Doctoral Dissertation

Metacognitive processes in bilingual second language learners

Procesos metacognitivos en bilingües aprendices en un segundo idioma

International PhD candidate

Marta Reyes Sánchez

Advisors

Mª Julia Morales Castillo

Mª Teresa Bajo Molina

PhD in Psychology

Department of Experimental Psychology

University of Granada

December, 2023

Editor: Universidad de Granada. Tesis Doctorales
Autor: Marta Reyes Sánchez
ISBN: 978-84-1195-206-4
URI: https://hdl.handle.net/10481/89858

A mami,
que siempre quiso que escribiera un libro.
No sé si esto es lo que esperaba, pero se lo dedico.

A papi,
por ser mi liebre favorita en todos los aspectos de la vida.

Agradecimientos

En el amplio panorama de la vida, mi etapa pre-doctoral ha sido probablemente, uno de los mayores retos que he enfrentado jamás. Quizás haya sido como emprender una carrera en un terreno montañoso técnico que requiere de velocidad, resistencia y determinación. O como un triatlón, donde la concentración, la disciplina y la precisión se ponen a prueba. Durante estos últimos cuatro años, he sentido la adrenalina de la investigación mientras me adentraba en una prueba repleta de subidas y bajadas. Cada experimento, cada análisis y cada palabra escrita eran pasos hacia la línea de meta, impulsados por la pasión, la persistencia y el deseo de superarme a mí misma. Siempre he dicho que el deporte me hace (más) fuerte, (más) valiente y (más) segura, habilidades trasversales que me han ayudado a trabajar incansablemente en mi tesis doctoral.

Así como en el deporte se necesita perseverancia y dedicación, no me habría sido posible completar la tesis sin el apoyo inquebrantable de mi equipo de entrenamiento: directoras, compañeras/os, familia y amigas/os. Con gratitud y humildad, celebro este hito en mi carrera académica y agradezco a todas las personas que me han animado y acompañado en cada kilómetro de este desafiante reto. Me siento muy afortunada de cruzar la meta rodeada de tan bonitas personas.

Gracias a mis directoras Teresa y Julia. Desde luego que sin vosotras no hubiera sido capaz de resistir la importante fatiga. Gracias, Teresa, por tu sabiduría, por saber leer mis necesidades y ofrecerme ayuda siempre. Gracias por potenciar incansablemente mi autonomía e independencia. Gracias, Julia, por tu apoyo pese a todo. Gracias por mentorizarme durante tu (bi)maternidad. Gracias por tu empatía y por mostrarte siempre disponible.

Me siento muy afortunada de haber podido crecer profesional (y personalmente) en el CIMCYC; muchas de las horas de trabajo se han desarrollado en este espacio que tiene una calidad humana inmensurable. Gracias a la gente de Neurociencia, por tan bonita acogida desde el primer momento y por mostrar tanto cariño. Chema, por tu franqueza, sinceridad y humor. Alba, por tu amistad más allá de la ciencia. Isma, por tus abrazos reconfortantes. Luis, por tu sosiego y conversaciones trascendentales. Carmen, por tu alegría y sonrisas infinitas. Greta, por la buena energía. Josu, por tu sinceridad y la positividad que aportas. Rafa, por ser Córdoba allí donde estés. Mar, por tu lucha feminista incansable, que me contagia y hace reflexionar. Cristina, por tu dulzura. Klara, por expresar tus dudas y tu arte. Ana Paqui, por tu optimismo y buen vivir. Omar, por enseñarnos otros caminos posibles. Belén, por compartir con entusiasmo y pasión. Georgia, por contagiar con tu expresividad. Gracias a la gente de Psicología Social: Sofía, Fran, Rocío, Danna, por su compromiso e iniciativas. Gracias a los *sex* Ana, Óscar y Laura, por su cercanía.

Gracias a la gente de 3MT. Esta experiencia ha marcado significativamente mi etapa pre-doctoral. Gracias, Susana y Carlos, por vuestras sugerencias y por ser tan buenas guías. Gracias compis, por la ayuda mutua y por crear un clima competitivo tan sano. Gracias en especial a Manu, por haber compartido kilómetros (literal y metafóricamente) durante algo más de 3 minutos. Gracias por haber fomentado un vínculo tan bonito entre Vera, tú y yo, alrededor de la montaña. Gracias por animarme y empujarme a seguir haciendo ciencia.

Sin duda, gracias superlativas a mi familia *sientífica*: el grupo Memoria y Lenguaje. Gracias Nuria Aguerre, por tus consejos de hermana mayor y por alimentar mi lado *hippie.* Gracias, Tania Valle, por enseñarme tanto, por ser una hermana mayor discreta y muy presente. Marta y Antonio, vuestro *tándem* es la esencia del grupo. Gracias, Marta, amiga, por estar, por estar siempre. Gracias, Antonio, por entenderme con la mirada, aconsejarme y apoyarme. Gracias, Cristina, por convivir, respetar y ser tan buen ejemplo en la tesis y en la vida. Gracias, Chus, Alba y Ana Belén, por ser referencia. Me siento afortunada de teneros de vuelta y haber podido *correr* mi última etapa con vosotras. Gracias, Raquel, por preocuparte por mí siempre (y por ser la *tita* que consiente a Vera). Gracias, Guille, por tu ingenio y autenticidad. Gracias, Melodie, por mostrar otras vías. Gracias, Miguel y Filip, por vuestro buen humor y buena disposición. Gracias, Fernando, por ser ejemplo de resiliencia en la etapa final de tesis. Gracias, Anikó y José Luis, por incorporaros al grupo con tanta energía positiva. Gracias, Nuria Montoro, por enseñarnos que la maternidad también tiene cabida en los últimos meses predoctorales. Gracias, amigas/os, por crear tantos recuerdos dentro y fuera del laboratorio. Ojalá ese espíritu de compañerismo y buen querer,

perdure y trascienda muchos años más. Gracias también a los estratos superiores, porque este grupo (y alrededores) se compone de grandes personas y profesionales. Gracias, Daniela, Antonio Bernal, Sandra, José, Ana Pérez, Carlos Gómez-Ariza, Pedro Macizo, Ian G. Simpson, Juan Linde, Carmen Aguirre, Patri Román, Luis Morales.

Mi gratitud cruza el Atlántico. Un trocito de mi corazón se quedó en Montreal. Todavía me recreo en los recuerdos de aquella gélida experiencia. Gracias Lola, por aparecer en el mejor momento y amortiguar la creciente ansiedad causada por un vuelo perdido rumbo a Canadá. Gracias, Cris y Nerea, por vuestra energía joven y vuestra madurez. Gracias, Héctor, por ofrecer siempre el *house mate* como refugio y por enseñarme que diversificar el amor, lo multiplica. Many thanks, Can, for such a warm welcome, for sharing a little time each evening over dinner, which made me feel at home. Thanks for contributing to make my stay a beautiful memory. Thanks, David, for being the perfect snail mail companion, for your compliments, and for teaching me to be “a one-day-old PhD student”. Thanks, Gigi, for letting me be at your lab, despite all the restrictions due to COVID.

Gracias a mi familia. Papá y mamá, gracias por ser mi avituallamiento emocional favorito. Gracias por vuestra paciencia y por mostrar interés en todo lo que hago. Gracias por haberme educado en valores de persistencia, humildad y trabajo, que me han traído hasta aquí. Gracias, hermana, por ser luz, diversión y risas. Gracias por aceptar(me) y amar(me). Gracias, Gabri, por quererme *tata*. Gracias, Bruno por tus abrazos y tu nobleza. Gracias, Vera por desafiar(me), desordenar(me) y flexibilizar(me).

Abuela Isabel, gracias por preguntarme cada día “¿cuándo vienes?” o “¿cuánto te queda en Granada?”. Qué forma más bonita tienes de expresar tu deseo de que esté cerca del calor familiar.

Gracias tito Juan Carlos, por estar siempre tan orgulloso de tu sobrina. Gracias, abuela Carmen por quererme tanto (y reñir a veces) y abuelo Rafael por tus bromas y tus risas. Gracias, tita Mamen, por tu cariño; tita Rosi, por tu sensatez y tu saber vivir; tito Antonio, por tus piropos y halagos; primas Alba y Blanca, por vuestra dulzura; primos Miguel Ángel y Fco. Javier, por querer esforzaros y hacer vuestro camino. Gracias, padrino, por creer en mi desde que ni recuerdo. Gracias, prima (y amiga) Isa, por todo, por ser y estar; por ser familia y amiga, por estar siempre presente.

Gracias a mis amigas de Córdoba, las de siempre, las de toda la vida. Gracias por apoyarme incondicionalmente, mostrarme vuestra admiración, respetarme y aceptar mis largas ausencias durante este intenso período. Vuestra comprensión y paciencia han sido fundamentales para mantener el equilibrio entre mis responsabilidades académicas y nuestra amistad sincera. Gracias, Lorena Jiménez, por tus asesorías jurídicas y por contagiar tu buena filosofía de vida. Gracias, Lorena Alcántara, por facilitarme la llegada a Granada y estar en todas partes conmigo. Gracias, Gloria, por estar siempre dispuesta a reír y bailar. Gracias Eli, porque siempre tienes palabras tiernas y dulces para regalarme, y por ser mi sirena de referencia. Gracias, Paloma, por sonreír venga lo que venga. Gracias, Virginia, por confiar siempre en mí, mostrar tu orgullo y ser nuestra médica de confianza. Gracias, Irene, por abrir camino con valentía. Gracias, Silvia, por tus palabras de aliento. Gracias, María, por estar siempre dispuesta a celebrar mis logros. Gracias María Jesús, por dialogar desde lo más profundo. Gracias Alba, por tu lealtad.

Gracias también a la familia mercedaria de Córdoba, mucho de lo que soy se lo debo a esa bonita comunidad. Gracias por haberme visto crecer, y haberme querido acompañar durante muchos años después. Gracias en especial a Manuel, (tito) Antonio Luis, (amiga) Ana Belén y (el gran compi) Paquito. Gracias. Que el cole se siga acordando de mí, me enorgullece y emociona a partes iguales.

Durante este tiempo de tesis, la vida también me ha regalado momentos bonitos con personas vitamina que me han impulsado y reseteado. Gracias a Mª Ángeles y Almudena, por ser vía de escape, por comprenderme aun siendo ajenas a este mundo de la investigación, por reírnos de ser *pardas*, por mostrarme apoyo y escuchar mis audios infinitos (¿o son *podcast*?). Gracias a Tania y Bea, por despertarme pasión por las carreras por montaña y por reestructurar mis creencias (infantiles) de amistad. Gracias a Óscar, por acompañarme en el camino e introducir el triatlón en mi vida. Gracias a Vicen, por nuestra amistad, sin más. Gracias a Simón, por darme aliento y crear una bonita tradición de almuerzo en el que se ha convertido en mi bar favorito de Granada (@cafébarlagoma). Gracias a Soraya y Vicky por regalarme un viaje eterno a nuestra infancia y adolescencia. Estar con vosotras es revivir momentos mágicos. Gracias por seguir veinte años después. Gracias a Sergio por impulsarme a muchas aventuras deportivas, y ser un buen referente de cómo hacer una vida (feliz) en torno a la investigación. Gracias a Maribel, por ser hermana. Gracias a Anabel, por ser tan espontánea y estar siempre dispuesta a compartir. Gracias a Belén y Miguel, por comprenderme tan bien en tantos aspectos de mi vida.

Gracias (miles) a Juanma, por tenerme paciencia, sacarme de mi zona de *confort*, invitarme a la pista y crearme antojo de fresas cada día. Gracias por haberte atrevido a acompañarme durante estos últimos e intensos kilómetros de tesis. Gracias por dejarme espacio, hacerme reír y quererme bonito.

Gracias también a las personas que ya no están, pero que me han acompañado en este gran desafío. Tania, amiga, todavía me pregunto por qué te fuiste tan pronto. Tenías un proyecto de vida por delante para disfrutar que, lamentablemente, dejaste a medias. No puedo entrar al CIMCYC sin pensar en ti. Gracias por cuidarme desde el primer día que llegué a Granada cargada de (bastantes) inseguridades y miedos. Tu despedida me causó mucho dolor. Gracias, supongo, por enseñarme a *vivir hoy.* Abuelo Juan, tú también marchaste antes de tiempo. Te he echado de menos en muchos momentos importantes. Hoy es otro de ellos. Gracias, porque tu recuerdo me hace sonreír.

No me cabe más gratitud. Cruzo la meta con los brazos en alto, llena de orgullo, satisfacción y amor. Gracias.

Marta Reyes.

Contents

Prefacio

La realidad es que mi vida ha estado rodeada de idiomas desde la infancia temprana. Bien charlatana desde pequeña, recuerdo jugar en mi habitación a "las profesoras" mientras "hablaba otros idiomas". Paradójicamente, a pesar de haber crecido en un entorno monolingüe, mis padres sintieron motivación para matricularme en clases de inglés a los seis años recién cumplidos. Porque claro, "aprender un segundo idioma es muy importante". Imagino que, al verme tan parlanchina, creyeron que sería de mi interés. O quizás, que me haría tener mejor porvenir. Honestamente, dudo que en aquel momento tuvieran conocimiento de las consecuencias directas e indirectas, sociales, cognitivas y personales, derivadas del aprendizaje y uso de un segundo idioma. Lo que estaba claro es que "al inglés hay que ir, Marta". Y Marta, capaz de sacarle el gusto a cualquier cosa, veinte años después, no solo sigue hablando inglés, sino que realiza con gran interés, una tesis en torno a eso.

El trasfondo poético de cómo y por qué he llegado hasta aquí lo desconozco, (¿casualidad, destino?), lo que está claro es que nadie imaginaba que aquella niña que pronunciaba palabras inventadas intentando asemejarlas a la fonología anglosajona, acabaría investigando cómo cambian los procesos metacognitivos y de metamemoria implicados en el aprendizaje cuando éste se produce en un segundo idioma.

Desde una perspectiva global, vivimos en una sociedad multilingüe donde la convivencia de idiomas es más la norma que la excepción. Concretamente en España, tenemos cuatro idiomas cooficiales además del castellano: catalán, valenciano, euskera y gallego. ¡Una riqueza cultural y lingüística inmensurable! Más allá de los idiomas cooficiales dentro del territorio español, la implantación de un modelo bilingüe ha hecho que la alternancia de idiomas dentro del entorno educativo sea una nueva realidad. Nadie cuestionará las ventajas que adquirir un segundo idioma tiene a medio y largo plazo. No obstante, puede que muchas personas se identifiquen con mi amiga Lidia y los diferentes retos a los que tuvo que hacer frente durante su etapa académica. Lidia estudió inglés en el colegio. Tenía un nivel medio (B2) que le permitía ver series y películas en versión original y se atrevía a explicar a cualquier persona extranjera que le preguntara por la Mezquita-Catedral en Córdoba. Sin embargo, Lidia entraba en pánico cada vez que se enfrentaba al examen de *Science* en inglés. Estudiar el contenido en una segunda lengua –inglés–, en la que no tenía el mismo nivel de dominio que su lengua materna, le hacía sentirse insegura y creer que no aprobaría el examen.

La historia de Lidia me acompañó durante muchos años de mi adolescencia. Más tarde, ya en la Universidad, el patrón se repetía con bastante frecuencia entre otras personas a mi alrededor. Muchas/os compañeras/os temían el momento de leer artículos científicos y trabajar sobre su contenido, ¡PORQUE ESTABAN ESCRITOS EN INGLÉS! Esto me hizo pensar que quizás Lidia no era la única que experimentaba algo así, sino que esa historia simplemente representaba la sensación general de otras muchas personas cuando se enfrentan al reto de estudiar contenidos nuevos en un segundo idioma, en el que quizás no tienen un alto dominio.

Apuesto a que muchas personas se sienten reflejadas en *Lidia* o por lo menos, conocen a alguna *Lidia* a su alrededor; prevalencia que manifiesta la importancia de saber qué ocurre a nivel cognitivo cuando mi amiga *Lidia,* o cualquiera de ustedes, estudia para su examen de *Science*. En mi tesis, investigo si las estrategias de aprendizaje que se ponen en marcha durante el estudio cambian cuando el aprendizaje tiene lugar en el contexto de un segundo idioma. Estas estrategias de aprendizaje son procesos metacognitivos y de metamemoria que nos permiten evaluar la dificultad del material, detectar si una parte no la hemos comprendido bien o decidir qué parte re-estudiar, por ejemplo. Estos procesos de *monitorización* y *control* están estrechamente ligados al éxito académico y para que tengan lugar de forma correcta, numerosos procesos cognitivos se ven implicados. Digamos que ya de por sí, poner en marcha estos procesos metacognitivos es costoso en términos cognitivos, ¡imagina hacerlo en el marco de un segundo idioma!

Investigaciones previas han constatado que cuando trabajamos en un segundo idioma, diferentes procesos de control cognitivo están actuando para controlar la interferencia del idioma dominante, integrar la información, o mediar la traducción de conceptos, por ejemplo. Ante este escenario de sobrecarga cognitiva, parece razonable pensar que los procesos metacognitivos de aprendizaje se vean comprometidos cuando estamos trabajando en un segundo idioma, que también es altamente demandante. No obstante, los resultados más relevantes de esta tesis ponen de manifiesto que personas con un nivel medio/alto de inglés como segundo idioma, son capaces de monitorizar –estrategia metacognitiva– el material de estudio de forma correcta tanto en su primer como en su segundo idioma. Este hallazgo tiene especial relevancia y aplicabilidad para todas las personas que en se enfrentan a materiales en inglés en el ámbito académico y/o profesional. ¡Estudiar en un segundo idioma no impide la correcta monitorización del material de estudio!

A continuación, encontrarás un intento de sintetizar de forma elegante cuatro años de intenso trabajo –formación, lectura, cursos, diseño experimental, selección de estímulos, programación, análisis– con la humilde intención de poder contribuir con un pequeño granito de arena en la generación de conocimiento. Ojalá esto solo sea el inicio de un largo y apasionante camino. Ojalá algún día, los resultados más relevantes de la investigación que llevo a cabo tengan aplicación en políticas educativas reales.

Preface

The truth is that my life has been surrounded by languages since early childhood. Being quite talkative from a young age, I remember playing in my room as "a teacher" while "speaking other languages." Paradoxically, despite growing up in a monolingual environment, my parents were motivated to enroll me in English classes when I just turned six because, of course, "learning a second language is very important." I imagine that, seeing how chatty I was, they believed it would be in my interest. Or perhaps, that it would secure a better future for me. Honestly, I doubt that at that time, they had knowledge of the direct and indirect, social, cognitive, and personal consequences stemming from the learning and use of a second language. What was clear, though, was that "you have to learn English, Marta." And Marta, capable of finding enjoyment in everything, continues not only to speak English twenty years later but also enthusiastically undertakes a thesis on the subject.

The poetic backstory of how and why I have reached this point remains unknown to me (coincidence, destiny?). What is clear is that no one could have imagined that the little girl, who used to utter invented words in an attempt to mimic Anglo-Saxon phonology, would eventually delve into researching how metacognitive and metamemory processes involved in learning change when occurring in a second language.

From a global perspective, we live in a multilingual society where the coexistence of languages is more the norm than the exception. Specifically in Spain, we have four co-official languages in addition to Spanish: Catalan, Valencian, Basque, and Galician. An immeasurable cultural and linguistic richness! Beyond the co-official languages within Spanish territory, the implementation of a bilingual model has made language alternation within the educational environment a new reality. No one will question the long-term advantages of acquiring a second language. However, many people might identify with my friend Lidia and the different challenges she had to face during her academic journey. Lidia studied English at school. She attained a moderate level (B2), which enabled her to watch TV series and movies in their original versions. She was also confident enough to explain to any foreign person who asked about the Mosque-Cathedral in Córdoba. However, Lidia would panic every time she faced a *Science* exam in English. Studying the content in a second language, in which she did not have the same level of proficiency as her native language, made her feel insecure and believe that she would not pass the exam.

Lidia's story accompanied me for many years during my adolescence. Later, at university, the pattern repeated itself quite frequently among other people around me. Many of my classmates feared the moment of reading scientific articles and working on their content, BECAUSE THEY WERE WRITTEN IN ENGLISH! This made me think that perhaps Lidia was not the only one experiencing something like this, but that her story simply represented the general feeling of many others when they confront the challenge of studying new content in a second language, in which they may not have a high level of proficiency.

I believe that many people can relate to my friend Lidia, or at the very least, know someone similar to her. This prevalence emphasizes the importance of comprehending what occurs at the cognitive level when my friend Lidia, or any of you, are studying for your *Science* exam. In my thesis, I investigate whether the learning strategies that come into play during studying change when learning takes place in the context of a second language. These learning strategies are metacognitive and metamemory processes that allow us to assess the difficulty of the material, detect if we have not understood a part well, or decide which part to review, for example. These monitoring and control processes are closely linked to academic success, and for them to occur correctly, numerous cognitive processes are involved. Initiating these metacognitive processes is cognitively demanding in itself; now, imagine doing so within the context of a second language!

Previous research has established that when we work in a second language, different cognitive control processes are at play to manage language interference, integrate information, or mediate the translation of concepts, for example. Given this scenario of cognitive overload, it seems reasonable to think that metacognitive learning processes may be compromised when we are working in the context of a second language, which is also highly demanding. However, the most relevant results of this thesis demonstrate that individuals with a moderate to high level of English as a second language are capable of monitoring - a metacognitive strategy - the study material correctly in both their first and second language. This finding is of particular relevance and applicability to all individuals who, in an academic and/or professional context, encounter materials in English. Studying in a second language does not hinder the proper monitoring of study material!

Below, you will find an attempt to elegantly synthesize four years of intensive work: reading, courses, experimental design, stimulus selection, programming, and analysis. All this with the humble intention of contributing a small grain of sand to the generation of knowledge. Hopefully, this marks just the beginning of a long and thrilling journey. Perhaps one day, the significant research findings I have pursued will have tangible applications in educational policies.

Part I

Introduction

Bilingualism and second language use

Over the past decades, the use of a second language (L2) has become an integral part of daily life for a significant portion of the global population (Grosjean, 2010). This linguistic phenomenon manifests itself in many different contexts where bilingual individuals, with diverse linguistic backgrounds and proficiencies, process information, acquire new knowledge, and make decisions through their second language.

Bilingualism could be conceived as a spectrum that spans a wide range of linguistic and cultural experiences. It is a term with a broad definition – dynamic and multifaceted, it varies depending on the unique characteristics and attributes of individuals (Luk & Bialystok, 2013). When discussing bilingualism, we need to consider a range of variables that help paint a more accurate picture of someone's language capabilities and experiences. These variables include, but are not limited to, factors such as the age at which a person acquired their second language, the percentage of time they use each language in their daily life, and objective measures of their language proficiency. Recognizing that there is no single definition allows us to appreciate the diversity of bilingual individuals and scenarios. From this perspective, bilingualism includes people who use two or more languages in their daily lives, even if they had not mastered them at a high proficiency level (Cenoz, 2003). In nowadays society, we can observe a plethora of scenarios in which the use of L2 is indispensable.

In the realm of business, multinational corporations frequently adopt English as the lingua franca for internal communications and international meetings. Similarly, in the domain of diplomacy and international relations, the use of multiple languages is common, reflecting the diversity of nations and cultures involved in global affairs. Furthermore, the digital age has given rise to a globalized online community where individuals from diverse linguistic backgrounds communicate through social media platforms, blogs, and forums. In the realm of entertainment and media, the global reach of films, television shows, and music often leads to the consumption of content in an L2.

Beyond this, and most importantly, English often serves as the predominant language of instruction and communication in academia and higher education, even for non-native speakers (Byun et al., 2011; Dafouz & Camacho-Miñano, 2016; Macaro, 2018; Pessoa et al., 2014). In fact, bilingualism has surged to the forefront of educational policies and practices around the world. This shift reflects the increasingly globalized nature of academia and higher education, where universities and institutions strive to connect with a diverse and international student body. English, in particular, has an established status as the lingua franca for academic communication, transcending national and linguistic borders (Dafouz & Camacho-Miñano, 2016). This trend is driven by various factors, including the desire to facilitate cross-cultural collaboration, attract international students, and enhance the global competitiveness of educational institutions.

The impact of L2 dominance in academia extends beyond the classroom. It also influences research collaboration, academic publishing, and the dissemination of knowledge on a global scale. Researchers and scholars, regardless of their native language, often publish their work in English to reach a wider international audience. While the promotion of bilingualism in education undoubtedly brings many advantages, it also poses unique challenges, especially for students who do not possess high levels of L2 proficiency. The adoption of an L2 as the medium of instruction in environments where it is not the students' first language (L1) can lead to several complexities. The rise of bilingualism in general, and in academia and higher education in particular, has reshaped the educational landscape on a global scale.

