3 euro inferiore rispetto a quello dello stesso orologio su *Amazon*. L' orologio normale da nuoto su *Amazon* costa 2 euro in più rispetto a un mini orologio da nuoto *Apple*. Quanto costa su *Amazon* un mini orologio da nuoto *Apple*?	(25 + 3) - 2 = 26	2	Yes - Yes	Sup1: (25 - 3) - 2 = 20 Sup2: (25 + 3) + 2 = 30 Sup3: (25 - 3) + 2 = 24

Tabla 3. Descripción de los problemas aritméticos verbales (AWP), *versión menos*, utilizados en el Estudio 2: enunciado del problema, operaciones aritméticas para resolverlo, número de operaciones, inconsistencias y posibles respuestas superficiales

#	Problem	Arithmetic	# Ops	Inconsistences	Superficial Responses
1	Da Conad, una bottiglia grande di Coca-Cola costa 2,20 euro. La stessa bottiglia di Coca-Cola costa 20 centesimi in più da Carrefour. Quanto costa al Carrefour una bottiglia di Coca-Cola?	2.20 + 0.20 = 2.40	1	No	--
2	Da Intersport, una palla di calcio costa 8 euro. Questo prezzo è di 3 euro superiore a quello della stessa palla di calcio da Decathlon. Quanto costa da Decathlon una palla di calcio?	8 - 3 = 5	1	Yes	8 + 3 = 11
3	Da Mediaworld, un tablet normale costa 46 euro e un tablet mini costa 4 euro in più. Lo stesso tablet mini costa da Unieuro 1 euro in più che da Mediaworld. Quanto costa da Unieuro un tablet mini?	(46 + 4) + 1 = 51	2	No	--
4	Da Buffetti una scatola di pennarelli normali costa 15 euro. Questo prezzo è di 5 euro superiore a quello della stessa scatola di pennarelli presso Cartorama. Da Cartorama, una scatola di pennarelli magnetici costa 2 euro in meno rispetto alla scatola dei pennarelli normali. Quanto costa una scatola di pennarelli magnetici da Cartorama?	(15 - 5) + 2 = 12	2	Yes - No	(15 + 5) + 2 = 22
5	Nel negozio GameStop, un joystick wireless per una console per videogiochi costa 30 euro. Lo stesso joystick wireless costa 6 euro in più da Expert. Da Expert, il joystick wireless costa 5 euro in meno rispetto alle cuffie con microfono. Quanto costano le cuffie con microfono da Expert?	(30 - 6) + 5 = 29	2	No - Yes	(30 - 6) - 5 = 19
6	Da Decathlon un orologio normale per il nuoto costa 25 euro. Questo prezzo è di 3 euro superiore rispetto a quello dello stesso orologio su Amazon. L' orologio normale da nuoto su Amazon costa 2 euro in meno rispetto ad un orologio da nuoto di base Apple. Quanto costa su Amazon un orologio da nuoto di base Apple?	(25 - 3) + 2 = 24	2	Yes - Yes	Sup1: (25 + 3) + 2 = 30 Sup2: (25 - 3) - 2 = 20 Sup3: (25 + 3) - 2 = 26

Tabla 4. Descripción de los problemas de reflexión cognitiva (CRT) utilizados en el Estudio 3: enunciado del problema, operaciones aritméticas para resolverlo, respuesta correcta y posibles respuestas superficiales.

#	Problem	Arithmetic	Correct Response	Superficial Response
1	En unas rebajas has comprado una pala y una pelota y las dos te han costado 1.10 Euros en total. La pala te ha costado 1.00 Euro más que la pelota ¿Cuánto te ha costado la pelota?	x + x = (1.10 – 1.00)	0.05	0.10
2	Has comprobado que 5 niños tardan 5 minutos en afilar 5 lápices ¿Cuánto tardarían 100 niños en afilar 100 lápices?	x = 5	5	100
3	En la montaña hay un lago que en verano siempre está siempre cubierto de flores. Tu maestra te ha dicho que al llegar la primavera las flores tardan en ocupar toda su superficie 48 días. También te ha dicho que la zona con flores del lago aumenta el doble su tamaño cada día ¿Cuántos días tardan en ocupar la mitad del lago?	48 - 1	47	24