Despite the ubiquity of L2 use across diverse contexts, research exploring whether learning processes change significantly when individuals operate in an L2 context, as opposed to their L1, remains scarce. Numerous studies have examined the impact of performing specific tasks in either an L1 or L2 on the ability to carry out a diverse array of activities: decision-making (Brouwer, 2021), visual attention (Chabal & Marian, 2015), the perception of multisensory emotions (Chen et al., 2022), long-term memory (Arndt & Beato, 2017; Marian et al., 2021), reading comprehension and inferences (Pérez et al., 2018), and even prospective memory (López-Rojas et al., 2022). Yet little is still known about whether metamemory processes involved in learning undergo significant changes when learning is carried out in L2. This knowledge gap becomes especially critical in educational settings, where understanding these mechanisms is crucial for addressing non-native primary students' potential underachievement (Buehler et al., 2021). As our world continues to become more interconnected, the role of L2 in various aspects of life is likely to expand, making it increasingly important to explore the multifaceted impact of language on cognition, and self-regulated learning.

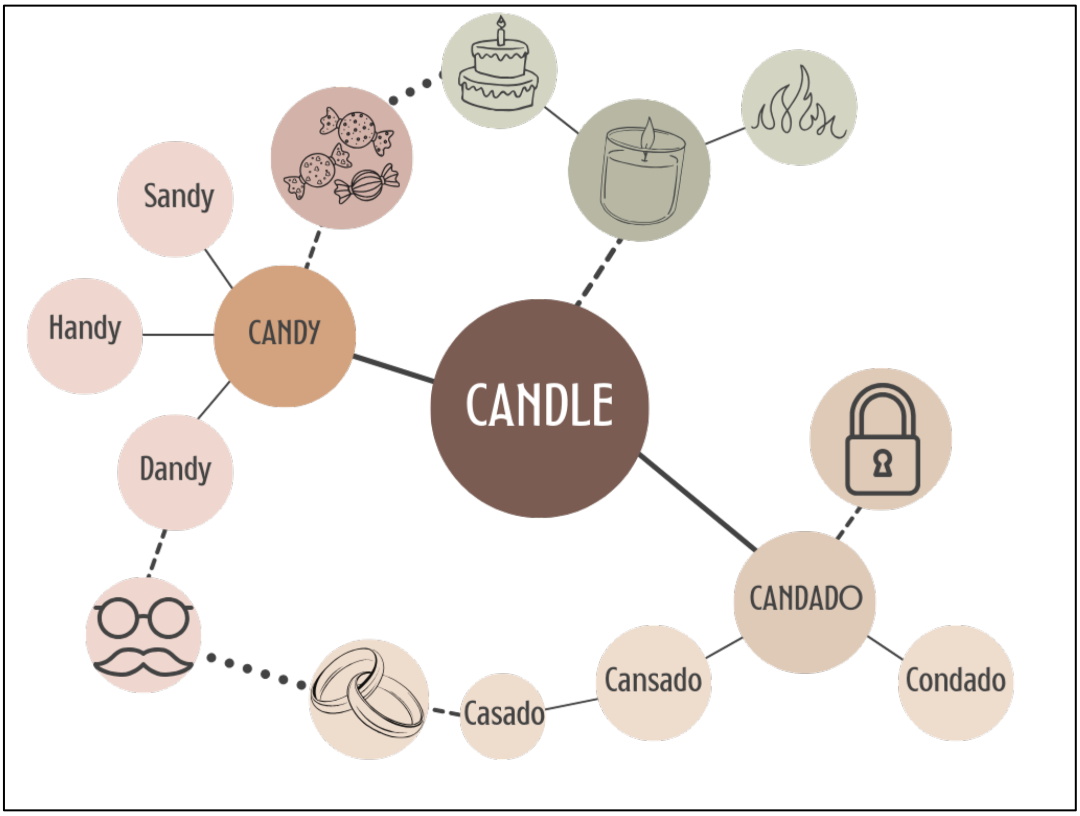
Cognitive trade-offs of bilingualism

Previous research have revealed that bilingualism can have an impact on our overall well-being by molding the outcomes of cognitive growth and the aging process (Bialystok, 2001; Bialystok et al., 2007, 2012, 2016). Bilingualism has also been reported to enhance metalinguistic skills (Campbell & Sais, 1995; Duncan, 2005; Galambos & Goldin-Meadow, 1990; Galambos & Hakuta, 1988; Gathercole et al., 2014; Titone, 1997; Woll, 2018), to improve performance on executive-function tasks (see Bialystok & Craik, 2022 for a review), to boost creativity and divergent thinking (Adi-Japha et al., 2010; Bialystok & Shapero, 2005; Ning et al., 2020) or to promote more logical decision-making (Hayakawa et al., 2016). Nevertheless, the condition of being bilingual also entails specific cognitive expenses due to the requirement of simultaneously managing two languages (Giezen et al., 2015; Marian & Spivey, 2003b).

Even for highly proficient bilinguals, working in an L2 can be cognitively challenging (Ma et al., 2014; S. Moreno et al., 2010; Pérez et al., 2018). A large number of studies have established that within the bilingual brain, both languages are co-activated during the production or comprehension of auditory, written, and bi-modal communication, even in contexts where only one of the languages is involved (Bialystok, 2017; Bialystok et al., 2012; Chen et al., 2017; Giezen et al., 2015; Hatzidaki et al., 2011; Iniesta et al., 2021; Kroll et al., 2014; Shook & Marian, 2012). Basically, when a bilingual is speaking or processing information in one language, the other language is simultaneously active. This co-activation can occur at various linguistic levels, including phonological (Marian et al., 2008; Shook & Marian, 2019), lexical (Shook & Marian, 2013), and grammatical (Lemhöfer et al., 2008; Morales et al., 2014). Imagine a bilingual individual conversing in English-L2. When the word *candle* appears in scene, their brain automatically will co-activate words with similar sounds in English (e.g., *candy*) but also words that share phonology in their other language (e.g., Spanish-L1, *candado–lock*). This situation can result in momentary lexical or syntactical interferences. Thus, not only would this bilingual person activate other similar-sounding words within the language that is being spoken, but translations of those words in other languages would

Figure 1.

Associative diagram of the word CANDLE.



also become activated (Shook & Marian, 2019). Similarly, reading a word in one language, triggers the written and spoken representations of both languages. Figure 1 represents an example of the spread co-activation across languages when a Spanish-English bilingual read or listen to the word *candle.*

Studies with homographs and cognates in reading (Dijkstra & Kroll, 2005; Libben & Titone, 2009; Schwartz & Kroll, 2006), listening (Marian & Spivey, 2003a), and even writing production (Iniesta et al., 2021) provide evidence for this co-activation effect. Homographs are words that share the same spelling (i.e., they look alike) but may have different meanings. For example, the word *library* means *biblioteca* in Spanish, not to be confused with *librería*, which means *bookstore*. Homographs create situations where bilinguals have the challenge of disambiguating and selecting the correct language-specific interpretation. Otherwise, cognates are words that share a common origin, have similar meanings and similar spellings across languages. For instance, in English and Spanish, *family* and *familia* are cognates. When bilingual individuals encounter cognates, both languages can aid each other in the lexical access of a term. In a typical experimental task, bilingual people show facilitation and interference –by means of response time– in cognates and homographs processing respectively. These two effects provide insights into the co-activation phenomenon (Kroll et al., 2015).

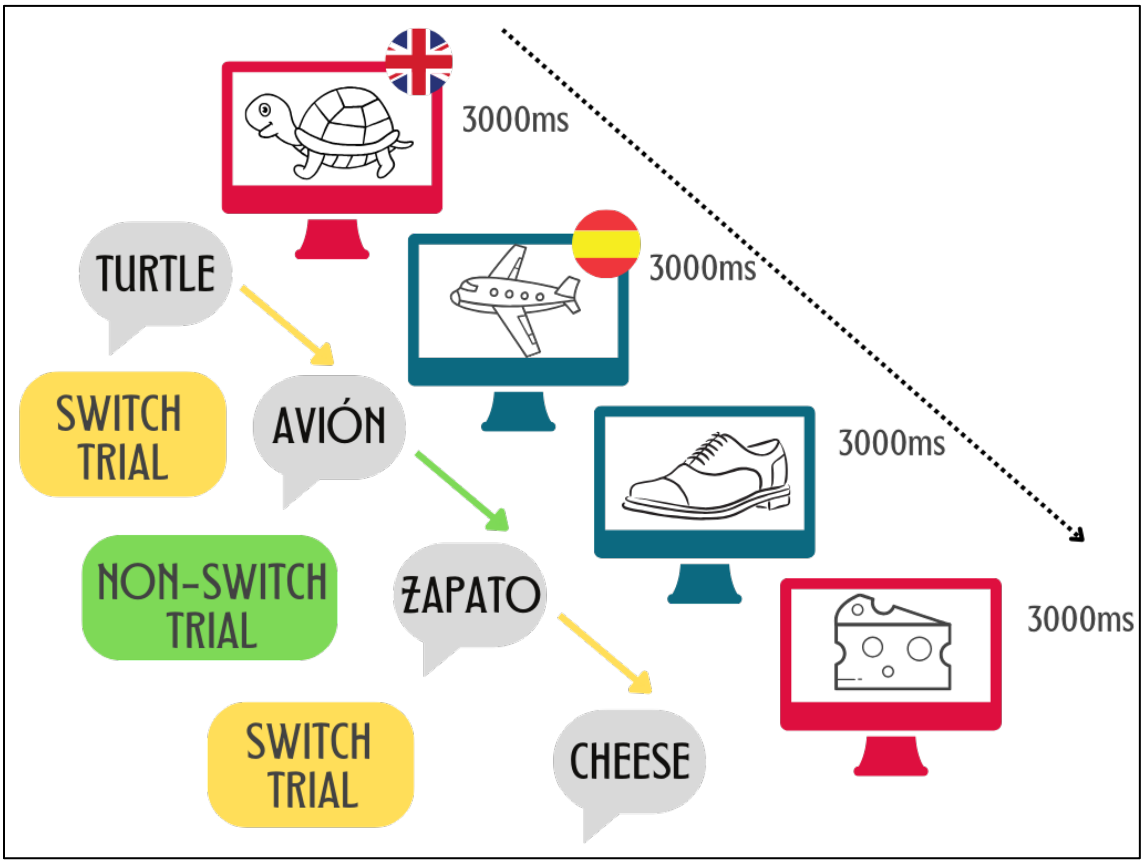
Further, research using *eye-tracking* have replicated the findings of parallel activation during language comprehension (Marian & Spivey, 2003b; Shook & Marian, 2019). For example, in a visual search experiment involving Russian-English bilinguals, participants were instructed to click on the item they heard (e.g., *speaker*) from a computer display containing various objects (e.g., *umbrella, matches,* *hammer*). Results showed that when asked in English to select the *speaker*, Russian-English bilinguals tended to look more frequently at the *matches* (*speachkey* in Russian, which shares phonological similarity with the target word *speaker*) compared to other distracting objects in the display, and more often than their English monolinguals counterparts did. Hence, all this evidence makes it clear that, even when not in explicit use, the second language remains consistently active and undergoes automatic processing within the bilingual brain.

If the alternative language remains active when using the context-appropriate language, additional cognitive resources are recruited to control the interference and actively select the desired language (Beatty-Martínez et al., 2020; Kroll et al., 2015; Macizo et al., 2010; Soares et al., 2019). These cognitive mechanisms recruited to manage the co-activation of languages and to select the appropriate one for a given context involves suppressing interference from the non-target language and ensuring that the intended language is used accurately and fluently (Declerck & Koch, 2023). On a basic level, language control allows bilinguals to maintain language boundaries, preventing lexical or grammatical blending. Then, when speaking in English-L2, a bilingual person must inhibit the urge to insert Spanish-L1 words or sentence structures unintentionally. In this respect, brain imaging studies have revealed that neural bases of bilingual language control share brain networks with processes that enable domain-general cognitive control (e.g., Calabria et al., 2018).

The effect of language inhibitory control has been classically explored with the language switching paradigm (Blanco-Elorrieta & Pylkkänen, 2017; Christoffels et al., 2007; Declerck et al., 2017; Reynolds et al., 2016; Verhoef et al., 2009; Yahya & Özkan Ceylan, 2022). Figure 2 illustrate a sequence of four trials of a naming switching task.

Figure 2.

Example of a trial sequence of the language switching paradigm.



Note: Each picture appears with a color frame as a cue that indicates the language in which the picture should be named.

In this paradigm, participants usually name pictures in either their L1 or L2 based on cues such as color, which signal the required language for that trial (e.g., red for English and blue for Spanish). Depending on the sequence of trials, there may be either a "switch trial", where participants need to change the language, or a "non-switch trial", where they continue using the same language (Costa & Santesteban, 2004; Meuter & Allport, 1999). The standard outcomes of such experiments reveal two patterns. First, responses in switch trials tend to be slower and more error-prone compared to non-switch trials, a phenomenon referred to as "language switching cost." This cost arises as a consequence of the increase in cognitive demand associated with managing two languages simultaneously (for a comprehensive review, see Declerck & Philipp, 2015). Second, there is often a more pronounced switching cost when switching from L1 to L2 than the opposite direction, known as "asymmetrical switching cost" (De Bruin et al., 2014; Macizo et al., 2012). This suggests the involvement of inhibitory processes aimed at preventing interference and competition between the two languages. The greater competition from L1 compared to L2 engages stronger inhibition of L1 when processing in L2. As a result, switching from L2 to L1 is more costly than the opposite. Furthermore, language control is highly context-sensitive and operates as a dynamic interplay between various cognitive processes. Bilinguals possess a remarkable ability to assess their immediate surroundings, the individuals they are interacting with, and the linguistic norms prevalent in a given situation (Beatty-Martínez & Dussias, 2017).

The adaptive control hypothesis, as proposed by Green & Abutalebi (2013), posits that bilingual individuals must continually engage control and meta-control processes to maintain their language goals, effectively monitor any arising conflicts, and suppress potential interference stemming from the co-activation of both languages. This intricate process of language control is highly contingent on the specific context and the unique language experiences of the bilingual individual (Beatty-Martínez et al., 2020). The fact that bilinguals can easily switch between languages while communicating effectively and avoiding confusion shows how flexible and skilled they need to be in handling different languages and situations.

Language control might be especially costly for unbalanced bilinguals (i.e., people with an asymmetric proficiency level in the languages they know), since the interference from the dominant L1 to the less-dominant L2 has been shown to be greater than the other way around (e.g., Contemori & Dussias, 2016; Meuter & Allport, 1999; Soares et al., 2019). Moreover, unbalanced and late bilinguals rely primarily on transfer from L1 to L2. Semantic representations are weaker in L2, and some concepts are activated through L1-L2 translation (Kroll & Stewart, 1994). Furthermore, some studies have shown that L2 processing takes more time, word recognition is slower, and a smaller amount of information is processed simultaneously compared to L1 (Dirix et al., 2020). In sum, bilingualism brings about cognitive challenges affecting language control. Below, we will explore how language control interplays with reading comprehension.

The dynamics of reading comprehension in L1 and L2

When we read, multitude cognitive and metacognitive processes are engaged to enable comprehension. At its core, reading involves visual recognition of written words, a bottom-up process that relies on a well-established mental lexicon, where words are stored with their associated meanings and pronunciations. Furthermore, comprehension goes beyond mere word recognition and it encompasses higher-level processes such as information updating, inferencing, integration with prior knowledge, and the ability to monitor and evaluate the difficulty of the text (Castles et al., 2018). Thus, text processing requires the construction of a mental representation (i.e., situation model, Van Dijk & Kintsch, 1983) based on the interplay between both: bottom-up –automatic– processes such as visual word recognition or connections between words-spelling-meaning, and top-down –reader-initiated– processes such as inferring meanings from the surrounding text (van den Broek & Helder, 2017).

These processes operate effortless and smoothly for proficient readers in their L1, facilitated by their deep understanding of the language's syntax, semantics, and pragmatics. In L1, these mechanisms have been developed and fine-tuned over years of exposure and practice, resulting in rapid and precise word recognition, information updating and inferencing. However, when individuals read in an L2, particularly when their L2 proficiency is not at a native-like level, language processing undergoes significant changes.

Firstly, processing information in an L2 imposes a higher cognitive load, leaving less working memory capacity for higher-order tasks like constructing mental models or making inferences (Francis & Gutiérrez, 2012a; Sandoval et al., 2010). Even basic linguistic processes, such as accessing words in the lexicon, lead to increased activation in brain regions associated with cognitive control when performed in L2 (Ma et al., 2014). Secondly, representations of L2 words are less robust and detailed than those of the L1 (see the weaker links hypothesis by Gollan et al., 2008). Lastly, although the linguistic networks of both languages are interconnected, the early-stage representations of L2 (orthographic/phonological) have a lower baseline activation level, thus requiring more time and cognitive resources to activate (see the bilingual interactive activation plus model, BIA+; Dijkstra & van Heuven, 2002).

Empirical evidence supports these assumptions and previous research exploring differences between languages in reading comprehension tasks shows indeed poorer comprehension in L2 compared to L1 (for a review, see Melby-Lervåg & Lervåg, 2014). For example, L2 words and sentences are typically processed at a slower pace compared to their L1 equivalents (Cop et al., 2015; Whitford & Titone, 2012), and students tend to have lower vocabulary retention for L2 words in comparison to their L1 counterparts (Gablasova, 2014). The word recognition process in L2 may require more effort and conscious attention, which can impact not only reading fluency, but also comprehension accuracy (Plat et al., 2018; Schwartz & Kroll, 2006). L2 readers may encounter words with unfamiliar spellings, pronunciations, or meanings, which necessitates greater cognitive resources for decoding and comprehension (Hosoda, 2017; Jeon & Yamashita, 2014). Also, since it takes longer to activate a word's meaning in L2, a poorer development of the mental model of a text might be expected.

The higher-order processes involved in comprehension, such as making inferences or integrating information with prior knowledge, become more challenging in an L2 context. It has been suggested that comprehension in L2 primarily relies on late integration processes. This involves passively integrating encountered words rather than actively predicting them, resulting in a relatively less proactive type of processing than when processing in L1. Accordingly, more resources are required to anticipate upcoming linguistic information during successful L2 reading comprehension (Foucart et al., 2016; Kaan et al., 2016; Pérez et al., 2018).

For example, Pérez et al. (2018) presented findings indicating that updating and inference processes in text comprehension were less efficient in L2 compared to L1. Furthermore, Dirix et al. (2020) in an *eye-tracking* experiment found that total reading times were about 20% longer and that 15% more fixations were made when studying in L2, especially when processing longer and syntactically complex sentences. Finally, recent studies have demonstrated a difficulty in L2 when evaluating the coherence between sentences that are far apart within the paragraph or text, while the coherence between sentences positioned next to each other remained unaffected when working in an L2 (Ushiro et al., 2022). This finding suggests that an L2 comprehension cost may arise when individuals need to monitor the coherence and other features of the text.

In addition, the level of proficiency in L2 may also have an impact on L2 text learning. The ability of L2 readers to construct the necessary inferences for forming situation models might be constrained by their proficiency in L2 reading. When L2 reading skills are limited, readers are compelled to prioritize fundamental reading processes (i.e., word decoding and syntactic parsing) over inferential processing, allocating their cognitive resources accordingly (Horiba, 1996; Hosoda, 2014). Given the challenges in inferential processing, less proficient L2 readers may find arduous to construct comprehensive situation models. This, in turn, hinders their ability to internalize the causal relationships presented in the texts as knowledge. While high-proficiency readers are more likely to effectively employ their prior knowledge to enhance their comprehension of the text, low-proficiency readers often struggle to construct accurate situation models, impeding their capacity to acquire the causal relations presented in the text (Hosoda, 2017).

To summarize, all the evidence reported so far suggests that L2 processing is more challenging and might take place within a presumably overloaded cognitive system (Adesope et al., 2010; Hessel & Schroeder, 2020, 2022; Pérez et al., 2018). The question is then whether this cognitive overload has consequences for learning strategies and cognitive resources allocated when processing, studying, or acquiring new information in L2. In the upcoming section, we provide an overview of the primary studies on self-regulated learning, the main theoretical frameworks within this domain, and the cognitive mechanisms underlying the learning process.

Metacognition in Self-regulated Learning

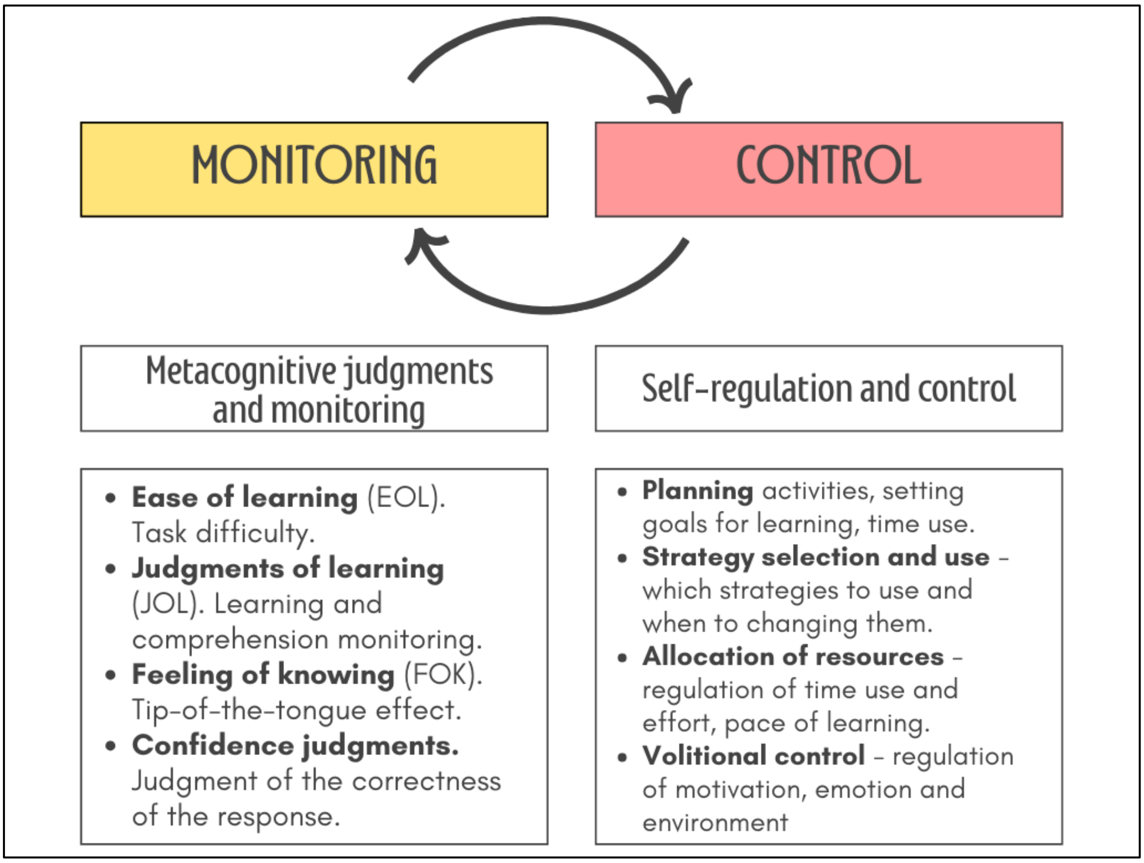
From a learning perspective, metacognition is a key function conceived as a feeling-of-knowing state that serves a self-regulatory purpose whereby the brain monitors the learning conditions and regulates the resources and processes devoted to learning. That is, through metacognitive strategies one can observe the ongoing processing, asses one's comprehension and/or learning, detect potential errors and decide what strategies need to be employed in order to improve learning outcomes. Therefore, metacognition is crucial for self-regulated learning (Panadero, 2017; Zimmerman, 2000, 2008) and it is involved in the development of successful learning strategies linked to academic achievement (Pintrich & Zusho, 2007; Zimmerman, 2008; Zusho, 2017).

From a broad perspective, self-regulated learning (SRL) represents a multifaceted construct at the heart of educational psychology, weaving intricate interactions with cognitive, motivational, and contextual dimensions (Zimmerman, 2000). SRL entails personal, behavioral, and environmental dimensions, and underscores the learner's intrinsic efforts in actively managing their own academic achievement across three distinct phases: planning, performance, and regulation. The planning phase encompasses goal orientations and self-efficacy beliefs, setting the stage for learners' educational pursuits. During the performance phase, learners employ monitoring judgments, strategies, and cognitive/metacognitive processes to guide their learning. Simultaneously, during the self-regulating and control phase, mechanisms such as volitational control –motivational beliefs–, planing activities, selection and use of strategies, and allocation of resources are taking place (Pintrich, 2000). **¡Error! No se encuentra el origen de la referencia.** 3 represents an overview of the processes and activities associated with metacognitive judgments and monitoring, and self-regulation and control, respectively. In essence, SRL embodies a proactive journey where learners establish goals for their learning endeavors and subsequently engage in a dynamic interplay of monitoring, regulating, and controlling their cognitive, motivational, and behavioral domains, all guided by their objectives and contextual factors in the learning environment.

There is compelling evidence supporting the efficacy of SRL in fostering proactive learning and enhancing academic achievement. In fact, numerous studies have demonstrated the substantial positive impact of integrating self-regulatory processes into domain-specific instruction within formal educational settings, cultivating active learners who consistently yield superior academic outcomes (e.g., Harris & Graham, 2009; Pintrich, 2004; Schunk & Greene, 2018).

Figure 3.

Components of Metacognition and Self-regulated learning.

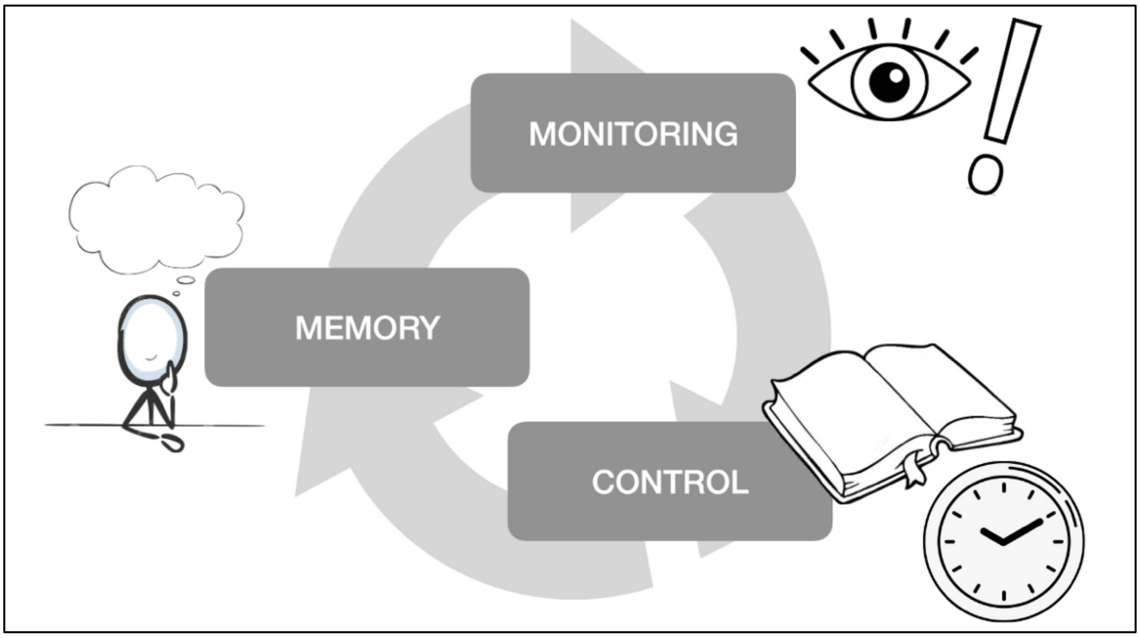


The term SRL was first introduced by educational and developmental psychologists and has been extensively investigated from the educational framework (see Schunk & Zimmerman, 1994; Zimmerman, 1986; Zimmerman & Schunk, 1989, 2011). Nevertheless, metacognition is an "older" term defined and used by cognitive psychologists (see Flavell, 1979), and so, has been subjected of much research. According to the classical model proposed by Nelson and Narens (1990), there are two mechanisms underlying metacognitive strategies in the learning process: *monitoring* and *control*. Metacognitive *monitoring* refers to the online supervision and assessment of the effectiveness of cognitive resources while metacognitive *control* refers to the management and regulation of such cognitive resources.

These two functions of metacognition –*monitoring* and *control*– are inextricably connected and, in turn, have consequences on memory. *Monitoring* processes, such as judgments regarding the ease/difficulty of the task or the level of learning achieved after studying, are crucial for guiding *control* processes, such as selecting a strategy, regulating the cognitive resources devoted to the task, and adjusting them when necessary (Nelson & Narens, 1990; Zechmeister & Shaughnessy, 1980a). For example, task's perceived difficulty, uncertainty, complexity, or novelty, may serve as cues to trigger *control* strategies. Thus, accurate *monitoring* is believed to be used to facilitate *control* for the benefit of learning and memory performance. Figure 4. shows how monitoring and control are interrelated in the learning process, and their interplay is assumed to have direct consequences on memory performance (Pieger et al., 2016).

Figure 4.

The interplay between monitoring, control and memory.



Nelson & Narens, (1990) proposed the strategic use of monitoring in self-regulation: people *monitor* –asses– their learning through various metacognitive judgments, and this monitoring affects their *control* –regulation– behavior (e.g., terminating study, selecting study strategies). Thus, metacognitive judgments play a critical role in the regulation of study and behavior. Learners tend to monitor the degree of learning and use their judgments as a basis for regulating the study of the materials. Consequently, if people prematurely terminate their study before they have mastered difficult materials, it is most likely due to inaccurate monitoring.

Nevertheless, the relationship between metacognitive *monitoring* and metacognitive *control* might be also bidirectional. Koriat et al., (2006) proposes that metacognitive judgments may also be based on feedback from the outcome of control operations. Monitoring may not occur prior to the controlled action, but rather, it might take place afterwards. According to this hypothesis, the difficulty of an item is monitored *ad hoc*: learners allocate the appropriate resources to an item based on its demands. They understand that a specific item will be challenging to remember when it demands a relatively higher level of effort to commit to memory.

These two models are not mutually exclusive and might be dynamically operating in a sequential mode: the monitoring-based control may lead to control-based monitoring. The initial assessment of a situation provides valuable information for executing control actions, and the feedback obtained from these actions can subsequently be used as a foundation for monitoring. This monitoring process, in turn, can guide future control operations, creating a cyclical relationship between monitoring and control. In other words, subjective experience informs the initiation and self-regulation of control operation that may in turn change subjective experience. For example, when people feel that they do not understand a text they have just read, or have the feeling they have not mastered it, they read it again until they feel more confident.

More recent theories on metacognitive regulation and self-regulated learning, in particular, suggest that individuals rely on continuous monitoring to determine the best course of action to achieve their learning goals (e.g., Dunlosky & Ariel, 2011; Dunlosky & Tauber, 2015; Koriat & Goldsmith, 1996; Metcalfe, 2009; Metcalfe & Finn, 2008; for a review see Panadero, 2017). For instance, identifying the difficult parts of a text correctly leads to appropriate effort regulation and strategy selection, which, consequently, results in greater comprehension and better memory (Follmer & Sperling, 2018). Thus, from a learning perspective, metacognitive strategies and self-regulation have been linked to academic achievement (Pintrich & Zusho, 2007; Zimmerman, 2008; Zusho, 2017), as they are critical for comprehending and memorizing information (e.g., Collins et al., 1996; Fukaya, 2013; Huff & Nietfeld, 2009; Krebs & Roebers, 2012; Thiede et al., 2003).

In this line, Deekens et al., (2018) investigated the relationship between the frequency of metacognitive monitoring and the utilization of surface and deep-level strategies. Surface-level strategies usually imply investing minimal time and effort to meet the requirements (e.g., rote learning or memorizing key concepts, Cano, 2007), whereas deep-level strategies involve, for example, paying attention to the meaning of the materials, relating ideas and integrating them with previous knowledge, to maximize understanding. Deep-level strategies are regarded as more effective strategies producing longer-lasting learning (Deekens et al., 2018;

Lonka et al., 2004; Vermunt & Vermetten, 2004). In their study, Deekens et al. (2018) found that students who enacted more frequent learning monitoring also engaged deep strategies more frequently than low-monitoring students, and this resulted in better performance on academic evaluations. This pattern suggests that the combination of metacognitive monitoring and deep-level learning strategies are intrinsically linked to successful academic achievement.

The interplay between *monitoring* and *control* is assumed to be effortful and cognitively demanding (Efklides, 2014). For these two processes to occur, sufficient cognitive resources must be available and executive control must be engaged, as the flow of bottom-up (*monitoring*) and top-down (*control*) processes is simultaneous (Nelson & Narens, 1990). Evidence from different kind of studies supports this assumption. For example, research comparing younger and older people (whose executive control skills might be in decline) suggests that metacognitive processes recruit cognitive resources. Stine-Morrow et al. (2006) showed that, when memory monitoring was required, learning was poorer in a group of older participants relative to a group of younger participants. This suggests that metacognitive monitoring might compromise performance in the age group. Similarly, Tauber and Witherby (2019) showed that, unlike younger adults, instructions to use metacognitive strategies did not improve memory performance in older adults. This suggests that age-related deficits might make it difficult for older participants to implement metacognitive strategies. In the same vein, neuropsychological studies provide evidence that the neural correlates of metamemory are driven by frontal lobes (Pannu & Kaszniak, 2005). Likewise, a review of brain imaging studies reveals that midfrontal and frontoparietal areas are involved in metacognition (Do Lam et al., 2012; Vaccaro & Fleming, 2018), which suggests that executive functioning is involved (Fernandez-Duque et al., 2000).

Overall, there is mounting evidence that the synergic interaction between monitoring and control processes plays a critical role in the self-regulated learning, and that their interplay is cognitive effortful, which may have consequences under certain circumstances of high cognitive load. Below, we will discuss the different experimental approaches to assessing the monitoring process and self-regulated learning.

Assessing the Monitoring Process and SRL

Traditionally, from an experimental approach, the monitoring process have been investigated through the assesment of four main metacognitive factors : (a) task difficulty (ease-of-learning judgments – EOL), (b) learning and comprehension monitoring (judgments of learning – JOL) , (c) feeling of knowing (FOK), and (d) confidence judgments.

We, as learners, may judge how hard or easy a task will be, a process referred to as ease-of-learning judgments (EOL), as described by Nelson and Narens (1990). EOL judgments happen before you start the task, in the learning phase and are based on previous knowledge and past experiences. Imagine, for example, you have been assigned the task of reading a new article on a certain topic for later discussion. You might instinctively use EOL judgments when you see the title of the article that you are about to read. Then, when reading such article, you might realize that you do not understand something you just read or that you are reading too fast for what you want to learn. These are judgments of learning (JOLs) and help comprehension monitoring. You might also make these judgments to figure out if you are ready to engange in the discussion on what you just learned. Basically, with JOLs, you predict which things you have learned and if you will be able to remember them later on.

Following up with our example, imagine the day after you read the paper, you attend the discussion with several collegues and you cannot remember a specific detail that you need to refer to, but you are sure you know it or have a strong feeling that you do. In research, this is known as the "tip-of-the-tongue" effect and it is related to feeling-of-knowing judgments (FOKs) (Koriat, 1993; Nelson & Narens, 1990). Finally, in the ebb and flow of conversation, you answer a question within the discussion and subsequently assess your confidence in the accuracy of your response. These confidence judgments reflect your level of certainty about the information you have just shared. They occur after you have attempted to remember something and provided an answer (Pressley et al., 1990).

The nature and time reference varied among these metacognitive judgments. Some of them are retrospective, as for example in confidence judgments, where individuals assess their performance after completing a task. Others, like JOLs, are prospective judgments made during the acquisition phase (encoding and storage). Some prospective judgments take place before the actual encoding of information, such as ease-of-learning judgments (EOL), while others such as feeling-of-knowing judgments (FOK) occur after unsuccessful retrieval attempts, helping individuals to predict whether they will remember the information in the future.

From an experimental approach, the monitoring process has been extensively investigated through the previously mentioned self-report judgments. For example, in a clasical task, participants are given some information to study that they have to remember later, like a list of words, and are asked to rate (usually with a percentage) the likelihood of remembering in the future the learning material they have just studied (i.e., JOL) (Koriat, 1997; Nelson & Leonesio, 1988; Nelson & Narens, 1990).

On the other hand, the assessment of learning strategies and self-regulation has given priority to the use of other methods such as think-aloud protocols, self-report questionnaires, and interviews. The think-aloud protocols require students to perform the actual task and simultaneously verbalize their thoughts. These protocols can offer "real-time" insights into metacognition and self-regulated learning in specific situations, yet implementing them on a large scale can be challenging and can compromise validity (Pintrich, 2000).

Moreover, a number of different questionnaires have been used to assess various aspects of regulation. One of the most well-established and widely-used questionnaires is the Motivated Strategies for Learning Questionnaire (MSLQ, Pintrich et al., 1993; Pintrich & De Groot, 1990). This instrument assesses students' cognitive strategy use and regulation of cognition. It delves into domain-specific details at the course level and evaluates the use of cognitive strategies with scales that measure rehearsal, elaboration, organization, and critical thinking.

Finally, the Self-Regulated Learning Interview Schedule (SRLIS, Zimmerman & Martinez-Pons, 1986, 1988) is the most standarised interview measure available. It is an individual-interview-based instrument that assesses self-regulated learning behaviors in various academic contexts (e.g., classroom discussions, writing assignments, and tests). SRLIS responses are categorized into different categories, covering knowledge, monitoring behavior, strategy use, and regulation aspects of self-regulated learning (e.g., organizing, keeping records, rehearsing, goal setting).

As in many other fields of psychology, using a single method in isolation may not yield a comprehensive understanding of the processes engaged on a certain task. Rather, the combination of multiple methods can provide more nuanced insights on the metacognitive mechanisms underlying learning. Recent research has successfully attempted to integrate various sources of information (e.g., metacognitive judgments, think-aloud-protocols, self-reported survey data, and eye-tracking) (Deekens et al., 2018; Dinsmore & Zoellner, 2018; Fox, 2009; Jordano & Touron, 2018). This integrated approach enables a more thorough disentanglement of the complexities involved in metacognition.

Even so, JOL is one of the most widely utilized methods for unraveling the intricacies of metacognitive monitoring. Consequently, the study of JOLs has garnered increasing attention within the field of self-regulated learning. Exploring the factors and modifiers that influence the magnitude of JOLs has become a central focus, shedding light on the nuanced dynamics at play in the realm of metacognition. This exploration enhances our comprehension of how learners perceive and manage their own knowledge under certain circumstance. Below, we will discuss some of the factors that may influence JOLs magnitude.

Factors and Modifiers Influencing JOLs magnitude

In experimental studies with JOLs, participants are asked to rate the likelihood (0-100%) that they will remember in the future information they have just studied. JOLs are inferential in nature and can be based on a full-list evaluation (e.g., lists of words, texts) or an item-specific assessment (e.g., single words, pictures.) (Rhodes, 2015). According to the cue utilization approach (Koriat, 1997), learners base their JOLs on different sources of information, namely intrinsic, extrinsic and mnemonic cues. Intrinsic cues refer to features of the learning material that indicate how easy or difficult it will be to learn, such as word frequency, associative strength or text cohesion. Extrinsic cues relate to the study environment, such as the use of interactive imagery, the time constrains, or repeated study trials. Mnemonic cues are internal states that provide information about how well an item has been learned, such as the subjective experience of processing an item fluently, past experiences in similar situations or beliefs. These include inherent features of the material, such as perceptual characteristics (e.g., size and clarity), association strength, word frequency, concreteness, or relatedness; conditions of encoding and testing (time frame, test format, presentation rate, retention interval, etc.); and one's own memorial experience of the material (Koriat, 1997).

Whether the effects of JOLs on memory performance rely on the ease of processing at encoding, on general beliefs, or on a combination of both factors is still under debate, but overall, the evidence suggests that people use cues at different processing levels to assess the difficulty of the learning process (e.g., Blake & Castel, 2018; Frank & Kuhlmann, 2017; Hu et al., 2015; Kornell et al., 2011; Li et al., 2021; Mueller et al., 2013; Mueller & Dunlosky, 2017; Wang & Xing, 2019).

Previous research has shown that there are different factors that can influence JOLs magnitude. For example, JOLs are sensitive to different cues and item-based features –variations of the difficulty– such as font type (e.g., Magreehan et al., 2016; Thompson et al., 2013), concreteness (e.g., Hertzog et al., 2003; Tauber & Rhodes, 2012), relatedness (e.g., Matvey et al., 2006; Soderstrom & McCabe, 2011), and emotionality (e.g., Groninger, 1976; Tauber & Dunlosky, 2012; Zimmerman & Kelley, 2010). Also, there are encoding and retrieval factors that influence JOLs: memory conditions, spacing and testing effects, timing (e.g., delaying JOLs), or the format of the memory test. Whatsmore, when multiple cues are manipulated in the same context, people can consider them all and integrate them in their JOLs. (e.g., Undorf et al., 2018).

Generally, JOLs tend to be accurate in predicting recall performance. For example, concrete and related words usually receive higher JOLs and are usually better remembered than abstract and unrelated words, respectively (Hertzog et al., 2003; Undorf & Erdfelder, 2015; Witherby & Tauber, 2017). However, some studies have demonstrated dissociations between JOLs and memory, with participants exhibiting overconfidence –JOLs are higher than actual performance– or underconfidence –performance surpasses JOLs– regarding their predictions about their success in remembering the targets (e.g., Koriat, 2015). Due to the fact that JOLs are inferential processes based on cues provided by the materials and tasks, mispredictions may arise because the cues used by the learner are not diagnostic, informative, or related to actual memory performance (e.g., see Kühl & Eitel, 2016, for mixed results in the fluency effect of JOLs and memory).

Indeed, some studies have reported that participants showed only partial accuracy for conditions that influenced encoding. For example, Dunlosky & Nelson (1994) found that JOLs were higher for items cued by categories and rhyming compared to items cued by specific letters. Surprisingly, the predictions about memory performance were similar for category and rhyming items, even though there was a clear memory advantage for category-related items. This indicates that participants often fail to effectively discriminate between encoding activities that they would result in successful learning and those that would not.

Another example comes from two of the most powerful methods of enhancing memory: spacing and testing. Spacing (see Cepeda et al., 2006 for a review) involves studying information multiple times with intervals in between (not consecutively), and it provides a memory advantage compared to massive learning, which occurs when information is studied consecutively. The testing effect (see Roediger & Butler, 2011, for a review), on the other hand, shows that recalling previously studied information enhances retention compared to simply rereading it. While both these methods are effective for memory, people's JOLs do not accurately reflect these advantages. In fact, participants often underestimate the significant benefits of spacing (Logan et al., 2012; see also Zechmeister & Shaughnessy, 1980b), and a similar pattern is observed in studies of the testing effect (King et al., 1980; Kornell & Rhodes, 2013; Roediger & Karpicke, 2006; but see also Tullis et al., 2013), wherein JOLs for tested information resemble those for restudied information, despite the clear memory benefits for the former.

In summary, the study of JOLs has gained significant importance and has come to the forefront in metacognitive research. The effects of item-based influences on JOLs have been extensively investigated at the word-unit level (e.g., Halamish, 2018; Hourihan et al., 2017; Hu et al., 2016; Li et al., 2021; Undorf et al., 2017, 2018; Undorf & Bröder, 2020; Undorf & Erdfelder, 2015; Witherby & Tauber, 2017) but also when learning larger chunks of information such as lists, paragraphs and texts (Ackerman & Goldsmith, 2011; Ariel et al., 2020; Lefèvre & Lories, 2004; Nguyen & McDaniel, 2016; Pieger et al., 2016; see Prinz et al., 2020 for a meta-analysis). Next, we will provide an overview of the experiments included in this doctoral dissertation, outlining their objectives and hypotheses.

Aims and outline of the present studies

Thus far, we have reviewed how in recent decades, the use of an L2 has become a part of daily life, and so, there has been a growing interest in exploring how bilingualism affects general cognitive abilities (see Bialystok & Craik, 2022 for a review). This focus involves an emphasis on exploring the diverse cognitive consequences resulting from various bilingual experiences (Luk et al., 2011; Sabourin & Viñerte, 2019). In fact, there are numerous contexts in which bilingual people with very different linguistic profiles process information, acquire new knowledge, or make decisions via an L2. For example, in academia and higher education, English is commonly used as the language of instruction and communication (Dafouz & Camacho-Miñano, 2016) even though English might not be students' L1 and they might not be proficient in it. Such students may face different challenges when studying and learning in L2 relative to contexts in which they study and learn in their L1. Despite the fact that bilingual instruction contexts tend to be the norm and that L2 use has grown over the years, research on whether the processes underlying learning undergo significant changes when the learning context is not L1 is still scarce (for a recent attempt at exploring the mechanisms underlying non-native primary students' underachievement, see Buehler et al., 2021).

In light of the above, the present dissertation focuses on investigating the factors that regulate metacognitive processes of learning in L2. In five experiments, we aimed to investigate the consequences of studying in L1 or L2 on the interplay between memory monitoring and control. We explored the effects of manipulating font type, concreteness, and relatedness at the word unit level, and cohesion at the text level, on JOLs and recognition to discover to what extent unbalanced bilinguals can monitor and control their learning both in L1 and L2. In addition, since proficiency in L2 could introduce subtleties and potentially exert a considerable influence on the cognitive mechanisms engaged in the learning process, we also explored the role of L2 proficiency in the underlying mechanisms of learning.

The interaction between monitoring and control processes, and learning, is significantly relevant when the materials to be learned vary in difficulty. In light of previous research on metacognition, people are sensitive to different features when assessing the likelihood of remembering the studied material. For example, research indicates that JOLs are influenced by changes in the to-be-studied material, including perceptual, lexical, and semantic features or the degree of coherence and elaboration. Therefore, people tend to judge that their memory will be better for easy-to-read items in contrast to *difficult-to-read* items (Yue et al., 2013); for concrete in contrast to abstract words (Hertzog et al., 2003; Tauber & Rhodes, 2012; Tullis & Benjamin, 2012), even under conditions of divided attention while studying (Pérez-Mata et al., 2002); for related pairs (Dunlosky & Matvey, 2001; Undorf & Erdfelder, 2015) and semantically related word lists (Matvey et al., 2006) in contrast to unrelated words; and for well cohesion texts in contrast to poorly cohesion texts (Carroll & Korukina, 1999; Lefèvre & Lories, 2004; Rawson & Dunlosky, 2002).

Thus, our aim was threefold. Firstly, we sought to investigate whether the manipulation involving perceptual, lexical, semantic, and coherence features was consistently observed in JOLs both L1 and L2. In other words, we aimed to understand how the effects of font-type, concreteness, relatedness, and cohesion manifested in the monitoring process while studying in both languages. Second, we aimed to assess whether participants adapted their overall perception of the learning context based on the language used. We also intended to examine the interplay between these two effects: language and the manipulation of the to-be-studied material. Finally, we aimed to explore the role that L2 proficiency may play on the process of SRL.

In five experiments, participants studied different sets of materials and provided JOLs for each of the study words (Experiments 1 and 2) [1], short word lists (Experiment 3) [1] or short texts (Experiment 4 and 5)[2]. For Experiments 1-4, participants were unbalanced late bilinguals, with medium level of English-L2. In experiment 5, we compared two groups of participants: a higher L2 proficiency group, and a lower L2 proficiency group. All participants performed the tasks in both L1 and L2, with language being a blocked

---

[1] Experiments 1-3 have been presented in a paper entitled "**Judgments of learning in bilinguals: Does studying in a L2 hinder learning monitoring?"** and has been published in *PloS ONE* (Reyes, Morales & Bajo, 2023).

[2] Experiments 4-5 have been presented in a paper entitled "**Self-regulated learning strategies in L1 and L2 reading**" and has been submitted for revision in *Bilingualism: Language and Cognition* (Reyes, Morales & Bajo, under review).

and counterbalanced variable in Experiments 1-4, and a mixed and randomized variable in Experiment 5. Across experiments, we varied the features of the materials that could be considered cues for metacognitive assessment. Experiment 1 concerned perceptual features (font type); Experiment 2 concerned lexical-semantic features (concreteness); Experiment 3 concerned semantic-relational features (relatedness among words in a short list) and Experiment 4-5 concerned organizational features (text cohesion).

Our manipulations involved different types of processing that may interact with the language of study in different ways (Dirix et al., 2020; Kuperman et al., 2021; Moreno et al., 2010; Pérez et al., 2018; Soares et al., 2019). Note that these manipulations are not equally predictive of learning success since, intrinsically, perceptual manipulation does not necessarily imply increased difficulty of the material neither when encoding nor during retrieval (Rummer et al., 2016). However, people do encode and retrieve concrete and categorically related words better than abstract and unrelated words, which makes them more memorable (Romani et al., 2008; Taconnat et al., 2020). Similarly, poor cohesion texts impose higher demands on readers who would need to produce inferences in order to create a meaningful representation of the information in the text (Best et al., 2005), and these texts are associated with poorer comprehension (Crossley et al., 2014, 2016; Hall et al., 2016). Hence, manipulations of the different features of the to-be-studied materials might reveal interesting interactions between language and monitoring.

Overall, our hypothesis was that the extra cognitive demands involved in L2 processing relative to L1 processing may reduce metacognitive processing and monitoring as they both (L2 processing and monitoring) require cognitive control. That is, when studying in

L2, people have less cognitive resources available to devote to metacognitive learning processes. Nevertheless, whether the language of study plays a part in the monitoring process and whether it interacts with other cues is yet to be known. We expected to observe less accurate use of possible cues for monitoring in L2 than in L1, meaning that manipulations to increase or decrease the intrinsic difficulty of the material –such as concreteness, relatedness or cohesion of the to-be-studied material– might not be detected when the task is performed in L2.

In addition, we wanted to assess whether participants adjust their overall perception of learning to the context provided by the language. Since each block in the Experiments 1-4 defined a linguistic context (either L1 or L2), it was possible to assess if the participants perceived learning success differently depending on the language and adjusted their JOLs consequently. Overall, we expected that participants considered learning in L1 to be easier and more successful than learning in L2, with higher JOLs in the L1 context than in the L2 context.

Part II

Experimental section

Experiment 1:

Easy-to-read vs. difficult-to-read font type

The font type effect refers to the phenomenon where the style or format of the to-be-studied material (e.g., font type) influences how individuals perceive their own learning (i.e., JOLs) and memory (Rosner et al., 2015; Susser et al., 2013; Thompson et al., 2013; but see Maxwell et al., 2021 for null results). For instance, studies suggest that using a *difficult-to-read* font may lead individuals to estimate their learning as more challenging, subsequently enhancing memory retention (Diemand-Yauman et al., 2011; Yue et al., 2013; but see Taylor et al., 2020 for null results). Conversely, easily legible fonts might instill a false sense of confidence in memory despite not necessarily improving actual retention. This effect demonstrates how a perceptive manipulation can subtly influence individuals' subjective assessments of their learning and later recollection. Experiment 1 represents an initial attempt to explore how the language of study interacts with this perceptual manipulation. With this purpose, we examined the effect of font type on JOLs and memory when the study materials were presented in L1 and L2.

Interestingly, the evidence for the effect of perceptual manipulations on JOLs and memory performance is mixed, with varying contributions across studies, and seems to depend on specific conditions (see Kühl & Eitel, 2016, for a review). For example, Magreehan et al. (2016) examined whether people use perceptual characteristics as a cue for JOLs when another cue such as item relatedness was present. For this, people studied related and unrelated pairs, with half of the pairs of each type presented in either perceptual format (easy-to-read vs. difficult-to-read). They found that participants focused on item relatedness as a cue, disregarding any impact of the perceptual manipulation. Nevertheless, when they eliminated item-relatedness and people only studied unrelated word pairs, they were able to obtain an effect of the perceptual manipulation on JOLs.

Regarding memory performance, results are also mixed. Some studies indicate that perceptual disfluency may function as a desirable difficulty since it provides a metacognitive cue for “difficulty,” leading to more effortful processing and, in turn, to better performance (Diemand-Yauman et al., 2011; Sungkhasettee et al., 2011; Yue et al., 2013). Yet, some other studies have concluded that perceptual disfluency does not affect memory (Rhodes & Castel, 2008; Rummer et al., 2016; Yue et al., 2013). Moreover, working memory capacity also seems to be a modifier in the effect. Some studies provide evidence that perceptual manipulation enhanced accuracy only for those of high cognitive ability (e.g., Alter et al., 2007; Thompson, Turner, et al., 2013; Yue et al., 2013). Then, the question is whether perceptual fluency might be boosted as a cue for learning monitoring when more effortful processing is dedicated to study due to the L2 context.

In Experiment 1, we varied font type (easy- vs. *difficult-to-read*) and language of study (L1 vs. L2) within participants, and language was blocked and counterbalance. Thus, within each block, font type was the main cue on which to base the JOLs, while language functioned as the contextual setting of learning (i.e., words within each block appeared in the same language). We wanted to examine not only whether there were differences in the way participants assessed their learning within each linguistic context, but also whether they adjusted their JOLs as a function of the linguistic context (e.g., higher JOLs for L1 than L2). Based on the literature reviewed, the effects of perceptual features on JOLs (and memory) are not robust and occur only under specific conditions (e.g., Diemand-Yauman et al., 2011; Yue et al., 2013; Taylor et al., 2020). The role of perceptual disfluency remains ambiguous and thus we feel cautious and tentative with regard to our expected results.

Concerning the language of study, we hypothesize that studying in L2 might pose greater challenges. Even for highly proficient bilinguals, working in L2 can be cognitively challenging (Ma et al., 2014; Moreno et al., 2010; Pérez et al., 2018). Language control poses a challenge, as research indicates greater interference from the dominant L1 to the less-dominant L2 compared to the other way around (e.g., Contemori & Dussias, 2016; Meuter & Allport, 1999; Soares et al., 2019). Additionally, unbalanced and late bilinguals heavily rely on transfer from L1 to L2, with weaker semantic representations in L2, often activating concepts via L1-L2 translation (Kroll & Stewart, 1994). Studies also reveal that L2 processing demands approximately 20% more time, exhibits slower word recognition, and processes a reduced amount of information simultaneously compared to L1 (Dirix et al., 2020). All this suggests that L2 processing is notably more demanding and may occur within a potentially overloaded cognitive system (Adesope et al., 2010; Hessel & Schroeder, 2020, 2022; Pérez et al., 2018). Thus, we expected that participants would use language as a diagnostic cue (Koriat, 1997) and assessed learning in L1 as easier and more successful compared to learning in L2. Consequently, they would assign higher JOLs in the L1 context than in the L2 context.

With respect to the perceptual manipulation, we may expect to find a significant effect of font-type on JOLs in L1. People may be sensitive to the perceptual cue and give higher JOLs to the easy-to-read words (Rosner et al., 2015; Susser et al., 2013; Thompson, Ackerman, et al., 2013). Nevertheless, in L2, it may be the case that the presence of two relevant cues (perceptual features and language-L2) may reveal an interesting pattern where one of them is more prevalent and overshadow the potential effect of the other. Previous evidence has been reported toward perceptual manipulations being ignored in JOLs under conditions of high cognitive load. For example, Luna et al. (2019) examined the font size effect (another perceptual manipulation that appeared to affect JOLs) with sentences that differed in their length (shorter vs. longer) and their relatedness. They found an effect of font size only for the shorter sentences and not the longer sentences condition. These outcomes align with the idea that the heightened cognitive load associated with processing longer sentences seems to reduce the use of font size as an indicator for JOLs. Whether font type serves as a distinctive cue (in L1 and L2) upon which people base their JOLs in remain to be seen.

Methods

Participants

We conducted a power analysis using G*Power (Faul et al., 2007) to determine the required sample size. We calculated it considering a mixed factor analysis of variance (ANOVA) with language and font type as repeated measure variables, and order of the language block as a between-participants variable. We estimated a required sample size of 28 participants, assuming a small-to-moderate effect size (partial eta squared of 0.05) to observe significant ($\alpha = 0.05$) effects at 0.8 power. This estimation applies for all three experiments, as they are all of similar design and characteristics. We recruited some more participants in order to ensure a representative sample after removing those who did not perform the task correctly: participants who did not vary JOLs across items, had a low hit and/or high false alarm proportion (d-prime < 0.5) in the recognition test (Macmillan & Kaplan, 1985), and/or had a rate of fast anticipatory responses (<300ms) of over 10% (Roessel et al., 2018) were excluded from the study. Participants had normal or corrected-to-normal vision and reported no neurological damage or other health problems. Participants gave informed consent before participating in the experiment. The experiment was carried out following the Declaration of Helsinki (Association, 2013). The protocol was approved by the institutional Ethical Committee of the University of Granada (857/CEIH/2019) and the Universidad Loyola Andalucía (201222 CE20371).

Thirty-seven psychology students from the University of Granada participated in Experiment 1. We removed from all analyses a participant who rate every item at the maximum possible value. We therefore had a total sample of 36 (18–40 years old, *M* = 23.67, *SE* = 5.34). Participants were tested in person and individually in the laboratory and received course credit as compensation.

We recruited non-balanced Spanish-English bilinguals who started acquiring English as their L2 during late childhood (*M* = 7.90, *SE* = 2.86). They were moderately proficient in English, as proven by subjective [Language Background Questionnaire, LEAP-Q;(Marian et al., 2007) and objective (MELICET Adapted Test, Michigan English Language Institute College Entrance Test, and verbal fluency task) language measures. Table 1 shows descriptive statistics for the participants' language measure.

Table 1

Participant information: Mean score (and standard deviations) for language measure.

	L1	L2
Self-reported measures		
Daily exposure L1 (%)	71.1 (14.7)	23.2 (10.9)
Age of acquisition (in years)		7.9 (2.9)
Age of becoming fluent (in years)		15.5 (5.0)
Speaking self-competence (0-10)		7.6 (0.9)
Reading self-competence (0-10)		8.2 (1.5)
Exposure to reading (0-10)		7.9 (2.0)
Learning by reading (0-10)		8.2 (1.5)
Language proficiency		
MELICET (0-50 points)		34.3 (7.2)
Verbal fluency L1	21.9 (4.1)	15.3 (4.5)

Note: Verbal fluency task shows the mean number of words elicited in each language condition. By error, we did not record the data for L1 in the LEAP-Q

Materials and Procedure

The experimental session lasted approximately 100 minutes. The main task consisted of a JOL task with a study phase, a distractor task, and a recognition test. Additionally, participants completed other cognitive tasks, a metacognitive questionnaire regarding the strategies used when studying the words, a language background and sociodemographic questionnaire (LEAP-Q; Marian et al., 2007), a verbal fluency test in L1 and L2, and the MELICET Adapted Test. We presented stimuli and collected data for all tasks with E-Prime Professional 2.0 software (Schneider et al., 2002).

The JOL task was modeled after Halamish (2018). This paradigm included two consecutive blocks with identical procedure with the exception of the language in which words were written (Spanish–L1 vs. English–L2). The assignment of L1 and L2 to the blocks was counterbalanced across participants as a between-subject factor. For each block, participants studied a list of words for a later recognition test and were informed that the words would be presented in two different font types. For each block, words could appear in either an easy-to-read (Arial 18 points black color, RGB (Decimal) 0, 0, 0 and RGB (Hex) 0x0, 0x0, 0x0) or *difficult-to-read* font (Monotype Corsiva 18 points silver color, RGB (Decimal) 192, 192, 192 and RGB (Hex) 0xC0, 0xC0, 0xC0), following the procedure established by previous studies (French et al., 2013; Seufert et al., 2017; Weissgerber & Reinhard, 2017). Each study phase lasted eight minutes, and the recognition test took approximately four minutes. During the study phase, words were presented one at a time in the middle of the computer screen. For each trial, a fixation point appeared for 500 ms; then, a slide with the study word remained for 5,000 ms. Immediately after the presentation of each word, participants gave a judgment of learning (JOL). They predicted the likelihood of remembering it on a 0–100 scale (0: not likely at all, 100: very likely). They typed in the JOL using a regular computer keyboard. This screen advanced automatically after the prescribed time (4,000 ms) or when the participant pressed ENTER.

For each block, the to-be-studied list comprised 44 words, with the first and last two words serving as the primacy and recency buffers and the remaining 40 as targets (testing materials are available at https://osf.io/vfykg). Language was blocked such that all words within each study/recognition block appeared either in L1 or L2. Within each language block (L1 or L2), participants studied 40 words (after removing primacy and recency buffers), half of them in an easy-to-read font and half of them in a *difficult-to-read* font type, which was counterbalanced across participants. For assignment to the easy/*difficult-to-read* conditions, we created two lists (list A and list B) of 20 words in each language. The assignment of each list to the easy-to-read or to the *difficult-to-read* font was counterbalanced across participants (see testing materials at https://osf.io/vfykg).

We selected English and Spanish words from the CELEX English Corpus (Baayen et al., 1995) and the LEXESP database (Sebastián et al., 2000) and used the N-Watch (Davis, 2005) and BuscaPalabras programs (Davis & Perea, 2005), respectively, to compute and control for psycholinguistic indices. Within and between lists and languages, words were matched for estimated frequency (L1-List A: $M = 2.2$, $SD = 0.4$; L1-List B: $M = 2.2$, $SD = 0.5$; L2-List A: $M = 2.3$, $SD = 0.4$; L2-List B: $M = 2.3$, $SD = 0.4$) and number of letters (L1-List A: $M = 4.8$, $SD = 1.0$; L1-List B: $M = 4.8$, $SD = 0.9$; L2-List A: $M = 5.1$, $SD = 1.1$; L2-List B: $M = 4.8$, $SD = 0.8$). Within the blocks, words were presented in a pseudo-random order, with the restriction that no more than three items from the same font type appeared consecutively.

In between the study phase and the recognition test of each block, participants did a distractor task for 10 minutes. We chose a short version of the AX- Continuous Performance Task [AX- CPT; (J. Morales et al., 2013), which is a cognitive control task with minimum verbal load.

With regard to the recognition test, studied words (excluding the primacy and recency buffers) appeared along with 40 new words in a random order. Studied and new words were two independent sets that remained constant for all participants, but they were matched for mean estimated frequency (L1: M = 2.1, SD = 0.3; L2: M = 2.3, SD = 0.3), and mean number of letters (L1: M = 4.7, SD = 1.2; L2: M = 4.4, SD = 0.8), so that any possible effect that may arise would not be explained by those psycholinguistic parameters (see Lanska et al., 2014; Wehr & Wippich, 2004; Yue et al., 2013) for a similar procedure). First, a blank slide was displayed for 100 ms. Then, the target stimulus remained on the screen for 3,000 ms or until the participant gave a response. For each word, participants indicated whether it had appeared in the study phase by pressing a 'YES' or 'NO' key. The assignment of the keys (Z and M) to the correct responses ('YES' and 'NO') was counterbalanced between subjects and kept constant across tasks.

Results

We performed 2 x 2 x 2 (language x font type x block order) mixed-factor ANOVAs for JOLs in the study phase and accuracy (d-prime) in the recognition test. Language (L1 vs. L2) and font type (easy-to-read vs. *difficult-to-read*) were within-subject factors, and block order (L1-first vs. L2-first) was a between-subject factor. We included block order in the analyses because participants' calibration and expectations when performing the task may vary as a function of whether the first block was performed in L1 or L2. For all analyses, the alpha level was set to 0.05, and we corrected by Bonferroni for multiple comparisons. Effect sizes are reported in terms of partial eta squared (η_p^2) for ANOVAs and Cohen's *d* for t-tests.

We also conducted the same analyses but including MELICET scores as a covariate (ANCOVAs), which yielded identical results. MELICET did not interact with language $F(1, 33) = 0.749$, $p = .393$, $\eta_p^2 = .002$, or font type $F(1, 33) = 0.05$, $p = .829$, $\eta_p^2 = .000$ for JOLs or recognition task – language $F(1, 30) = 0.29$, $p = .593$, $\eta_p^2 = .001$ and font type effect $F(1, 30) = 0.02$, $p = .887$, $\eta_p^2 = .000$). Therefore, for the sake of simplicity, we report ANOVAs outcomes in the main text.

We removed three duplicate items in the L1 block, which resulted in 76 valid trials (37 studied and 39 new items) in the L1 block and 80 valid trials (40 studied and 40 new items) in the L2 block. Measures were adjusted for the total number of valid items in each language block.

Study phase (JOLs). To evaluate the effect of language and font type on the magnitude of JOLs, we computed the mean across participants after removing trials with blank responses (0.63%) and trials with responses over 100 (0.77%, presumably due to typography errors, as participants were instructed to rate their JOLs on a 0–100 scale by key-pressing a value).

We found no significant main effects of language ($F(1, 34) = 0.63$, $p = .432$, $\eta_p^2 = .02$), font type ($F(1, 34) = 2.18$, p = .149, $\eta_p^2 = .06$), or block order ($F(1, 34) = 0.01$, $p = .922$, $\eta_p^2 = .000$). We observed a significant interaction between language and block order ($F(1, 34) = 9.47$, $p = .004$, $\eta_p^2 = .22$). Post-hoc comparisons revealed that when L2 was studied first, JOLs for L2 items were lower (*M* = 57.8, *SE* = 4.7) than for L1 items (*M* = 66.8, *SE* = 3.9), although these differences were marginal ($t(34) = 2.74$, $p = .059$). In contrast, when participants started with the L1 block, they rated comparably the probability of remembering items in both languages (*M* = 60.2, *SE* = 3.9 and *M* = 65.5, *SE* = 4.7 for L1 and L2 items, respectively; $t(34) = -1.61$, $p = .696$). No other interaction was significant: font type did not interact with block order ($F(1, 34) = 0.61$, $p = .44$, $\eta_p^2 = .02$) or with language ($F(1, 34) = 0.39$, $p = .535$, $\eta_p^2 = .01$). Block order and language did not interact neither (reporter) and the three-way interaction between font type, language, and block order was not significant ($F(1, 34) = 0.47$, $p = .499$, $\eta_p^2 = .01$). Overall, it seemed that font type did not have an effect in either L1 or L2.

Recognition test (accuracy). Following the procedure of Undorf & Zander (2017), we removed from the analysis trials with a reaction time shorter than 300 ms (0.14% of the total number of trials). We calculated d-prime as a sensitivity index on the basis of hits and false alarms. See S1 Table in the Supporting Information for estimated means (and standard deviations) for hits, false alarms, misses and correct rejections by language, font type and block order. Greater d-prime indicates better discrimination between studied and new items. We followed Hautus (Hautus, 1995) and the 1/(2N) rule to apply corrections for extreme false-alarm or hit proportions (p = 0 or p = 1). Due to a technical error, we did not record data from the recognition test for the first three participants. Therefore, we analyzed data from 33 participants in this measure.

The analysis showed that d-prime did not differ across conditions. Neither language ($F(1, 31) = .61$, $p = .441$, $\eta_p^2 = .02$), font type ($F(1, 31) = 0.70$, $p = .41$, $\eta_p^2 = .02$), nor block order ($F(1, 31) = 3.55$,

p = .069, η_p^2 = .10) reached significance. None of the interactions were significant. Block order did not interact with language (F(1, 31) = .15, p = .701, η_p^2 = .01) or font type (F(1, 31) = .10, p = .751, η_p^2 = .003). Neither did font type interact with language (F(1, 31) = 1.85, p = .184, η_p^2 = .06). The three-way interaction between the factors did not reach significance (F(1, 31) = 0.72, p = .40, η_p^2 = .02). Participants recognized items similarly across conditions. See Table 2 for estimated means and standard deviations for JOLs and d-prime.

Table 2.

JOLs and d-prime across conditions.

		Easy-to-read		Difficult-to-read	
Language	Block order	JOL	d-prime	JOL	d-prime
L1	L1-first	60.0 (4.0)	2.9 (0.2)	60.4 (4.1)	2.9 (0.2)
	L2-first	67.9 (4.0)	2.4 (0.2)	65.8 (4.1)	2.5 (0.2)
L2	L1-first	66.4 (4.6)	2.8 (0.2)	64.7 (4.7)	2.7 (0.2)
	L2-first	58.8 (4.6)	2.5 (0.2)	56.8 (4.7)	2.4 (0.2)

Goodman–Kruskal gamma correlation. We used a Goodman–Kruskal (GK) gamma correlation (Nelson, 1984) as a nonparametric measure of the association between JOL and subsequent recognition. This analysis permitted us to examine participants' metamemory accuracy –resolution– across conditions. We calculated one gamma correlation for each participant in each of the four conditions of interest (L1 easy-to-read, L1 *difficult-to-read*, L2 easy-to-read, L2 *difficult-to-read*). We then ran mixed-factor ANOVAs to examine whether the GK gamma correlations differed across conditions, including block order as a between-subject variable. Note that the degrees of freedom may differ from the previous analyses because the correlation cannot be computed when there is not enough variance in participants' responses (Blake & Castel, 2018).

We found a marginal effect of font type ($F(1, 16) = 4.18$, $p = .058$, $\eta_p^2 = .207$), with the *difficult-to-read* font type ($M = 0.3$, $SE = 0.1$), having better resolution than the easy-to-read font type ($M = 0.2$, $SE = 0.1$). This was mediated by a marginal interaction between language and font type ($F(1, 16) = 4.18$, $p = .058$, $\eta_p^2 = .207$). We observed a tendency towards having better resolution for the difficult materials in L2, whereas easy-to-read and *difficult-to-read* materials did not differ in L1 (see Table 3). The main effect of block order was not significant ($F(1, 16) = 0.04$, $p = .85$, $\eta_p^2 = .002$). Neither was any of the other interactions (all *p*s > .05).

Table 3.

Goodman–Kruskal gamma correlations across conditions.

		Easy-to-read	Difficult-to-read
Language	Block order	JOL	JOL
L1	L1-first	0.2 (0.2)	0.2 (0.2)
	L2-first	0.3 (0.1)	0.2 (0.2)
L2	L1-first	0.0 (0.2)	0.3 (0.2)
	L2-first	0.1 (0.1)	0.3 (0.2)

With regard to the learning strategies used in the study phase, in this experiment participants reported grouping words by their semantic meaning (86.1%), creating mental images (69.4%) and rehearsal of words (52.8%) as the strategies most used.

Discussion

In Experiment 1, we were interested in two possible effects. First, we wanted to observe if a perceptual cue, such as font type, produced different JOLs and recognition accuracy and if they differed as a function of the linguistic context in which the task was performed. Second, we were interested in assessing if the linguistic context (L1 or L2) had an effect on the overall perceived learning difficulty of the task.

Regarding the perceptual manipulation, we did not find an effect of font type on JOLs or memory in either language (L1 or L2). Participants predicted similar memory performance for words in a *difficult-to-read* and easy-to-read font. Correspondingly, recognition accuracy was similar for both font conditions. This pattern of results is in line with Magreehan et al. (2016), who only observed the effect of perceptual fluency when item-relatedness was eliminated as a cue for JOLs. It can be argued that with our design and materials, font-type was the only cue available for the participants since language was blocked. However, participants had been fully informed of the procedure, and they knew from the beginning that they were going to study words in two languages and that within each language block, words could appear in two different font types. They were instructed to judge their learning based on the difficulty perceived with all the information available, which includes font type and language. Results also align with Luna et al. (2019), who could not find an effect of font size under a high cognitive load condition. With respect to the lack of an effect of JOLs on memory performance, other studies have found a beneficial impact of *difficult-to-read* items on memory, particularly when participants could adjust their reading times. In fact, previous research has reported additional and beneficial encoding of the blurred words when the processing time was long enough (Yue et al., 2013). However, this was not possible under the time constrain of our design, and therefore this may have obscure possible effects.

In fact, despite participants reporting similar JOLs in both font type conditions, the gamma correlations revealed an intriguing pattern. It appeared that they displayed a more effective monitoring process for the more challenging material compared to the easier material. This effect seems to suggest better adjustment between perceived degree of learning and actual recognition performance. That is, when participants performed the task in L2, JOLs for remembered *difficult-to-read* items were higher than for unremembered ones, suggesting they monitored difficult materials better. However, this interaction was only marginal, and it should be considered with caution. One possible explanation might derive from the fact that L2 processing is more demanding than L1 processing. Thus, even basic language tasks like lexical access result in an increased activation in brain regions linked to cognitive control when performed in L2 (Ma et al., 2014). Research also indicates that L2 processing requires greater working memory resources compared to L1 processing (Dussias & Piñar, 2010; Ransdell et al., 2001). These increased cognitive demands might prompt heightened monitoring and more profound engagement with the material during L2 study. Consequently, the L2 learning block within our study could have elicited a more profound level of control compared to the L1 block, possibly leading to a perceptual effect that was slightly apparent when measuring accuracy in L2.

Interestingly, the language effect was dependent on the order in which the languages were presented. Thus, participants' JOLs increased for L1 when L2 was presented first, whereas when the L1 block preceded the L2 block, differences in the perceived difficulty of the language did not reach significance. Although block order did not modulate the GK correlations, still this effect might be due to the possible anchor point for further comparison provided by the first block. It is possible that participants were cautious in judging their degree of learning during the initial L2 block and increased their JOLs when confronted with the following, easier L1 block. This increase in perceived learning for the second block was not evident when the second block was L2. Note, however, that L2 receiving lower JOLs when L2 was presented first did not correspond with performance in the recognition test, since recognition did not vary with language or block order. Participants' recognition performance remained consistent regardless of the language or the order in which the blocks were presented. It seems that people could overcome the self-perceived difficulty of L2-first blocks to achieve successful learning. Previous research exploring memory recall and retention of study materials in both L1 and L2 suggests that individuals might remember study materials similarly in both languages, or even have a general bilingual advantage for novel word learning (Kaushanskaya & Marian, 2009).

In sum, the results of Experiment 1 suggest that when font type was manipulated, there were only very small variations in JOLs depending on the block order. Recognition accuracy did not vary with language or font type. These findings emphasize the limited impact of font type manipulation on JOLs and recognition accuracy regardless of the linguistic context. Next, in Experiment 2 we further explore whether a lexical-semantic manipulation, which induce a deeper level of processing, may interact with the language of study in JOLs and recognition. Previous research has accounted for differences observed in conceptual processing between languages (Farley et al., 2012; Kaushanskaya & Rechtzigel, 2012; Paolieri et al., 2019). For example, there are weaker associations between words and their meanings in L2 compared to the L1, particularly for unbalanced bilinguals, who tend to exhibit asymmetric language processing favoring the more proficient language (see the Revised Hierarchical Model -RHM- by Kroll & Stewart, 1994). The dominant and more proficient language has stronger connections to semantic representations and heightened conceptual access. In contrast, L2 often relies more on translation mechanisms or transfer from the L1 during language processing. Therefore, semantic access in the L2 is mediated by the L1 in unbalanced bilinguals with asymmetrical language abilities. Given the differences in lexical-semantic processing between languages, the language of study could significantly influence the learning process when the to-be-studied materials trigger a lexical-semantic processing.

Experiment 2:

Concrete vs. abstract words

The concreteness effect represents how the level of tangibility or abstractness in information influences cognitive processing, mental representations and language comprehension. In Experiment 2, we introduced a lexical-semantic manipulation by including concrete and abstract words in the study list. Concrete words represent tangible objects and usually invoke familiar mental images. According to the Dual Coding Theory, concrete words have richer and more interconnected semantic representations than abstract words (Paivio, 1991), and have been shown to enhance item memory in tasks such as free, serial, and cued recall (e.g., Holmes & Langford, 1976; Richardson, 2003; Romani et al., 2008; Walker & Hulme, 1999) and recognition (e.g., Glanzer et al., 1993; Glanzer & Adams, 1985; Hirshman & Arndt, 1997), as well as in lexical processing tasks (e.g., Schwanenflugel et al., 1988; van Hell & de Groot, 1998, 2008).

The concreteness effect have been demonstrated not only in memory but also in JOLs (Hertzog et al., 2003; Witherby & Tauber, 2017). Previous evidence suggests that the mechanisms that drive the concreteness effect on JOLs are twofold. On the one hand, concrete words are more easily processed, and this encoding fluency serves as a cue for metacognitive judgments, such as JOLs (Hertzog et al., 2003). On the other hand, it has been proposed that beliefs about own's memory play a significant role as a primary mechanism driving the impact of concreteness on JOLs. That is, people have preexisting beliefs about the impact of concreteness on memory performance. For example, to evaluate beliefs, Witherby & Tauber (2017) asked participants to predict their memory performance by providing a pre-study JOL for each item they were about to study (concrete and abstract words). Participants did not have access to the word pairs during these pre-study JOLs. Thus, they likely relied on their beliefs regarding how different types of items (concrete vs. abstract) might affect memory. Results indicated that participants' pre-study JOLs were higher for concrete compared to abstract words, indicating that they applied their beliefs about concreteness on an item-by-item basis when constructing their JOLs. Regardless of the underlying mechanism of this effect, the impact of concreteness on JOLs is evident—participants tend to assign higher JOLs to concrete words compared to abstract words (e.g., Begg et al., 1989; Hertzog et al., 2003; Tauber & Rhodes, 2012; Tullis & Benjamin, 2012).

We expected that this manipulation might interact with language, since previous studies have revealed variations in conceptual processing across languages (Farley et al., 2012; Kaushanskaya & Rechtzigel, 2012; Paolieri et al., 2019). Thus, associations between words and their meanings have been shown to be weaker in L2 than in L1, especially for unbalanced bilinguals (Kroll & Stewart, 1994), and this may have an effect when the cue for learning monitoring also involves conceptual processing. According to the Revised Hierarchical Model (RHM) by Kroll & Stewart (1994), unbalanced bilinguals tend to exhibit asymmetrical language processing and the more proficient language is accessed more readily and efficiently. This dominant language holds a stronger connection to semantic representations and conceptual access. In contrast, the less proficient language might rely on translation mechanisms or transfer to the dominant language during language processing. That is, the semantic access in L2 is mediated by the L1 in asymmetrical and unbalanced participants.

Our prediction was that, consistent with previous studies, participants might give higher JOLs to concrete relative to abstract words in L1 (Witherby & Tauber, 2017), but in L2 we might find two different trends. On the one hand, individuals might have difficulties in monitoring their learning accurately due to potentially higher cognitive demands imposed by L2 processing (Adesope et al., 2010; Hessel & Schroeder, 2020, 2022; Pérez et al., 2018) and to the fact that the semantic access in L2 is less direct, as it is mediated by the L1 (Kroll & Stewart, 1994). This could lead to less apparent effects of concreteness on JOLs and memory.

On the other hand, research in bilingualism has shown that, when learning L2 vocabulary, concrete words are more easily learned than abstract words (Altarriba & Bauer, 2004). Moreover, Kaushanskaya & Rechtzigel (2012) showed that bilingual individuals benefited from semantic details linked to new words that needed to be learnt. They manipulated the concreteness of the referent in a word-learning task and they found that the advantage seen in bilinguals was stronger for newly acquired words associated with concrete referents rather than abstract ones. This suggests that the effects of concreteness on memory and language processing might be more pronounced in bilingual individuals compared to monolinguals due to the increased activation of the lexical-semantic network. Thus, it seems plausible that L2 processing might be sensitive to our lexical-semantic manipulation, and that people might still be able to detect the difference in difficulty and adjust their JOLs accordingly.

Regarding the language of study, we hypothesize that studying in L2 could present increased difficulties (e.g., Ma et al., 2014; Moreno et al., 2010; Pérez et al., 2018). We predicted that participants would use language as a diagnostic cue (Koriat, 1997) and perceive learning in their L1 as more straightforward in comparison to learning in their L2. As a result, they were expected to assign higher JOLs in the L1 than in the L2 context. In addition, we wanted to explore if, similar to Experiment 1, we would observe a language-by-block interaction, indicating that participants adjust their JOLs depending on the anchor point provided by the language of the first block.

Methods

Participants

Participants were selected following the same criteria and procedure described in Experiment 1. Thirty-nine psychology students from the University of Granada participated in this experiment. We removed a participant who gave JOL values of 100% to all items, suggesting he/she did not perform the main task correctly, resulting in a final sample of 38 (18–31 years old, $M = 21.37$, $SD = 2.86$). Participants were tested individually in an online experiment and received course credit as compensation. In this experiment, we included self-reported measures for L1 in LEAP-Q. Comparisons of all self-reported measures and of the verbal fluency test results showed that participants were unbalanced and significantly more fluent in L1 than in L2. All *p* values were below .05. See Table 4. for descriptive statistics.

Table 4.

Participant information: Mean score (and standard deviations) for language measure.

Self-reported measures	L1	L2
Daily exposure L1 (%)	83.4 (17.4)	40.3 (27.8)
Age of acquisition (in years)	2.8 (1.7)	5.8 (2.9)
Age of becoming fluent (in years)	5.1 (2.2)	12.6 (3.2)
Speaking self-competence (0-10)	9.6 (0.9)	7.8 (1.4)
Reading self-competence (0-10)	8.9 (0.9)	8.9 (0.9)
Exposure to reading (0-10)	8.6 (1.9)	7.3 (2.2)
Learning by reading (0-10)	9.2 (1.4)	8.6 (1.4)
Language proficiency		
MELICET (0-50 points)		36.2 (5.3)
Verbal fluency L1	25.5 (4.2)	18.0 (3.9)

Note: Verbal fluency task shows the mean number of words elicited in each language condition.

Materials and Procedure

Participants were tested in a single online remote session that lasted approximately 120 minutes. We programmed, presented the stimuli and collected data for all tasks with Gorilla Experiment Builder, an online platform (Anwyl-Irvine et al., 2020). Participants accessed the experiment individually and on their own. They were forced to full-screen presentations so as to prevent them from opening other windows in the computer while doing the tasks. Recent research supports the validity and precision of experiments run online (Anwyl-Irvine et al., 2020, 2021; Gagné & Franzen, 2023).

The procedure was similar to that of Experiment 1, since the same cognitive and linguistics tasks were administered, although they were administered through an online platform in this case. In addition, for the memory and JOL tasks, word concreteness was manipulated. In this experiment, for the JOL task, participants responded by using the mouse to move a handle slider to the desired number and pressed the spacebar to continue to the next word. As in Experiment 1, the language of the study phase and test (L1 and L2) was blocked and counterbalanced. The study lists were composed of 44 nouns (4 primacy and recency buffers and 40 targets), and the subsequent recognition tasks included 80 words (40 targets and 40 new words). Half of the study and recognition words were concrete (concreteness for L1: $M = 5.8$, $SD = 0.5$; L2: $M = 4.6$, $SD = 0.4$) and half were abstract (L1: $M = 3.8$, $SD = 0.7$; L2: $M = 2.6$, $SD = 0.7$), and they were presented in random order. See testing materials at https://osf.io/3e2vy. We selected words from Brysbaert et al. (Brysbaert et al., 2014) and translated them to obtain words in Spanish–L1. Across participants, the L1-L2 versions of the words were counterbalanced in such a way that words that appeared in L1 for one participant would not appear in the L2 block. All selected words were composed between 3 and 7 letters and medium frequency. Within languages, concrete and abstract studied and new words were matched in estimated frequency (L1: $M = 2.0$, $SD = 0.3$; L2: $M = 2.2$, $SD = 0.3$) and numbers of letters (L1: $M = 5.3$, $SD = 1.2$; L2: $M = 5.0$, $SD = 1.2$). Note that we used two language specific norms to select the words. Concreteness ratings for English words were based on Brysbaert et al. (2014) using a 5-point scale for English, whereas values for the Spanish words were based on LEXESP (Sebastián et al., 2000) using a 7-point scale. Thus, the descriptive statistics are in different scales (see S4 Table in the Supporting Information for concreteness ratings –mean and SD– of each list). However, the criteria to consider a word abstract or concrete was equivalent for both data set. We calculated the concreteness mean for each language, and words with ratings above the means in both languages were considered concrete, while words with values below the means were considered abstract.

Results

As in Experiment 1, we report mixed-factor ANOVAs for JOLs and d-prime in the recognition test. Language (Spanish–L1 vs. English–L2) and concreteness (abstract vs. concrete words) were within-subject factors, and block order (L1-first vs. L2-first) was a between-subject factor. As in Experiment 1, we conducted the same analyses but including MELICET scores as a covariate (ANCOVAs), which yielded identical results. Scores in the MELICET did not interact with language $F(1, 35) = 2.67$, $p = .112$, $\eta_p^2 = .071$, or concreteness $F(1, 35) = 0.66$, $p = .420$, $\eta_p^2 = .019$ for JOLs or for recognition –language $F(1, 33) = 0.11$, $p = .739$, $\eta_p^2 = .003$, and concreteness effect $F(1, 33) = 3.56$, $p = .068$, $\eta_p^2 = .097$).

We removed two items that were erroneously duplicated in both blocks (Spanish–L1 and English–L2). This resulted in 79 valid trials (40 studied and 39 new items) in both blocks. Measures were adjusted for the total number of valid items in each language block. We removed from all analyses a participant who did not vary the JOLs across the items, suggesting he/she was not performing the task correctly.

Study phase (JOLs). The results showed no significant main effects of language ($F(1, 36) = 0.17$, $p = .684$, $\eta_p^2 = .005$) or block order ($F(1, 36) = 0.15$, $p = .904$, $\eta_p^2 = .000$). We observed a significant main effect for concreteness ($F(1, 36) = 18.34$, $p < .001$, $\eta_p^2 = .34$) such that concrete words received higher JOLs ($M = 59.4$, $SE = 3.0$) than abstract words ($M = 55.0$, $SE = 2.8$).There was a marginal interaction between language and block order ($F(1, 36) = 4.00$, $p = .053$, $\eta_p^2 = .10$). Follow-up tests revealed non-significant effects. Neither was there a significant difference between languages depending on whether the L1 block (L1: $M = 54.5$, $SE = 4.3$; L2: $M = 59.1$, $SE = 4.5$) or L2 block (L1: $M = 59.1$, $SE = 4.1$; L2: $M = 56.0$ $SE = 4.3$) was placed first. However, we observed a tendency of crossover effects such that L1 had lower JOLs when the L1 block was placed first and L2 received lower JOLs when the L2 block was placed first. No other interaction was significant: concreteness did not interact with block order ($F(1, 36) = 0.44$, $p = .513$, $\eta_p^2 = .012$), or with language ($F(1, 36) = 2.08$, $p = .158$, $\eta_p^2 = .055$). The three-way interaction was not significant ($F(1, 36) = 0.00$, $p = .931$, $\eta_p^2 = .000$).

Recognition test (accuracy). Following the procedure in Experiment 1, we filtered out trials with fast responses (<300ms, 0.21%). We removed a participant who had a reaction time below 300ms in more than 10% of trials (Roessel et al., 2018) and another participant whose d-prime was below 0.5 (low hit or high false alarm proportion) (Macmillan & Kaplan, 1985). See S2 Table in the Supporting Information for estimated means (and standard deviations) for hits, false alarms, misses and correct rejections by language, concreteness and block order

For d-prime, the analysis showed a significant main effect of concreteness ($F(1, 34) = 29.95$, $p < .001$, $\eta_p^2 = .468$). Participants recognized concrete words (*M* = 2.6, *SE* = .12) better than abstract words (*M* = 2.3, *SE* = .11). In addition, there was a significant language-by-block order interaction effect ($F(1, 34) = 6.96$, $p = .013$, $\eta_p^2 = .170$). Thus, words in L2 were better recognized (*M* = 2.7, *SE* = 0.17) than words in L1 (*M* = 2.3, *SE* = 0.17), but only when the L2 block was placed first ($t(34) = -3.03$, $p = .028$). When participants started with the L1 block, they recognized items in both languages similarly (L1: *M* = 2.5, *SE* = 0.18; L2: *M* = 2.4, *SE* = 0.18, $t(34) = 0.76$, $p = .100$). We did not observe other significant main effects or interactions ($p > .1$ for all). See Table 5 for estimated means and standard deviations for JOLs and d-prime.

Table 5.

JOLs and d-prime across conditions.

		Concrete		Abstract	
Language	Block order	JOL	d-prime	JOL	d-prime
L1	L1-first	56.9 (4.6)	2.7 (0.2)	52.1 (4.2)	2.4 (0.2)
	L2-first	62.1 (4.4)	2.5 (0.2)	56.1 (4.0)	2.1 (0.2)
L2	L1-first	60.5 (4.7)	2.5 (0.2)	57.8 (4.5)	2.3 (0.2)
	L2-first	58.1 (4.4)	2.8 (0.2)	53.9 (4.3)	2.6 (0.2)

Goodman–Kruskal gamma correlation. In the mixed-factor ANOVA, we found no significant effect of block order ($F(1, 19) = 0.19$, $p = .665$, $\eta_p^2 = .01$) (L1-first: *M* = 0.2, *SE* = 0.1; L2-first: *M* = 0.2, *SE* = 0.1), language ($F(1, 19) = 0.60$, $p = .449$, $\eta_p^2 = .031$) (L1: *M* = 0.2, *SE* = 0.1; L2: *M* = 0.3, *SE* = 0.1), or concreteness ($F(1, 16) = 0.14$, $p = .709$, $\eta_p^2 = .008$) (concrete: *M* = 0.2, *SE* = 0.1; abstract: *M* = 0.2, *SE* = 0.1). None of the interactions were significant.

With regard to the learning strategies used in the study phase, in this experiment participants reported creating mental images (68.8%), words rehearsal (59.4%), grouping words by their semantic meaning (59.4%), and relating words to personal experiences (56.3%) as the strategies most used in the study phase.

Discussion

The results of Experiment 2 replicated the concreteness effect that has been commonly reported in previous metamemory studies with monolingual participants; that is, concrete words produced higher JOLs and better recognition rates than abstract words (Undorf et al., 2018; Witherby & Tauber, 2017). For bilingual individuals, the relationship between word concreteness and memory can be influenced by various factors, including the context of learning, language dominance, bilingual experience, and the degree of overlap between the two languages. For instance, Kaushanskaya & Rechtzigel (2012) explored how bilingual individuals may be benefited from semantic details related to new words. They modified the concreteness of the reference when learning new words. Findings revealed that the bilingual advantage was larger for newly learned words associated with tangible references than with abstract ones. This implies that the impact of concreteness on memory and language processing is probably more marked in bilingual individuals than in monolinguals due to heightened activation of the lexical-semantic network. Nevertheless, results of our study failed to confirm this prediction, such that the strength of the concreteness effects in both, JOLs and recognition test did not differ across languages. This seems to suggests that the memory representations of concrete words differed from those of abstract words in both languages, enabling their use as cues for participants' JOLs. Although assessment of the representations of concrete and abstract words in the two languages based only on behavioral data is not feasible, our findings offer evidence supporting their usage as cues in both languages.

Consistent with previous studies (Pelegrina et al., 2000), despite judging abstract words as more difficult than concrete words, participants did not seem to allocate sufficient resources to compensate and achieve the same recognition rates as concrete words. Importantly, these effects were evident independently of whether participants performed the tasks in L1 or L2. The absence of supporting evidence for compensation aligns with prior research that suggests individuals do not entirely offset the effects of item difficulty through self-regulation. It is possible that although the mental representations for abstract words might be objectively more difficult to encode and retrieve from memory than concrete words (Mueller et al., 2013; Romani et al., 2008; Soderstrom & McCabe, 2011), they might not provide a sufficient level of awareness to induce participants to engage in control strategies for compensation.

As for the language cue, there was a tendency to judge L1 words as better learned than L2 words. Interestingly, when the L2 block was placed first, participants seemed to compensate and achieve better memory for L2 words than for L1 words. In this respect, there was a discrepancy between what participants predicted (i.e., JOLs) and what they actually remembered (i.e., recognition test). Although there is a positive trend in successful resolution regarding the lexical-semantic manipulation, it was not as prominent concerning the language cue. Consequently, the Goodman-Kruskal correlation did not indicate a significant effect.

The fact that L2 words were better recognized than L1 words when L2 was studied first, aligns with previous research findings that demonstrate enhanced recognition in the less proficient language

(Francis & Gutiérrez, 2012b; Francis & Strobach, 2013). This effect has been previously linked to factors such as the increased episodic distinctiveness associated with L2 words, reduced familiarity of these words, and the heightened requirement for cognitive resources in L2. For example, Francis & Strobach (2013) asked two groups of participants, Spanish-English bilingual and monolingual individuals, to study high- and low- frequencies words on each language. Both monolinguals and bilinguals displayed better recognition for low-frequency words compared to high-frequency ones. More interestingly, though, bilinguals showed better recognition rates in L2 than in L1. Furthermore, the recognition of bilinguals in their L2 surpassed that of monolinguals. They suggested that the advantage seen in bilinguals' L2 recognition mirrors the effects of word frequency in recognition and is attributed to the heightened distinctiveness of L2 words compared to L1 words in memory. Thus, L2 words may stand out more due to their novelty, lower familiarity or less frequent use, leading to increased attention during encoding and better recognition of L2 words compared to more familiar L1 words. This phenomenon carries significant implications, particularly in contexts where people need to study new materials on an L2 basis. It suggests a potential advantage or heightened sensitivity towards L2 words during recognition tasks, which could hold promise for educational settings. Individuals learning in their L2 might demonstrate improved memory recognition for acquired materials in that language.

Moreover, it is possible that participants devoted more resources to what they perceived as slightly more difficult (L2), and ultimately achieved better learning outcomes. Interestingly, this compensatory effect was evident in Experiment 2 but not in

Experiment 1. The distinction between both experiments lies in the nature of the cues manipulated in the to-be-studied materials. Experiment 2 included concrete and abstract words, likely prompting semantic processing, which may have driven this effect. Conversely, the materials in Experiment 1 might not have elicited a sufficiently deep processing to enhance retention of L2 words. The lexical-semantic manipulation in Experiment 2 might have induced deep processing, which could account for the advantage observed in recognizing L2 words. In Experiment 3, we further evaluated this explanation by introducing a relational-semantic dimension within the study materials. This manipulation was intended to prompt participants to engage in a deeper level of processing. Individuals might establish connections, associations, or relationships among the elements presented, thereby potentially amplifying and making more apparent any compensatory effects observed in the previous experiment.

Experiment 3:

Words grouped into semantic categories vs. unrelated words

In Experiment 3, we introduced a semantic manipulation, namely the level of semantic organization within a list. The degree of within-list semantic organization requires relational processing and semantic integration (Taconnat et al., 2020). Relational processing and organization are among the most efficient processes for learning (Bousfield & Cohen, 1956). Organization involves awareness of the sematic relations of the material during encoding and the use of this organization at retrieval. Hence, organization as a learning strategy involves a high degree of metacognitive processing at both encoding (assessment of possible word relations) and retrieval (controlled organizational strategies). Previous research on metamemory has reported on the effect of relatedness in memory and JOLs. People systematically give higher JOLs, and indeed recall and recognize related information (pairs or lists), better than information that is unrelated (Janes et al., 2018; Mueller et al., 2013; Undorf & Erdfelder, 2015). However, research across older and younger participants has also shown that organization at encoding and retrieval is often impaired in older participants (Denney, 1974; Howard et al., 1981; Taconnat et al., 2009, 2020; West & Thorn, 2001; Zivian & Darjes, 1983), suggesting that the use of these strategies involves the engagement of fully intact control processes. Previous research on L2 language processing has also shown that processes such as inferencing or mental model updating during text comprehension are less efficient when the texts are presented in L2 relative to L1 (Pérez et al., 2018). Hence, it is possible that engagement of language control during L2 processing might also reduce the use of costly encoding and retrieval strategies relative to L1 processing.

Consistent with previous literature, we expected to find an effect of relatedness in both JOLs and recognition in L1. That is, related word lists might receive higher JOLs and would indeed be better recognized in a later memory test than unrelated words (Janes et al., 2018; Mueller et al., 2013; Undorf & Erdfelder, 2015). However, we might expect a different trend in L2. As previously mentioned, the lexical-semantic access in L2 is mediated by the L1 (Kroll & Stewart, 1994). In this line, findings from an event-related potentials (ERP) study conducted by Opitz & Degner (2012) indicated that the processing of affective valence in L2 words occurs less immediately compared to L1 words. They conclude that this delay might imply a lag in lexical access during L2 processing. Similarly, Zhang et al. (2020) asked native English speakers and L2 English speakers to judge the semantic relatedness of English words while scanned in functional magnetic resonance imaging (fMRI). They found that, at the behavioral level, L1 speakers performed the task more quickly and accurately as compared to L2 English speakers. Moreover, neurocognitive data indicated that L2 processing of action words induced greater brain activation compared to object words within crucial brain language regions. Notably, while both L1 and L2 processing engaged extensive brain networks, noticeable differences emerged. For L1 processing, words engaged a more interconnected brain network linking language centers with semantic integration nodes. However, during L2 processing, the connections with the semantic integration hub was less strongly activated, suggesting a differential engagement of brain regions.

As our material requires lexical-semantic integration and deeper associative processing compared to those in Experiments 1 and 2, we expected that it would be affected by the possibly costlier monitoring and memory processes in L2. Thus, the relatedness effect on JOLs and recognition might be less pronounced in L2, as the increased cognitive demands associated with processing an L2 (compared to an L1) might diminish metacognitive processing and monitoring. Based on previous studies, people might not detect the semantic-relational manipulation and might judge both types of word lists –semantically related and unrelated– similarly (Opitz & Degner, 2012; Zhang et al., 2020). This will suggest that the language cue is more salient, potentially overshadowing semantic sources of information for learning monitoring. As to between language comparisons, based on results from the previous two experiments, we expect people to use language as a diagnostic cue (Koriat, 1997) and to give lower JOLs to L2 blocks, which might make people unfold control processes and show a recognition memory advantage for L2 materials (Francis & Strobach, 2013).

Methods

Participants

Forty-two psychology students from the University of Granada (45.2%) and Universidad Loyola Andalucía (54.8%) participated in this experiment. They were recruited and selected following the same procedure and criteria as in Experiments 1 and 2. We removed a participant who recorded the default JOL value of 50% for all items, suggesting he/she did not correctly perform the main task. This resulted in a final sample of 41 (18–29 years old, $M = 20.54$, $SD = 2.41$). Participants were tested individually in two remote sessions and received course credit as compensation. Comparisons for all self-reported measures and for the verbal fluency test showed that participants were unbalanced and significantly more fluent in L1 than in L2. All p values were below .05. See Table 6. for descriptive statistics.

Table 6.

Participant information: Mean score (and standard deviations) for language measure.

Self-reported measures	L1	L2
Daily exposure L1 (%)	81.3 (18.0)	33.2 (24.6)
Age of acquisition (in years)	2.8 (1.3)	7.0 (2.0)
Age of becoming fluent (in years)	5.3 (2.5)	12.6 (2.7)
Speaking self-competence (0-10)	9.7 (0.7)	7.7 (1.2)
Reading self-competence (0-10)	9.8 (0.4)	8.4 (1.1)
Exposure to reading (0-10)	9.1 (1.3)	7.0 (2.4)
Learning by reading (0-10)	9.5 (1.0)	8.1 (2.0)
Language proficiency		
MELICET (0-50 points)		36.2 (4.6)
Verbal fluency L1	24.7 (6.6)	16.3 (4.7)

Note: Verbal fluency task shows the mean number of words elicited in each language condition.

Materials and Procedure

The experiment consisted of two online sessions, each of which lasted approximately 60 minutes. The procedure was similar to that of Experiment 2, with the same cognitive and linguistics tasks programmed and administered with the same experiment builder (Anwyl-Irvine et al., 2020). However, the procedure for the memory task differed in the materials used and the moment when JOLs were solicited. In this case, participants studied 10 short lists of six words for a later recognition test and gave a JOL after the study phase for each list (see testing materials at https://osf.io/t2jg4). We used an adapted procedure, modeled after Matvey et al. (2006). Lists comprised either words grouped into a semantic category (e.g., musical instruments: horn, bass, drum, keyboard, harp, saxophone) or unrelated words (e.g., hole, blind, tower, kingdom, wheel, bishop). Nevertheless, unlike Matvey et al. (2006), participants in our study gave JOLs after studying each list instead of giving JOLs after each target word. Note that the relatedness manipulation affects the complete list (related word lists vs unrelated-word lists), differently from experiments 1 and 2 were the manipulation affected specific words within the list (e.g., concrete words vs. abstract words), and therefore, in this case, we could assess the difficulty of the list as a whole.

Within each list, words were presented randomly, one at a time, in the middle of the computer screen. Within each session, participants completed the JOL and recognition task in the two language blocks (L1 vs. L2). Similar to previous experiments, the order of the language blocks was counterbalanced across participants.

Semantic categories were selected from Van Overschelde et al. (2004) for English words and Marful et al. (2015) for Spanish words. We excluded English and Spanish cognates and filtered categories with less than six exemplars. Unrelated and new words were randomly selected from Brysbaert et al. (2014). Studied semantic-category words and unrelated studied and new words within and between languages were matched for estimated frequency and number of letters. S5 Table in the Supporting Information shows frequency and number of letters of each study list.

Materials for the study phase consisted of 20 lists of six words. For half of the lists, words belonged to the same semantic categories, whereas for the other half, the lists were composed of unrelated words. We randomly assigned five semantic-category lists and five unrelated lists to the Spanish–L1 and English–L2 block. List assignments were counterbalanced across participants, and participants were randomly assigned to one of the counterbalanced conditions. The lists within a counterbalanced condition were pseudo-randomly presented with the restriction that no more than two consecutive lists belonged to the related or unrelated condition. For the recognition task, participants were presented with all studied words and 60 unrelated new words, for a total of 120 words. Note that new words in the recognition task were always unrelated because given the restriction in the selection procedure, there were not enough categories to be used as new-distractor words. However, this was true for the two language conditions, and therefore the critical between-language comparison was fully controlled. Words appeared randomly one by one in the center of the screen, regardless of the condition (words grouped into semantic related categories and unrelated words).

Results

As in Experiments 1 and 2, we introduced the JOLs from the study phase and d-prime for recognition into ANOVAs with language (Spanish–L1 vs. English–L2) and relatedness (related list vs. unrelated list) as within-subject factors and block order (L1-first vs. L2-first) as a between-subject factor. As in Experiment 1 and 2, we conducted the analyses but also including MELICET scores as a covariate (ANCOVAs), which yielded identical results. Scores in the MELICET did not interact with language $F(1, 38) = 3.30$, $p = .007$, $\eta_p^2 = .080$, or type of list $F(1, 38) = 0.001$, $p = .972$, $\eta_p^2 = .000$ when considering both the JOLs and recognition test – language $F(1, 31) = 0.42$, $p = .520$, $\eta_p^2 = .013$, and type of list $F(1, 31) = 0.001$, $p = .970$, $\eta_p^2 = .000$.

Following the same exclusion criteria as in the previous experiments, we removed one participant from analyses.

Study phase (JOLs). The results of the ANOVA yielded a significant main effect of language ($F(1,39) = 9.29$, $p = .004$, $\eta_p^2 = .192$), with L1 lists ($M = 64.4$, $SE = 2.5$) receiving higher JOLs than L2 lists ($M = 59.3$, $SE = 2.1$). We also observed a main effect of relatedness ($F(1,39) = 84.10$, $p < .001$, $\eta_p^2 = .683$), such that related lists ($M = 71.6$, $SE = 2.4$) received higher JOLs than unrelated lists ($M = 52.1$, $SE = 2.4$). The main effect of block order was significant ($F(1,39) = 5.41$, $p = .025$, $\eta_p^2 = .122$). JOLs tended to be higher when the L2 block was placed first ($M = 66.9$, $SE = 3.12$) compared to JOLs when L1 was first ($M = 56.8$, $SE = 3.1$). There were no significant interactions ($p > .1$ for all).

Recognition test (accuracy). In this task, we excluded two participants for having a fast response (<300ms) in more than 10% of trials (Roessel et al., 2018) and five participants with d-prime below 0.5 (Macmillan & Kaplan, 1985), resulting in a sample of 34 participants. We filtered out 0.16% trials with reaction times shorter than 300 ms. S3 Table in the Supporting Information shows estimated means (and standard deviations) for hits, false alarms, misses and correct rejections by language, type of list and block order.

We found a language effect ($F(1, 32) = 5.51$, $p = .025$, $\eta_p^2 = .147$), with L2 lists ($M = 2.4$, $SE = 0.15$) being recalled better than L1 lists ($M = 2.2$, $SE = 0.13$). This effect was mediated by a significant interaction effect of language and block order ($F(1, 32) = 7.29$, $p = .011$, $\eta_p^2 = .186$). Thus, participants recognized L2 words ($M = 2.7$, $SE = 0.21$) better than L1 words ($M = 2.1$, $SE = 0.18$) only when the L2 block was placed first. We also observed a type-of-list effect ($F(1, 32) = 35.56$, $p < .001$, $\eta_p^2 = .526$). Words grouped into semantic categories ($M = 2.5$, $SE = 0.13$) were better recognized than unrelated words ($M = 1.9$, $SE = 0.13$) regardless of language or block order. There were no other significant main effects of block order or interactions ($p > .1$ for all). See Table 7 for estimated means and standard deviations for JOLs and d-prime.

Table 7.

JOLs and d-prime across conditions.

		Semantic category		Unrelated words	
Language	Block order	JOL	d-prime	JOL	d-prime
L1	L1-first	69.9 (4.0)	2.3 (0.2)	50.0 (4.0)	1.8 (0.2)
	L2-first	77.4 (4.1)	2.4 (0.2)	60.2 (4.1)	1.8 (0.2)
L2	L1-first	63.1 (3.2)	2.3 (0.2)	44.1 (3.4)	1.7 (0.2)
	L2-first	75.9 (3.3)	2.9 (0.2)	54.2 (3.5)	2.6 (0.2)

Goodman–Kruskal gamma correlation. In order to calculate the Goodman–Kruskal gamma value for each subject and condition, we correlated the JOLs with the proportion of words correctly recognized in each list. The mixed-factor ANOVAs revealed no significant difference across conditions. There was no significant main effect of block order ($F(1, 25) = 2.48$, $p = .128$, $\eta_p^2 = 0.09$) (L1-first: $M = -0.2$, $SE = 0.1$; L2-first: $M = 0.0$, $SE = 0.1$), language ($F(1, 25) = 0.95$, $p = .339$, $\eta_p^2 = 0.037$) (L1: $M = -0.2$, $SE = 0.1$; L2: $M = -0.1$, $SE = 0.1$), or relatedness ($F(1, 25) = 0.02$, $p = .903$, $\eta_p^2 = 0.001$) (words grouped into semantic categories: $M = -0.1$, $SE = 0.1$; unrelated words: $M = -0.1$, $SE = 0.1$). None of the interactions were significant ($p > .1$ for all).

With regard to the learning strategies used in the study phase, in this experiment participants reported words rehearsal (76%), grouping words by their semantic meaning (68.3%), and creating mental images (56.1%) as the strategies most used in the study phase.

Discussion

We successfully replicated the relatedness effect previously reported in the literature among monolingual individuals, in both JOLs and memory performance (Janes et al., 2018; Mueller et al., 2013; Undorf & Erdfelder, 2015). In our study, participants consistently assigned higher JOLs and demonstrated better recognition for semantically related word lists compared to unrelated word lists, in L1. These findings are consistent with prior research indicating that greater semantic relatedness among items tends to make them more perceivably learnable and memorable, and indeed usually result in improved retrieval (Janes et al., 2018; Mueller et al., 2013; Undorf & Erdfelder, 2015).

More interestingly, this was also true for L2-word lists despite the fact that JOLs were lower for L2 than for L1 lists. This suggests that the processes operating when processing lists of words with different semantic relatedness in L2 mirror those in L1. Bilingual individuals were able to monitor the study materials in L2 as well as they did in L1. They adjusted their JOLs to the objective semantic difficulty of the materials and showed, in fact, an advantage in memory for the easier condition (semantically related word lists). This finding suggests that participants correctly monitored the materials in L2 and identified unrelated words as more challenging. This pattern does not rule out the possibility of differential brain regions engagement or different underlying mechanisms that we might not capture with behavioral measures (see Opitz & Degner, 2012, for a delay in lexical access during L2 processing; and Zhang et al., 2020, for a weaker activation of the semantic integration hub, indicating a different engagement of brain regions during L2 processing). Indeed, other studies have shown that bilinguals demonstrated similar behavioral patterns but displayed different brain activation while performing cognitive tasks (e.g., prospective memory; López-Rojas et al., 2022).

Regarding the language cue, people perceived studying in L1 easier than studying in L2 (i.e., lower JOLs values for L2 blocks). Nevertheless, this pattern did not align with the actual results of the recognition test. Surprisingly, participants exhibited better recognition of L2 words, especially when the L2 block was placed first. This finding mirrors the L2 recognition advantage observed in previous studies (Francis & Gutiérrez, 2012b; Francis & Strobach, 2013) and also in Experiment 2. This advantage in bilinguals' recognition memory for L2 has been attributed to various factors. For instance, Francis & Strobach (2013) suggested that unbalanced bilinguals have had fewer exposures to L2 words over their lifetime compared to L1 words, which functionally exhibit lower frequency, supporting the observed bilingual L2 advantage in recognition over both bilingual and monolingual L1 performance. In their study, Francis & Strobach (2013) examined two participant groups: Spanish-English bilinguals and monolingual individuals. Both groups studied high- and low-frequency words in each language. Both monolinguals and bilinguals showed superior recognition of low-frequency words compared to high-frequency ones. Notably, bilinguals demonstrated better recognition rates in L2 compared to L1. Furthermore, bilinguals' L2 recognition exceeded that of monolinguals. This advantage in bilinguals' L2 recognition mirrors word frequency effects in recognition and is attributed to the heightened distinctiveness of L2 words in memory relative to L1 words. Consequently, due to their novelty, lower familiarity, or less frequent use, L2 words might capture increased attention during encoding, resulting in better recognition compared to more familiar L1 words.

The discrepancy between lower JOLs assigned to L2 words and the observed better recognition rates suggests an inconsistency in terms of the language cue's resolution, as indicated by the Goodman-Kruskal correlations. People thought they had not learnt L2 materials as good as L1 materials, yet that was not the case. The perceived difficulty might have triggered some kind of deeper processing to compensate and achieve successful learning. It is plausible that control regulatory mechanisms are engaged when performing the task in L2, even though the monitoring process did not clearly identify any potential learning deficits (L2 blocks received lower JOLs only when they were placed first). In fact, this results might support Koriat et al. (2006) suggestion that the link between monitoring and control processes originates from metacognitive judgments resulting from control processes during learning. Thus, monitoring does not necessarily precede controlled actions, but they might occur after control has been exerted. In line with this idea, item difficulty is dynamically monitored: learners assign suitable cognitive resources to an item depending on its demands, recognizing an item's challenge in memory when they perceive a greater effort is needed to memorize it. In our experiment, participants might have allocated more attention to L2 words, assuming they would be less likely to be remembered, inadvertently leading to a more effective learning of the materials. If this mechanism were at play, the challenges posed by the L2 could be viewed as a desirable difficulty that enhances the learning process for simple sets of materials (Bjork & Bjork, 2020). In our next experiment, we explore this hypothesis by utilizing more complex materials, specifically expository texts, which aligns with those commonly encountered in academic and professional contexts.

Experiment 4:

High vs. low cohesion texts

Text cohesion refers to linguistic cues that help readers to make connections between the presented ideas. Examples of cohesion cues include the overlap of words and concepts between sentences and the presence of discourse markers such as *because*, *therefore*, and *consequently* (Crossley et al., 2016; Halliday & Hasan, 1976). Poor cohesion texts impose higher demands on readers who would need to produce inferences in order to create a meaningful representation of the information in the text (Best et al., 2005), and these texts are associated with poorer comprehension (Crossley et al., 2014, 2016; Hall et al., 2016).

Importantly, text cohesion has been shown to influence JOLs magnitude. Some studies show that participants are able to monitor text cohesion and adjust their JOLs accordingly (Carroll & Korukina, 1999; Lefèvre & Lories, 2004; Rawson & Dunlosky, 2002). For example, Lefèvre & Lories (2004) manipulated cohesion by introducing or omitting a repetition of the antecedent to vary ambiguity in anaphoric processing. That is, they modified the complexity of resolving references to previously mentioned entities in the text. They observed that participants provided lower immediate JOLs for low than for high cohesion paragraphs. In addition, they found significant correlations between JOLs and comprehension scores. These results suggest that metacognitive monitoring is sensitive to the cohesion features of a text, as people reported that they were poorly learning the low-cohesion texts that, indeed, were comprehended worse than high-cohesion texts. Similarly, Rawson & Dunlosky (2002) varied coherence by manipulating causal relatedness across sentence pairs and by altering the structure of sentences within paragraphs. They also found that both predictions and memory performance were significantly lower for low coherence pairs than for moderate-high coherence pairs. Finally, Carroll & Korukina (1999) manipulated the sentence order in narrative texts to create different coherence versions. They found a significant main effect of text coherence on both judgments and memory, as the ratings and the proportion of items that were immediately recalled were significantly greater for ordered texts than for non-ordered texts.

In Experiment 4, we intended to extend our previous results regarding learning monitoring in L2, with more complex materials such as expository texts. These materials are closer in nature to those encountered in academic and professional settings. We manipulated cohesion to vary texts difficulty and assessed participants' monitoring of the difficulty of the presented texts (i.e., judgments of learning), their actual learning (i.e., open-ended questions) and the learning strategies used (i.e., self-report questionnaire). The evaluation of learning strategies and self-regulation has historically emphasized methods like self-report questionnaires. A combination of different methodologies (e.g., prospective judgments –JOLs– and self-report questionnaires) becomes crucial to attain a deeper understanding. Integrating diverse methods will enriches our understanding will offer a more comprehensive perspective on regulation and metacognition in learning contexts.

In line with our prior studies, our main hypothesis was that participants might adjust their overall perception of learning according to the linguistic context. We expected that participants would use language as a diagnostic cue (Koriat, 1997) and assessed learning in L1 as easier and more successful compared to learning in L2, and therefore, they would provide higher JOLs in L1 than in the L2 context. Several studies have consistently argued that the success of L2 readers' comprehension largely depends on those readers' L2 reading proficiency (e.g., Horiba, 2000; Hosoda, 2014, 2015; Yoshida, 2003). As our participants are unbalance L2 English speakers, we predict that reading in L2 might impose greater challenges compared to reading in L1.

On the other hand, studying from texts in L2 may compromise the correct functioning of the processes implicated in self-regulated learning. Thus, we expected to observe less accurate assessment of other cues that influence the material's difficulty (i.e., texts cohesion) in L2 compared to L1. The challenges associated with studying from texts in L2 could potentially be so demanding that individuals might struggle to accurately detect text cohesion as a valuable cue for assessing their learning in that language. For example, Lefèvre & Lories (2004) demonstrated that a disruptive task conducted immediately before judging the cohesion of a critical paragraph whose cohesion had been manipulated, eliminated the impact of cohesion. Moreover, it has been suggested that working memory capacity (WMC) had indirect impacts on L2 reading comprehension primarily with greater WMC leading to better L2 reading (and inferential) comprehension (e.g., Alptekin & Erçetin, 2010, 2011; Kim, 2023). Specially, WMC plays a critical role when dealing with complex tasks. For example, Jung (2018) investigated whether cognitive task complexity affects L2 reading comprehension and whether WMC moderates the influence of task complexity. For this, bilingual participants were randomly assigned to either the simple or complex condition and read two short passages while answering reading comprehension questions. Simple versions included coherent texts, whereas complex versions contained texts whose paragraphs were disarranged and required participants to order them coherently. They found that under the complex condition, participants benefited from higher WMC when answering reading comprehension questions. Thus, in our case, the complexity of L2 processing might overshadow the proper assessment of text cohesion and might make it harder to use both cues (language and cohesion) as diagnostic information for learning.

Methods

Participants

We conducted a power analysis using G*Power (Faul et al., 2007) to determine the required sample size. We calculated it considering a mixed factor analysis of variance (ANOVA) with language and cohesion as repeated measure variables, and order of the language block as a between-participant variable. We estimated 30 participants, assuming a small to moderate effect size (partial eta-squared of 0.07) to observe significant ($\alpha = 0.05$) effects at 0.8 power. Due to an error in the texts counterbalancing procedure, we had to recruit more participants to ensure a representative sample in each of the counterbalance lists. Participants had a normal or corrected-to-normal vision and reported no neurological damage or other health problem. Participants gave informed consent before performing the experiment that was carried out following the Declaration of Helsinki (World Medical Association 2013). The protocol was approved by the institutional Ethical Committee of the University of Granada (857/CEIH/2019) and the Universidad Loyola Andalucía (201222 CE20371).

Sixty-eight psychology students from Universidad Loyola Andalucía (51.5%) and the University of Granada (48.5%) participated in this experiment. We removed from all analyses (1) a participant who did not vary the percentage given as a JOL in any of the texts and left it at the default value, and (2) two participants who gave answers in Spanish for both L1 and L2 block. We therefore had a total sample of 65 (18–28 years old, M = 19.92, SD = 1.76). Participants were tested remotely and individually in a two-session experiment and received course credit as compensation.

Participants were non-balanced Spanish-English bilinguals although they started acquiring English as their L2 during childhood (M = 6.75, SD = 2.75). Subjective (Language Background Questionnaire, LEAP-Q; Marian, Blumenfeld, & Kaushanskaya, 2007) and objective (MELICET Adapted Test, Michigan English Language Institute College Entrance Test) language measures indicated that the sample had an intermediate proficiency level in English (M = 28.79). See Table 8. for descriptive statistics.

Table 8.

Participants' information for demographic and language measures.

	Mean (SD)	
Age	19.72 (1.76)	
O-Span	0.62 (0.22)	
Self-reported measures	L1	L2
Daily exposure (%)	83.03 (15.02)	37.55 (21.00)
Age of acquisition (in years)	2.88 (1.13)	6.75 (2.76)
Age of becoming fluent (in years)	5.71 (2.29)	11.60 (3.43)
Speaking self-competence (0-10)	9.69 (0.68)	7.44 (1.46)
Reading self-competence (0-10)	9.51 (0.97)	8.16 (1.22)
Exposure to reading (0-10)	9.00 (1.29)	6.23 (2.54)
Learning by reading (0-10)	9.45 (1.06)	7.73 (2.03)
Language proficiency		
MELICET (0-50 points)		28.79 (8.47)

Note: T-tests for paired sample showed significant differences between languages in all the measures (all *p*s < .001).

Materials and Procedure

The experiment consisted of two online sessions that lasted 120 and 90 minutes respectively. We programmed tasks, presented stimuli and collected data with Gorilla Experiment Builder, (Anwyl-Irvine et al., 2020). Participants accessed the experiment remotely and individually. In order to ensure that participants did not open other windows in the computer while doing the tasks, they were forced to full-screen presentations. Recent research supports the validity and precision of experiments run online (Anwyl-Irvine et al., 2020, 2021; Gagné & Franzen, 2023).

The main task in both sessions was a *learn-judge-remember* task with a study phase and a recognition test. It simulated a learning task in a classroom environment in which students needed to learn and remember information from texts either in Spanish-L1 or in English-L2, depending on the session. Additionally, participants completed the MELICET Adapted Test (Michigan English Language Institute College Entrance Test) as an objective L2 proficiency measure.

Furthermore, at the end of the second session they fulfilled a customized metacognitive questionnaire regarding the strategies used when studying the texts in both languages (see selected items at https://osf.io/83st7), a language background and sociodemographic questionnaire (LEAP-Q, Marian et al., 2007), and the Spanish version of the Operational Digit Span task (O-Span) to assess that all participants ranged within normal standardized values of working memory capacity (Turner & Engle, 1989). We used a shortened version adapted from Oswald et al. (2014) in which participants are presented with series of math problems followed by a to-be-remembered target letter. We calculated a working memory index by multiplying the mean proportion of successfully recalled letters and the mean proportion of correctly solved arithmetic equations (Conway et al., 2005).

In the *learn-judge-remember* task**,** participants were instructed to give a *judgment of learning* (JOL), and to answer some questions about the text they read. We manipulated the language (Spanish-L1 vs. English-L2) and the cohesion of the texts (high- vs. low-cohesion) as within-subjects factors. Language was a blocked variable and the assignment of L1 or L2 to the first or second session was counterbalanced across participants. Both high- and low-cohesion texts appeared along the study phase for each language block so that half of the texts within a block were high-cohesion and the other half was low-cohesion.

In each session, participants were informed to read comprehensively ten short texts for a later learning assessment test. Texts of high- and low- cohesion were presented in a pseudo-random order one at a time in the middle of the computer screen and remained 3 minutes for self-reading. Immediately after the presentation of each text, participants gave a JOL to predict the likelihood of remembering the information they have just read on a 0–100 scale (0- *not all likely,* 100- *very likely)* by moving a handle slider to the desired number. This screen advanced when participants pressed ENTER.

After studying and judging each text, participants answered three open-ended questions as an objective measure of their learning. Previous research exploring the consequences of studying in L1 vs. L2 on memory found different effects depending on the type of test. For example, Vander Beken et al. (2020) and Vander Beken & Brysbaert (2018) found that essay questions hindered performance in L2 presumably due to difficulties in writing production while no differences between L1 and L2 performance were found with open-ended questions and true/false recognition items. This suggests that language proficiency and background would make the writing process more complex and challenging in L2 than in L1. In order to avoid confounding effects with writing complexity, we discarded essay and chose open-ended format to better discriminate and prevent a possible ceiling effect that may appear with True/False recognition items. Questions covered a range of information from general ideas to examples or brief descriptions. Participants could respond with a single word, a noun phrase, or a concise sentence (e.g., "What type of teeth do meat-eating mammals have?"; see S6 Table in the Supporting Information for a detailed example). Open-ended questions were corrected automatically using a Python script that matched a rubric criterion developed a priori (available at https://osf.io/w3dc5). This script has been checked against a manual revision leading to a higher reliability in the final score/mark. We gave 1 point for fully correct answers in the language required in the question and 0 point for incomplete or incorrect answers. We provided full score if the key concepts in the rubric were included in the answer, accepting spelling mistakes in both languages. We calculated the mean proportion of correct recall for each participant and condition.

We selected twenty-two English texts from different books and previous studies: two of them were used as examples and the rest were testing materials (see S7 Table in the Supporting Information for references). We translated them into Spanish and manipulated their cohesion following norms from previous studies (see Table 9). This resulted in four different versions for every text, one per language-cohesion condition: L1 high- and L1 low-cohesion and L2 high- and L2 low-cohesion. Materials are available at https://osf.io/vfykg.

Table 9.

Norms for manipulating text cohesion.

High-cohesion texts	References
Including connectors.	(Gasparinatou & Grigoriadou, 2013; Ozuru et al., 2012)
Increasing noun overlap.	(Hall et al., 2016)
Shortening sentences.	(Soemer & Schiefele, 2019)
Low-cohesion texts	
Lengthening sentences.	(Soemer & Schiefele, 2019)
Including subordinate phrases.	
Replacing nouns with pronouns to create ambiguity.	(Hall et al., 2016; Soemer & Schiefele, 2019)
Using low-frequency synonyms for key concepts.	(Hall et al., 2016; Soemer & Schiefele, 2019)
Using the passive voice.	
Having abrupt gaps between sentences.	(Soemer & Schiefele, 2019)

We then created a total of four counterbalanced texts lists where only one version of the text was included. Each of these lists contained five texts per language-cohesion condition. Therefore, each participant was presented with 20 different texts: 10 in the L1 session and 10 in the L2 session, of which five were low- and five high-cohesion texts. Repeated measures ANOVA (cohesion and language) showed that the texts were matched in length (number of words) between condition, as no main effects or interaction were significant [all p values were above 0.05; L1: high- (M = 142.6, SD = 22.8) and low- cohesion (M = 141.0, SD = 28.4); L2: high- (M = 141.0, SD = 26.4) and low- cohesion (M = 140.3, SE = 29.6)]. See S6 Table in the Supporting Information for an example of a high- and low- cohesion version of a text in L2 and its associated open-ended questions.

For the *learn-judge-remember task*, we analyzed JOL responses in the study phase and the proportion of correct answers for the open-ended questions, grouped by condition (language and text cohesion).

In each language block, after the *learn-judge-remember* task, participants answered a customized metacognitive self-report questionnaire. We combined items from two different inventories into a single set of 8 questions. We selected items from the NASA Task Load Index (NASA-TLX; Hart & Staveland, 1988) and the Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich et al., 1991). Participants rated themselves on a seven-point Likert scale from "not at all true of me" to "very true of me”.

With this questionnaire, we assessed cognitive and metacognitive learning strategies, effort regulation, mental demand, and self-perceived performance. Originally, some items referred to a general learning setting, so we adapted some of the expressions to the task. We also translated the items into Spanish since we administered the questionnaire in the language the session was taking place. The item referring to metacognitive self-regulation was reversed and thus we inverted their punctuation. For this questionnaire, we compared the score for items in L1 and L2. See **¡Error! No se encuentra el origen de la referencia.** to check the set of questions included into the questionnaire.

Results

We performed a 2 x 2 x 2 (language x text cohesion x block order) mixed-factor ANOVAs for JOLs in the study phase and for learning assessment test. We included language order in the analyses, since previous research (Reyes et al., 2023) suggests that it might influence performance. Language (L1 vs. L2) and text cohesion (high vs. low-cohesion) were within-subject factors and block order (L1-first vs. L2-first) a between-subject factor. For all analyses, the alpha level was set to 0.05 and we corrected by Bonferroni for multiple comparisons. All effect sizes are reported in terms of partial-eta-squared (η_p^2) for ANOVAs and Cohen's *d* for t- tests.

Study phase (JOLs). To evaluate the effect of language and text cohesion on the magnitude of JOLs, we computed the mean across participants' JOLs for each condition (see Table 10 for partial means).

We found significant main effects of language, $F(1, 63) = 20.13$, $p <.001$, $\eta_p^2 = .24$, and cohesion, $F(1,63) = 15.35$, $p <.001$, $\eta_p^2 = .20$. Texts in L1 ($M = 68.8$, $SE = 1.58$) received higher JOLs than texts in L2 ($M = 61.6$, $SE = 1.95$). Similarly, high-cohesion texts received higher JOLs ($M = 66.5$, $SE = 1.60$) than low-cohesion texts ($M = 63.8$, $SE = 1.64$). The main effect of block order was not significant, $F(1,63) = 0.06$, $p =.808$, $\eta_p^2 = .001$.Texts received comparable JOLs regardless of the language order (L1 first: $M = 64.8$, $SE = 2.22$; L2 first: $M = 65.5$, $SE = 2.26$). None of the interactions were significant (all *p*s > .05).

Learning assessment test (accuracy). To evaluate learning performance, that is, how much participants remembered from the texts, we computed the proportion of correct responses in the open-ended questions across participants (see Table 10).

Again, the analysis showed a significant main effect of language, $F(1, 63) = 6.41$, $p = .014$, $\eta_p^2 = .10$, and cohesion, $F(1, 63) = 24.66$, $p < .001$, $\eta_p^2 = .28$. That is, people remembered information from texts in L1 ($M = 0.71$, $SE = 0.02$) better than from texts in L2 ($M = 0.65$, $SE = 0.02$); and from high-cohesion ($M = 0.71$, $SE = 0.02$) better than from low-cohesion ($M = 0.65$, $SE = 0.02$) texts. The main effect of block order did not reach significance, $F(1, 63) = 1.62$, $p = .21$, $\eta_p^2 = .03$. People's accuracy in the learning assessment tests did not depend on which language block they performed on the first or second place (L1 first: $M = 0.65$, $SE = 0.03$; L2 first: $M = 0.70$, $SE = 0.03$). Similarly, none of the interactions were significant (all *p*s > .05).

Table 10.

Mean (and standard deviations) for JOLs scores (1-100 scale) and learning performance (proportion of correct responses) across language, cohesion and block order conditions.

		High-cohesion		Low-cohesion	
Language	Block order	JOL	Learning	JOL	Learning
L1	L1-first	69.4 (2.27)	0.74 (0.03)	66.2 (2.32)	0.65 (0.03)
	L2-first	70.9 (2.31)	0.75 (0.03)	68.6 (2.35)	0.70 (0.03)
L2	L1-first	63.0 (2.77)	0.65 (0.04)	60.5 (2.96)	0.58 (0.03)
	L2-first	62.8 (2.81)	0.71 (0.04)	59.9 (3.01)	0.65 (0.03)

Language metamemory accuracy –resolution–. In order to examine participants' metamemory accuracy –resolution– across languages (i.e., to check whether participants' JOLs discriminate between the information recall of one text relative to another), we used a Goodman–Kruskal (GK) gamma correlation (Nelson, 1984) and a language-accuracy index correlation. We calculated one gamma correlation for each participant in each of the language conditions. We then ran a *t-test* to examine whether the GK Gamma correlations differed across languages. No significant effects were found, $t(64) = -0.76$, $p = 0.45$, $d = -0.09$.

We also performed a language-accuracy index correlation as an additional measure of metamemory accuracy, which allows to further explore participant's overall resolution in L1 and L2. To do so, we first calculated a language index for JOLs and for learning accuracy, by subtracting the mean scores in L2 from the mean scores of L1 of JOLs and learning accuracy respectively, and then performed correlations analyses of the two indexes. Interestingly, JOLs index correlated with accuracy index ($r = 0.6$), suggesting that people's predictions during the study phase about what they would remember later agreed with what they actually recalled in the learning assessment test.

Customized metacognitive self-report questionnaire. We analyzed the questionnaire regarding the participants' learning strategies in the study phase (see Table 11). We conducted *t-tests* to compare the frequency of each strategy in L1 and L2. Overall, we found that people employed some strategies more frequently in L1 than in L2. Namely, elaboration ("When reading the texts, I tried to relate the material to what I already knew"), $t(64) = 2.17$, $p = .033$, $d = 0.27$, metacognitive self-regulation ("When studying the materials in the texts, I often missed important points because I was thinking of other things"), $t(64) = 2.71$, $p = .008$, $d = 0.34$, and effort regulation ("I worked hard to do well even if I didn't like what I was studying in the texts"), $t(64) = 2.06$, $p = .044$, $d = 0.26$. No differences were found in terms of critical thinking strategy, $t(64) = -0.60$, $p = .55$, $d = -0.07$, or rehearsal, $t(64) = -0.59$, $p = .56$, $d = -0.07$. No strategy was more frequently used in L2 than in L1 either. As expected, and consistent with JOLs, participants reported that they experienced significantly higher mental demand in L2 than in L1, $t(64) = -8.81$, $p < .001$, $d = -1.09$. Similarly, people felt their performance had been better in L1 than in L2, $t(64) = 2.17$, $p = .034$, $d = 0.27$.

Table 11.

Means (and standard deviations) of the different dimensions of the metacognitive self-report questionnaire regarding learning strategies during the study phase by language block.

Strategies	L1	L2
Rehearsal	5.72 (1.40)	5.83 (1.25)
Elaboration	5.86 (1.08)	5.58 (1.19)
Critical Thinking	4.43 (1.51)	4.33 (1.25)
Metacognitive self-regulation	5.19 (0.92)	4.92 (0.88)
Effort regulation	5.57 (0.93)	5.29 (0.94)
Self-perceived…		
Mental demand	3.77 (1.50)	5.34 (1.19)
Performance	5.09 (1.20)	4.69 (1.15)
Effort	4.60 (1.47)	4.86 (1.45)

Note: Participants self-rated the frequency they used each of the strategies in a 7-point Likert scale.

Discussion

In Experiment 4, the focus was twofold. Initially, we wanted to observe if a cue such as text cohesion yielded variations in JOLs and learning. Previous studies had reported that low cohesion texts are associated with poorer comprehension (Crossley et al., 2014, 2016; Hall et al., 2016); and that participants are able to monitor text cohesion and adjust their JOLs accordingly (Carroll & Korukina, 1999; Lefèvre & Lories, 2004; Rawson & Dunlosky, 2002). We also wanted to examine whether these differences depended on the language environment (L1 or L2) within which the task was executed. Additionally, we aimed to evaluate whether that linguistic context exerted an influence on the global perception of task complexity.

Regarding text cohesion, we found a cohesion effect both in JOLs and in the learning assessment test. Participants predicted better performance for high-cohesion texts and, correspondingly, learning rates were higher for them compared to low-cohesion texts. Thus, we replicated what had previously been reported in the monolingual text monitoring and comprehension literature (Carroll & Korukina, 1999; Crossley et al., 2014, 2016; Hall et al., 2016; Lefèvre & Lories, 2004; Rawson & Dunlosky, 2002). More importantly, we expanded these findings to a bilingual sample, suggesting that the processes operating in L2 are similar to those in L1 concerning text monitoring and comprehension.

As to the language effect, people judged L1 materials as easier to learn –giving higher JOLs– than materials in L2. This language effect in JOLs is consistent with the results of our previous studies involving single words and short lists (Experiments 1-3). This result seems to suggest that people use the language of study as a diagnostic cue (Koriat, 1997) and adjust their JOLs according to the difficulty perceived, no matter of the nature of the materials. Note that, unlike our previous experiments, participants encountered significantly more difficulty in remembering information from texts presented in L2, as evidenced by the learning assessment test. In this case, participants could not overcome the challenge impose by L2 to achieve successful learning. Previous research exploring the consequences of studying in L1 vs. L2 on memory observed different effects depending on the test type used. For example, Vander Beken et al. (2020) and Vander Beken & Brysbaert (2018) found that essay questions hindered performance in L2 presumably due to difficulties in writing production while no differences between L1 and L2 performance were found with open-ended questions and true/false recognition items. Nevertheless, our participants did show an L2 recall cost despite the fact that we chose an open-ended format in order to avoid confounding effects with writing complexity, and that our rubric was not very strict and accepted grammatical, syntactic, or orthographic errors (note that language mistakes are not punished – as long as they do not obscure meaning– in international reading comprehension assessments like PISA). Hence other factors related to the type of processing or strategies used during L1 and L2 may have produced differences in L1 and L2 memory performance.

Overall, the pattern of results suggests that people were accurate in predicting their performance and were able to monitor their learning correctly both in L1 and L2. In fact, this is supported by the positive and significant correlation found between JOLs and memory accuracy. Despite participants judging texts in L2 as more challenging to learn, this perceived difficulty did not hinder their ability to employ monitoring processes effectively. They were still capable of identifying difficult materials and accurately assessing them as less likely to remember. Moreover, what they gave lower JOLs to (low-cohesion texts and L2 texts) was indeed worse comprehended, which suggest good metamemory accuracy. Therefore, L2 context did not impede their monitoring processes.

Nevertheless, results from the qualitative questionnaire indicated that participants employed distinct learning strategies when studying in L1 and L2. Overall, participants used deep-level strategies (e.g., elaboration, metacognitive self-regulation, and effort regulation) more frequently in L1. One possible explanation is that the selection of deep-level strategies might require extra study time, which was not possible under the time constrain of our experiment. Thus, people might have used deep-level strategies more frequently in L1 because they might have needed longer study time allocation for using them in L2 learning (Stoff & Eagle, 1971). What is evident, anyway, is that participants approach L2 learning with different strategies as opposed to L1, suggesting that the mechanisms underlying self-regulated learning in L2 might differ from those utilized in L1.

In sum, the results of Experiment 4 indicate that people with an intermediate level of English–L2 were able to use intrinsic cues such as text cohesion and language simultaneously, to monitor their learning. However, in previous experiments we did not manipulate participants' L2 proficiency, and it was plausible that differences in language proficiency and exposure could have influenced the monitoring behavior of the bilingual individuals. To explore this possibility, we conducted a last experiment to investigate the effects of L2 proficiency levels on monitoring and control processes of self-regulated learning. For this, we intentionally recruited participants with higher and lower L2 proficiency.

Experiment 5:

High vs. low cohesion texts in two proficiency groups

Recent research suggests that proficiency in L2 might introduce some nuances and might significantly impact the cognitive processes involved in learning from text. Several studies have explored the inferencing process of L2 readers, and they have consistently argued that the success of L2 readers' inference making largely depends on L2 reading proficiency (e.g., Horiba, 2000; Hosoda, 2014, 2015; Yoshida, 2003). The argument is that when L2 proficiency is limited, bilinguals have to dedicate more of their cognitive resources to fundamental reading processes, like interpreting individual words and phrases. Consequently, this heightened focus on basic reading tasks diminishes the cognitive resources available for more advanced discourse-level processing, including the generation of inferences. In a study by Hosoda (2014), bilingual university students, with varying levels of proficiency in L2 reading, were compared concerning their ability to generate causal inferences from explanatory texts. The findings indicated that readers with lower proficiency were notably less capable than those with higher proficiency in both on-line (during-reading) and off-line (task-induced) inference generation. Pérez et al. (2023) also present results on inferencing and revising processes involved in L1 and L2 text comprehension. They showed that L2 proficiency predicts better text comprehension and faster reading times in the comprehension question, suggesting a superior capacity to create anticipatory inferences. On the contrary, lower proficiency L2-English speakers showed longer reading times specially when the sentence was incongruent with the expectancies previously evoked in the text, suggesting a lower efficiency in the high-level cognitive processes during L2 processing. All this seems to suggest that higher proficiency in an L2 might lead to more efficient comprehension and processing of texts, allowing for a better use of monitoring and control strategies to optimize learning. On the contrary, lower proficiency in the L2 might introduce challenges to accurately monitor reading comprehension or needing more effort to control learning strategies and to compensate for linguistic limitations.

In Experiment 5, we investigated the influence of L2 proficiency on the on the dynamic relationship between monitoring and control during text-based learning. To achieve this, we recruited a sample that included both lower and higher English-L2 proficiency levels. We hypothesized that individuals in the lower-proficiency group might encounter challenges in effectively monitoring their learning process due to allocating a greater share of cognitive resources to language control compared to their higher-proficiency counterparts (Francis & Gutiérrez, 2012a; Hosoda, 2014; Pérez et al., 2023; Sandoval et al., 2010). As a result, the cohesion effect in JOLs, which manifests as higher JOL values for high-cohesion texts than for texts with low-cohesion, could potentially diminish within the lower-proficiency group in L2. This possible reduced effect might arise from the substantial cognitive load imposed by learning in a demanding L2 context, potentially overshadowing the sensitivity to detect nuanced differences in text cohesion. As previously mentioned, lower proficient L2 bilinguals showed less efficient monitoring and revising processes when making inferences in ambiguous sentences (Pérez et al., 2023). In essence, we posit that texts in L2 may present inherent challenges for individuals with lower-proficiency, regardless of their cohesion status.

In line with our prior studies, the language cue would influence JOls values. Our main hypothesis was that participants might adjust their overall perception of learning according to the linguistic context, and thus we expected that participants would use language as a diagnostic cue (Koriat, 1997). People with lower L2 proficiency level would assess learning in L1 as easier and more successful compared to learning in L2, and therefore, they would provide higher JOLs in L1 than in the L2 context. We might not observe this difference for the higher-proficiency group.

In addition, we introduced two further modifications to the procedure: (i) as block order was not significant in Experiment 4 during the study phase, in this experiment we presented high- and low-cohesion texts in L1 and L2 in a pseudorandom order; (ii) participants attended an in-person session at the laboratory for the second part of the experiment. The remaining conditions were held constant, mirroring the setup employed in Experiment 4.

Methods

Participants

Instructions for recruitment indicated that participants needed to have some English knowledge, although we did not specify the threshold for participation. Fifty-seven psychology students from the University of Granada (63.17%) and Universidad Loyola Andalucía (36.84%) enrolled in the experiment. Participants were tested individually in two sessions (a remote and an in-person session) and received course credit as compensation. We divided our sample into two independent groups by scores in MELICET. Based on previous studies (Kaan et al., 2020; López-Rojas et al., 2022), we established scores of 30 or above as the criteria to be included in the higher-proficiency group (n = 23, 18–23 years old, M =19.05, SD = 1.36), and scores of 25 or below were classified into the lower-proficiency group (n = 24, 18–49 years old, M =21, SD = 6.81). Participants with in-between scores were not included in the analyses. In addition, we removed two participants who did not vary the percentage given as a JOL in any of the texts and left it at the default value, so this resulted in a total sample of 49. No differences were found in the O-Span index (following the same calculation described in Experiment 1) between groups $t(45) = 1.49$, $p = .14$, $d = 0.43$ (higher-proficiency group: M = 0.60, SE = 0.04; lower-proficiency group: M = 0.51, SE = 0.05) suggesting that any possible difference between groups were not due to differences in working memory capacity. Comparisons between languages for all self-reported linguistic measures within groups showed that participants in both groups were unbalanced and significantly more fluent in L1 than in L2. All p values were below .05. See

Table 12 for further details.

Table 12.

Participants information for demographic and language measures divided by proficiency group (higher- and lower-proficiency).

	Higher-proficiency		Lower-proficiency	
Age	19.05 (1.36)		21.64 (6.80)	
O-Span index	0.60 (0.20)		0.50 (0.22)	
Self-reported measures	L1	L2	L1	L2
Daily exposure (%)	70.3 (16.50)	32.6 (17.00)	90.2 (9.52)	22.80 (22.70)
Age of acquisition (in years)	2.84 (1.45)	5.98 (2.18)	3.14 (1.92)	7.41 (2.54)
Age of becoming fluent (in years)	5.41 (2.10)	10.80 (3.89)	5.92 (2.91)	13.90 (7.56)
Speaking self-competence (0-10)	9.74 (0.70)	8.09 (1.16)	9.62 (1.42)	6.35 (1.77)
Reading self-competence (0-10)	9.70 (0.47)	8.41 (1.40)	9.58 (0.81)	7.23 (1.42)
Exposure to reading (0-10)	8.57 (1.47)	6.52 (2.78)	9.19 (1.41)	4.84 (2.62)
Learning by reading (0-10)	9.35 (1.03)	8.22 (2.35)	8.96 (1.51)	6.81 (2.32)
Second language proficiency				
MELICET (0-50 points)	-	38.70 (5.26)	-	17.04 (5.15)

Note: *Higher-proficiency* group (n = 23) scored 30 or more in MELICET (M = 38.7, SE = 1.1) while *lower-proficiency* group (n = 24) scored 25 or less (M = 17.7, SE = 0.85) and significant differences between groups in this measure was found $t(45) = 15.2$, $p < 0.01$, $d = 4.44$. (*) Significant differences between groups ($p < .05$).

Materials and Procedure

The experiment consisted of two sessions for which participants received course credit as compensation. We programmed and administered the tasks with the same experiment builder as in Experiment 4 (Anwyl-Irvine et al., 2020). The first session lasted 30 minutes and was administered remotely. Participants had to complete some questionnaires, the LEAP-Q (Marian et al., 2007) and an objective L2 proficiency measure (MELICET). Then participants came to the laboratory to complete the second session in person, which lasted 120 minutes. The procedure was similar to that of Experiment 4, with a *learn-judge-remember* task, a customized metacognitive questionnaire regarding the strategies used when studying the texts and the standard Operational Digit Span task (O-Span) to assess that all participants ranged within normal standardized values and that groups did not differ in working memory (Oswald et al., 2014).

Results

We report a 2 x 2 x 2 (language x text cohesion x proficiency group) mixed-factor ANOVAs for JOLs in the study phase and for accuracy in the learning assessment test. Language (Spanish–L1 vs. English–L2) and text cohesion (high vs. low) were within-subject factors, and proficiency group (higher vs. lower) was a between-subject factor. As in Experiment 4, the alpha level was set to 0.05 and we corrected by Bonferroni for multiple comparisons, in all analyses. All effect sizes are reported in terms of partial-eta-squared (η_p^2) for ANOVAs and Cohen's d for t- tests.

Table 13 shows partial means for JOLs and accuracy in the learning assessment test.

Study phase (JOLs). Analysis on the JOLs showed significant main effects of language, $F(1, 45) = 26.41$, $p < .001$, $\eta_p^2 = .37$, and cohesion, $F(1, 45) = 8.47$, $p = .006$, $\eta_p^2 = .16$. As expected, texts in L1 received higher JOLs ($M = 63.6$, $SE = 2.16$) than texts in L2 ($M = 55.4$, $SE = 2.37$). Similarly, texts with high-cohesion received higher JOLs ($M = 61.0$, $SE = 2.13$) than texts with low-cohesion ($M = 58.1$, $SE = 2.22$). The main effect of group did not reach significance, $F(1, 45) = 1.07$, $p = .31$, $\eta_p^2 = .02$. Overall, the higher-proficiency group ($M = 61.7$, $SE = 3.03$) gave similar JOLs values than the lower-proficiency group ($M = 57.3$, $SE = 2.97$).

The main effect of language was modulated by an interaction with proficiency group, $F(1, 45) = 23.48$, $p < .001$, $\eta_p^2 = .34$. Post-hoc comparisons showed that people in the higher-proficiency group did not differ in their JOLs between languages, $t(45) = -0.77$, $p = 1.00$ (L1: $M = 62.0$, $SE = 3.08$; L2: $M = 61.5$, $SE = 3.39$) whereas people in the lower-proficiency group gave significantly higher JOLs for texts in L1 ($M = 65.3$, $SE = 3.02$) than for texts in L2 ($M = 49.4$, $SE = 3.32$), $t(45) = 7.13$, $p < .001$. Overall, this trend suggests that proficiency was modulating JOLs values when studying in L2.

The interaction between language and cohesion was also significant, $F(1, 45) = 66.4$, $p < .001$, $\eta_p^2 = .60$. Post-hoc comparisons showed that JOLs for low-cohesion texts differed between languages, $t(45) = 8.31$, $p < .001$, while no significant difference between languages was found for JOLs in high-cohesion texts, $t(45) = 0.17$, $p = 1.00$. That is, participants gave higher JOLs in L1 for low-cohesion texts ($M = 66.1$, $SE = 2.2$) than in L2 ($M = 50.1$, $SE = 2.63$). This difference was not significant for high-cohesion texts, that received comparable JOLs in both languages (L1: M = 61.1, SE = 2.26; L2: M = 60.8, SE = 2.36). This suggests that the cohesion effect is significantly more salient in L2 and texts with low-cohesion in L2 was the most difficult condition among all four.

Learning assessment test (accuracy). As for accuracy**,** analysis showed a similar pattern. We found a significant main effect of language, $F(1, 45) = 22.77$, $p < .001$, $\eta_p^2 = .34$, with texts in L1 (M = 0.64, SE = 0.02) receiving higher scores in the learning assessment test than texts in L2 (M = 0.51, SE = 0.03); a significant main effect of cohesion, $F(1, 45) = 18.82$, $p < .001$, $\eta_p^2 = .30$, with high-cohesion texts receiving higher scores (M = 0.62, SE = 0.02) than low-cohesion texts (M = 0.53, SE = 0.02); and a significant main effect of proficiency group, $F(1, 45) = 17.9$, $p < .001$, $\eta_p^2 = .29$, with people in the higher-proficiency group (M = 0.66, SE = 0.03) achieving overall better learning than people in the lower-proficiency group (M = 0.49, SE = 0.03).

More interestingly, the main effect of language was modulated by an interaction with proficiency group, $F(1, 45) = 24.65$, $p < .001$, $\eta_p^2 = .35$. Post-hoc comparisons showed that the higher-proficiency group had comparable accuracy in both languages, $t(45) = -0.14$, $p = 1.00$, so their learning was similar in L1 (M = 0.66, SE = 0.03) and L2 (M = 0.67, SE = 0.04). In contrast, accuracy for people with lower-proficiency did differ between languages, $t(45) = 6.96$, $p < .001$, and they achieved a significantly better learning for texts in L1 (M = 0.63, SE = 0.03) than for texts in L2 (M = 0.35, SE = 0.04). It seems that proficiency plays a role when learning in L2 and lower-proficiency level might hinder learning.

The interaction between language and cohesion was marginally significant, $F(1, 45) = 2.88$, $p = .096$, $\eta_p^2 = .06$. Post hoc comparisons showed that accuracy for texts in L2 differed between cohesion condition, $t(45) = 4.50$, $p < .001$. Participants were more accurate in high-cohesion texts ($M = 0.56$, $SE = 0.03$) than in low-cohesion texts ($M = 0.45$, $SE = 0.03$). Nevertheless, no significant difference between cohesion condition was found for texts in L1, $t(45) = 2.36$, $p = .14$ (high-cohesion: $M = 0.67$, $SE = 0.02$; low-cohesion: $M = 0.61$, $SE = 0.03$). Again, it seems that the cohesion effect is significantly more salient in L2 not only in JOLs but also in learning.

Table 13.

Mean (and standard deviations) for JOLs scores (1-100 scale) and learning performance (proportion of correct responses) across language, cohesion and proficiency group conditions.

		High-cohesion		Low-cohesion	
Language	Proficiency group	JOL	Learning	JOL	Learning
L1	Higher-proficiency	59.5 (3.24)	0.70 (0.03)	64.5 (3.14)	0.62 (0.04)
	Lower-proficiency	62.8 (3.17)	0.65 (0.03)	67.8 (3.08)	0.60 (0.04)
L2	Higher-proficiency	66.3 (3.37)	0.74 (0.05)	56.7 (3.76)	0.59 (0.04)
	Lower-proficiency	55.3 (3.30)	0.39 (0.05)	43.5 (3.68)	0.31 (0.04)

Language metamemory accuracy –resolution–. We run a mixed-factor ANOVA and we found no significant main effect of Goodman-Kruskal Gamma correlations $F(1, 44) = 2.20$, $p = .15$, $\eta_p^2 = .05$, nor proficiency group, $F(1, 44) = 0.00$, $p = .96$, $\eta_p^2 = .00$. However, the interaction between both factors was significant, $F(1, 44) = 4.93$, $p = .03$, $\eta_p^2 = .10$. Post-hoc comparisons showed a marginal tendency for lower-proficiency group to have better resolution in L2 ($M = 0.32$, $SD = 0.1$) than in L1 ($M = -0.00$, $SD = 0.09$), $t(44) = -2.62$, $p = 0.07$, while no difference was found for the high-proficiency group (L1: $M = 0.18$, $SD = 0.09$; L2: $M = 0.12$, $SD = 0.1$, $t(44) = 0.52$, $p = 1.00$).

We also run JOL and accuracy language index correlations with the two proficiency groups independently, and we observed that a significant positive correlation appeared for people with higher-proficiency ($r = 0.42$, $p = .05$) while people with lower-proficiency showed a significant negative correlation between JOL and accuracy index ($r = -0.43$, $p = .04$).

Customized metacognitive self-report questionnaire. As in Experiment 1, we also explored whether the strategies differed between proficiency groups (see Table 14 for partial means in the questionnaire). We run repeated measures ANOVAs for each of the items in the questionnaire with language as a within-subject factor and proficiency group as a between-subject factor.

We found a significant interaction between **elaboration** strategy and proficiency group, $F(1, 45) = 4.37$, $p = .04$, $\eta_p^2 = .09$. However, post-hoc analysis showed no significant difference in any of the comparison (all $p > .05$).

We also found a significant difference in the use of **metacognitive self-regulation** between languages $F(1, 45) = 4.12$, $p = .05$, $\eta_p^2 = .08$, with such strategy being more prevalent in L2 ($M = 4.11$, $SD = 0.27$) than in L1 ($M = 3.73$, $SD = 0.25$) regardless of the proficiency group.

Regarding the use of **effort regulation**, we found a marginal significant main effect of proficiency group, $F(1, 45) = 3.85$, $p = .06$, $\eta_p^2 = .08$, that was mediated by a significant interaction, $F(1, 45) = 11.41$, $p = .002$, $\eta_p^2 = .20$. Post-hoc comparison showed a significant difference between proficiency groups in the use of effort regulation in L2 (high-proficiency group: $M = 2.43$, $SE = 0.39$; low-proficiency group: $M = 4.13$, $SE = 0.39$), $t(45) = -3.07$, $p = .02$.

The main effect of **mental demand** was also significant, $F(1, 45) = 49.1$, $p < .001$, $\eta_p^2 = .52$, and was mediated by a significant interaction with proficiency group $F(1, 45) = 10.5$, $p = .002$, $\eta_p^2 = .19$. People in the lower group reported higher mental demand in L2 ($M = 5.63$, $SE = 0.24$) than in L1 ($M = 3.38$, $SE = 0.3$), $t(45) = -7.33$, $p < .001$. Such difference was only marginally significant for the higher-proficiency group (L1: $M = 3.91$, $SE = 0.32$; L2: $M = 4.74$, $SE = 0.25$, $t(45) = -2.63$, $p = .07$).

We found a significant main effect of **performance**, $F(1, 45) = 7.05$, $p = .003$, $\eta_p^2 = .18$, with people reporting better self-perceived performance for texts in L1 ($M = 5.02$, $SE = 0.17$) than in L2 ($M = 4.47$, $SE = 0.19$), regardless of their proficiency level.

Similarly, we found a significant main effect of **effort**, $F(1, 45) = 4.47$, $p = .04$, $\eta_p^2 = .09$, with people reporting having made higher effort for texts in L2 ($M = 4.95$, $SE = 0.19$) than in L1 ($M = 4.55$, $SE = 0.21$), regardless of their proficiency level.

Table 14.

Means (and standard deviations) of the different dimensions of the metacognitive self-report questionnaire regarding learning strategies during the study phase by language block and proficiency group in Experiment 2.

	Higher-proficiency		Lower-proficiency	
Strategies	L1	L2	L1	L2
Rehearsal	5.39 (0.35)	5.43 (0.26)	5.58 (0.26)	5.46 (0.33)
Elaboration	5.30 (0.29)	5.74 (0.24)	5.83 (0.25)	5.33 (0.34)
Critical Thinking	3.26 (0.40)	3.26 (0.38)	3.71 (0.38)	3.79 (0.36)
Metacognitive self-regulation	3.91 (0.32)	4.13 (0.33)	3.54 (0.38)	4.08 (0.42)
Effort regulation	3.09 (0.40)	2.43 (0.33)	3.46 (0.42)	4.13 (0.44)
Self-perceived…				
Mental demand	3.91 (0.3)	4.74 (0.28)	3.38 (0.31)	5.63 (0.21)
Performance	5.09 (0.22)	4.78 (0.23)	4.96 (0.26)	4.17 (0.30)
Effort	4.57 (0.30)	4.65 (0.30)	4.54 (0.30)	5.25 (0.23)

Note: Participants self-rated the frequency they used each of the strategies in a 7-point Likert scale.

Discussion

In Experiment 5, we aimed to explore whether the effects encountered in Experiment 4 varied as a function of L2 proficiency level. Overall, we observed a similar pattern of results in both studies wherein we replicated the effects of language and cohesion in both JOLs and memory. Specifically, individuals assigned lower JOLs when studying texts in L2 (as opposed to texts in L1) and for low-cohesion texts (in contrast to high-cohesion texts), respectively. However, different patterns emerged between the higher- and lower-proficiency groups in both the monitoring measure (JOLs) during the study phase and in the subsequent learning assessment test.

First, the cohesion effect interacted with the language effect so that JOLs for texts with high-cohesion did not differ between languages, whereas JOLs for texts with low-cohesion were significantly higher in L1. The observed interaction suggests that high-cohesion texts may create a perception of learnability regardless of the language in which they are presented. However, the significant difference in JOLs between languages for low-cohesion texts indicates that the impact of cohesion on perceived learning difficulty might vary considerably between L1 and L2 and people found the L2 low-cohesion texts as the most difficult condition among all. For learning outcomes, this interaction revealed that people achieved similar learning within cohesion conditions for texts in L1, but high-cohesion texts were favored, compared to low-cohesion texts, in L2. Results from previous studies seem to contradict this pattern. For example, Jung (2018) investigated whether cognitive task complexity affects L2 reading comprehension and found that task complexity did not affect reading comprehension scores, although participants perceived the complex tasks significantly more demanding. Nevertheless, bilingual participants in Jung's (2018) study reported staying in English speaking countries, at least 6 months. In our experiment, even people in the high-proficiency group were moderate in English-L2 and reported significantly less frequent exposure and use of English-L2 compared to Spanish-L1. This makes the results of the two studies difficult to compare.

More interestingly, the language effect in JOLs was modulated by the proficiency level. The differences encountered in JOLs between languages were only evident for the lower-proficiency group, who predicted greater difficulties in L2 compared to L1, as opposed to the higher-proficiency group, who predicted similar performance in both languages. This was exactly what the learning assessment test revealed. People with higher-proficiency learnt information from texts in L1 and L2 equally. However, people with lower-proficiency level showed an L2 cost for learning. It seems that the proficiency level plays a crucial role in L2 self-regulated learning (SRL). These results go in line with previous research that had already reported that less proficient L2-English speakers needed longer time for reading, particularly when encountering sentences that conflicted with the previously established expectations in the text. This indicates a decreased efficacy in the high-level cognitive processes involved in L2 processing (Pérez et al., 2023). On the contrary, in their study higher-proficient participants showed better text comprehension, and better ability to generate predictive inferences. Thus, they concluded that linguistic proficiency makes a difference in high-ordered processes such as inferential evaluation, revision and text comprehension (Pérez et al., 2023). It seems that when studying in L2, lower-proficiency learners might encounter greater challenges, especially when the to-be-study material is ambiguous or incongruent.

In sum, it seems that people could correctly monitor their learning irrespectively of their proficiency level. People were able to detect the difficult parts of the material and to adjust their judgments accordingly. Moreover, the learning assessment was consistent with their predictions. This is true even when they are not highly proficient, as shown by the accuracy resolution. The interaction between the Goodman-Kruskal correlations and the proficiency group showed that the lower proficiency group was more accurate in L2 texts than in L1, suggesting that their performance in L2 was consistent with what they had previously predicted (JOLs) and reported in the questionnaire (higher mental demand L2 condition).

Finally, results from the qualitative questionnaire showed an interesting pattern and suggested that participants engaged different learning strategies depending on the language (L1 vs L2) and their L2 proficiency level. For example, metacognitive self-regulation was more frequently used in L2 than in L1 in both proficiency groups. More interestingly, lower-proficiency participants used effort regulation more frequently in L2. This may suggest that people had enough cognitive resources available and could use them to select efficient learning strategies even when studying in L2. Studies have highlighted the flexibility of cognitive processes, indicating that individuals can flexibly allocate cognitive resources depending on task demands and situational factors (Broekkamp & Van Hout-Wolters, 2007; Panadero et al., 2019). This adaptability allows individuals to optimize their learning strategies, even in challenging L2 learning contexts. Apparently, although participants could devote sufficient cognitive resources so as to unfold metacognitive strategies and correctly monitor their learning, L2 proficiency seems to play a critical role for learning outcomes.

Part III

Discussion and conclusions

General Discussion

In a world where English-L2 is increasingly used as the medium of instruction, many students with lower proficiency may struggle when studying materials in a language which they may not be proficient in. Thus, addressing how metacognitive and learning strategies unfold when studying in L2 is paramount.

The goal of this doctoral dissertation was to explore the consequences of studying in L2 on the metacognitive processes required for successful learning (Experiments 1-4), and whether they varied as a function of L2 proficiency (Experiment 5). For this, we conducted a set of behavioral experiments involving university students that were second-language learners. They were asked to study materials that varied its difficulty in L1 and L2. Overall, we found that studying in L2 did not compromise the monitoring process of learning. Participants, regardless of their proficiency level, could judge the materials accurately in both L1 and L2, by using language and other features as cues to assess the difficulty of the materials. Across experiments, participants tended to predict worse performance (i.e., gave lower JOLs) in L2 yet this pattern was mediated by block order effects specially with simple studying materials (Experiments 1-3). More interestingly, data from the memory tests broadly validated the pattern observed in the study phases with the JOLs, showing accurate monitoring in L1 and L2.

This section outlines the main discoveries derived from the five experiments of this doctoral dissertation, shedding light on their theoretical significance concerning metacognitive learning strategies and bilingualism. Below, we will examine the use of different cues in the monitoring process both in L1 and L2. We will discuss how these cues were informative regardless of the L2 proficiency level, suggesting that the monitoring process was accurate in both languages. We will also address compensatory effects and the learning strategies that learners unfold when studying complex materials such as texts in each language. Finally, we will draw a general conclusion derived from the results obtained in the present series of experiments.

Within language cues

The five experiments reported in this dissertation differed in the type of cues provided by the materials to guide learning monitoring: perceptual (font type, Experiment 1), lexical-semantic (concreteness, Experiment 2); and semantic-relational (relatedness, Experiment 3) for single words or short lists; and organizational (cohesion, Experiment 4-5) for short expository texts, both in L1 and L2.

At the word unit level, we did not find differences in JOLs due to perceptual cues. In Experiment 1, people did not find the *difficult-to-read* font less likely to remember than the easy-to-read font. The effect of font type on JOLs is not consistently observed in previous studies on L1 (Magreehan et al., 2016), and it does not always appear in recognition (Oppenheimer & Alter, 2014; Rummer et al., 2016; Xie et al., 2018). Although it has been claimed that that the perceptual cue might have a direct effect on JOLs and memory by affecting the ease of processing of an item (Yue et al., 2013), the significant impact of perceptually degrading materials is limited and occurs in specific conditions. Thus, perceptual difficulty seems to be moderated by a variety of factors and it seems to have relevance in specific situations and among certain individuals. For instance, presenting materials perceptually degraded enhanced the retention and understanding of expository text for individuals with high working-memory span (as suggested by Lehmann et al., 2016; although, Strukelj et al., 2016 have differing views). Moreover, it might impact individuals' assessments of their learning (as observed in Magreehan et al., 2016, Experiments 4 & 5) and in the allocation of study time in self-pace learning (e.g., Rummer et al., 2016). Test type might also play a role – disrupting the visual appearance of an item within a list may interfere with conceptual processing and item elaboration, effects that tend to be observed in a free recall test, rather than in a recognition test (Nairne, 1988). Regarding JOLs, *difficult-to-read* font type was found to decrease JOLs relative to the ease-to-read font, only when both of them are compared in mixed lists (Maxwell et al., 2021), and when immediate rather than delayed JOLs were solicited (Luna et al., 2018). Although the literature presents mixed results, it appears that font type variations have limited influence on both the subjective judgments and the actual storage or retention of information in memory. Despite using a within-participants manipulation and obtaining immediate JOLs to increase the likelihood of observing the font type effect, we did not find it in our Experiment 1. The reasons for this null effect are not readily apparent. They could be associated with the bilingual status of our participants or with the presence of language cues across blocks of trials. In any case, the most interesting pattern is the consistency of outcomes observed in L2 regarding the perceptual cue, mirroring those identified in L1. Thus, our results replicated some of the previous studies indicating lack of font type effects (e.g., Magreehan et al., 2016; Oppenheimer & Alter, 2014) both in L1 and L2.

In contrast, participants provided differential JOLs when lexical and semantic cues were introduced: lower JOLs to abstract and unrelated words than to concrete and related words in Experiments 2 and 3, respectively. These results are in line with previous research—concreteness and relatedness effects consistently appear in the JOL literature (e.g., Matvey et al., 2006; Witherby & Tauber, 2017). There are several cognitive mechanisms that contribute to the phenomenon where concrete words and semantically related words are often rated as more memorable and indeed are better remembered in memory assessments. Concrete words are characterized by their distinctiveness, imagery and contextual embedding (see the Dual Coding Theory, Paivio, 1991; and the Context-Availability Theory, Schwanenflugel et al., 1992). They often evoke more vivid mental images, representations and sensory experiences compared to abstract words (according to the Dual Coding Theory; Paivio, 1991) and they tend to have more contextual associations or connections to real-world experiences, making them easier to relate to and embed within existing knowledge structures (according to the Context-Availability Theory, Schwanenflugel et al., 1992). Semantically related list of words, on the other hand, share connections in meaning or context, leading to a network of associations (Collins & Loftus, 1975). This interconnectedness aids in memory processes as the activation of one word or concept can trigger the activation of related information, facilitating memory retrieval (e.g., Buchler et al., 2008; Desaunay et al., 2017; Neely & Tse, 2011). Hence, both concrete and semantically related words benefit from their sensory richness and shared semantic connections, being often processed more quickly and efficiently. Results from our studies confirm that this is also true for L2 processing: participants were sensitive to these differences in L1 and L2. This carries significant implication as it had been hypothesized that semantic activation in bilingual individuals' L2 could be weaker and that the lexical access in L2 is mediated by the L1, and thus is more indirect (Kroll & Stewart, 1994). Previous research indicated a delay in lexical access during L2 processing (Opitz & Degner, 2012) or a weaker activation of the semantic integration hub, indicating a different engagement of brain regions during L2 processing (Zhang et al., 2020). Despite this, our participants correctly monitored the materials in L2, identifying concreteness and relatedness cues.

At the text level, the results were in the same vein. People judged high-cohesion texts as easier to learn than low-cohesion texts in both languages. This observation aligns with established findings from previous research focusing on monolingual text comprehension and learning assessment (Carroll & Korukina, 1999; Lefèvre & Lories, 2004; Rawson & Dunlosky, 2002). We replicated these findings with a bilingual population, which highlights the robustness and generalizability of cohesion effects across linguistic contexts and the good metamemory resolution. Previous research has accounted for reading processes to be different in L1 and L2 (e.g., Cop et al., 2015; Dirix et al., 2020; Pérez et al., 2018; Whitford & Titone, 2012). For example, Whitford & Titone (2012) observed longer gaze durations and longer sentence reading times on embedded target words in L2 sentences; and Cop et al. (2015) suggested that the slower sentence reading times found in L2 were due to a higher number of fixations, which were longer and closer together, and to the fact that fewer words were skipped in L2 compared to L1 reading. Moreover, Pérez et al. (2018) demonstrated that high-ordered cognitive processes such as inferential revision was less efficient in the L2 compared to the L1. Thus, we might have expected that the metacognitive monitoring unfolded while reading might be impaired in L2, as the reading process is less automatic and more effortful. Nevertheless, it seems that how individuals assess and perceive text coherence and learning difficulty is independent of language. The results also support the notion that the influence of cohesion on learning assessment transcends language differences and remains consistent across diverse language backgrounds. From the self-regulated learning framework, metacognitive processes are assumed to be effortful and cognitively demanding (Efklides, 2014). Indeed, Stine-Morrow et al. (2006) showed that metacognitive monitoring might compromise memory performance in older people (whose executive control skills might be in decline). Similarly, Tauber and Witherby (2019) showed that age-related deficits might make it difficult for older participants to implement metacognitive strategies. Our data, nevertheless, suggests that these processes are not impaired in an L2 learning context. In fact, metacognitive monitoring was as effective in L2 as in L1. It is possible that the cognitive control dedicated to managing the demands of L2 and to facilitating metacognitive processes mutually supported each other rather than competed for resources. And thus, the diagnostic cues were monitored successfully. However, we must be cautious with our assumptions as we present only behavioral data, which has obvious limitations in terms of providing insight into underlying cognitive mechanisms.

Language as a cue

Regarding the language in which the study took place, participants found studying in L2 more difficult than in L1. However, this effect varied across experiments, such that in Experiments 1 and 2, where JOLs involved single words, this pattern was only evident when the L2 block was placed first. In contrast, in Experiment 3, where JOLs involved judging short lists, the language effect appeared independently of block order. It is possible that when considering complete lists (Experiment 3), not only might have learners activated mental representations of the word but they might have also accessed representations of associated words. Memory links between words and conceptual representation have been shown to be stronger in L1 than in L2 (Kroll & Stewart, 1994). Therefore, words might have stronger links in L1 than in L2, and this, in turn, might manifest in differential judgments for L1 and L2. Different relational representations for L1 and L2 would be independent of whether the L2 block was presented first or second, and therefore, the language effect in Experiment 3 is independent of the block order. In contrast, when the task involved single words (Experiments 1-2) and no further relational processing was required, JOLs depended on the calibration between the two blocks; thus, participants considered L1 to be easier (increased their JOLs) after first experiencing the more difficult L2 language condition. Hence, it seems that participants used the first block as a baseline for comparison and considered L1 to be easier when contrasted with the more difficult L2 block. The reason why this contrast effect was only obtained when participants were first presented with L2 is not evident and might be due to a number of reasons (e.g., more interference from L1 that rules out the possible benefits of greater effort for L2). Nevertheless, in line with many previous bilingual studies (Beatty-Martínez et al., 2020, 2021; Beatty-Martínez & Dussias, 2017), these results provide evidence that the observed L2 effects are dependent on the context in which L2 learning is achieved.

People also judged texts in L2 as more difficult than texts in L1, both when language was blocked (Experiment 4, although the language effect did not interact with block order) and also when language was mixed within the study phase (Experiment 5). This language effect was evident for high- and low-cohesion texts in Experiment 4 but was modulated by the interaction with cohesion in Experiment 5. This interaction reflected that there was no difference in L1 and L2 for high-cohesion texts, yet for low-cohesion texts, L2 condition received significantly lower JOLs. This interaction did not appear in Experiment 4. Note that a difference between both experiments was that language was blocked in Experiment 4, whereas it was random in Experiment 5. Previous studies support the view that recognizing and integrating a linguistic code different from that most recently encountered entails a processing cost. Furthermore, code mixing demands a high level of proficiency and cognitive control in both languages (e.g., Costa & Santesteban, 2004; Grainger & Beauvillain, 1988; Grainger & O'Regan, 1992; Green & Wei, 2014; Johns et al., 2019; Moreno et al., 2002). Code-switching costs typically arise due to the cognitive demands required to navigate between the two languages. When bilingual individuals are forced to switch between languages within a context, they need to manage different linguistic systems, select appropriate words or structures, and inhibit interference from the non-relevant language. This process demands cognitive resources and the need to manage and control two linguistic systems simultaneously can have a significant impact on the efficiency in language use (Costa & Santesteban, 2004; Green & Wei, 2014). In Experiment 5, mixing languages in text presentation may have heightened the salience of language as a cue due to the need of language control derived from the switching. More cognitive resources were devoted to the task because of the necessary language control. Consequently, the cohesion cue was more prominent under the more demanding language condition (L2), which made the interaction between the two factors (cohesion x language) appeared.

Monitoring in L1 and L2

The most remarkable pattern is that language did not impede monitoring under any of our manipulations. Even though participants judged studying in L2 as more challenging, this difficulty did not preclude the use of monitoring processes that allowed them to detect difficult material and to accurately judge it as more challenging to learn. Monitoring was similarly performed in L1 and in L2. Our unbalanced bilingual participants demonstrated similar ability to assess their learning levels for features like concreteness, relatedness, and cohesion in both L1 and L2. This indicates that, at least for bilinguals with intermediate proficiency level, learning monitoring and control are not impaired by the L2 context. People assessed their learning correctly –monitoring process– (as shown by the JOLs) and could compensate in some circumstances –control process– (as shown by the memory tests; see the subsection of *Compensatory effects* for further explanation).

Data from the memory tests validated the pattern of results found in the study phase (JOLs) for the five experiments, suggesting a good metamemory resolution and an accurate use of the cues in both languages. Indeed, participants recalled the words similarly regardless of the font type (Experiment 1). However, they remembered concrete words better (as opposed to abstract ones, Experiment 2), semantically related words (as opposed to unrelated words, Experiment 3), and high-cohesion texts (as opposed to low-cohesion texts, Experiment 4-5), which goes in line with previous research on memory (e.g., Begg et al., 1989; Crossley et al., 2014; Undorf & Erdfelder, 2015).

The absence of an effect with font type variations in memory tests suggests that font type itself might not inherently introduce difficulty, hence the lack of a noticeable impact, which was also reflected in the JOLs (e.g., Magreehan et al., 2016; Oppenheimer & Alter, 2014). Conversely, the differing patterns observed for concrete vs. abstract words, semantically related vs. unrelated words, and high-cohesion vs. low-cohesion texts underscore a distinct variation in the difficulty of the materials. Concrete words, semantically related words, and high-cohesion texts inherently present an objectively ease of processing for storage and retrieval, supported by participants' memory patterns on both recognition tests and open-ended comprehension questions. Concrete words have richer and more interconnected semantic representations than abstract words (Paivio, 1991), which boost encoding and retrieval. Moreover, it is well-established that lexico-semantic processing of a word is facilitated when it is preceded by a semantically related word (Moreno & van Orden, 2001; Mummery et al., 1999; Wagner & Koutstaal, 2002). The semantic relatedness between words tends to strengthen the associative memory representation, establishing a link between two items, thereby aiding in encoding and retrieval processes. When reading the first word in a related list, this facilitation extends to the subsequent words in the sequence, creating a chain effect. Thus, semantically related word lists are bettered remembered than unrelated word lists (Janes et al., 2018; Mueller et al., 2013; Undorf & Erdfelder, 2015). Finally, poor cohesion texts impose higher demands on readers as they need to produce inferences in order to create a meaningful representation of the information in the text (Best et al., 2005). Consequently, these texts are associated with poorer comprehension (Crossley et al., 2014, 2016; Hall et al., 2016). This indicates that certain conditions within the study materials are indeed objectively more difficult than others. The features manipulated in our studies (except for the perceptual features) were informative for memory, and JOLs were also sensitive to these subtleties. We must highlight that this was the case for L1 and also L2 materials, which carries significant implications concerning metacognitive processes – the pattern of results presented so far supports the idea that metacognitive monitoring is not impaired due to L2 instruction.

Compensatory effects

Compensatory effects for L2 language difficulties. With regard to language compensation, it appears that when dealing with simpler materials like individual words (Experiments 2 and 3), participants could potentially compensate for the inherent difficulty posed by learning in L2. This compensation was evident in the better recognition rates observed for words in L2 compared to L1 in Experiments 1 and 2. However, as the complexity of the material increased, such as texts (Experiments 4 and 5), participants seemed unable to adequately compensate for the inherent difficulty of learning in L2 to achieve a comparable level of comprehension as in L1, especially at lower L2 proficiency levels (Experiment 5). Thus, our data suggest that lower proficiency individuals can unfold monitoring processes effectively, however they might struggle to compensate for the perceived L2 difficulty and achieve successful learning. Previous studies have provided evidence that the cost of studying in L2 depends on several factors such as study time, test requirement, and language proficiency level (Vander Beken et al., 2020), a notion supported by our results.

The lack of compensatory effects with complex text material might be related to the inherent difficulty of the processes involved in text comprehension (vs. single words). Thus, with single words, cognitive resources might have been sufficient to manage the language difficulty. In contrast, texts demand deeper semantic processing, inferencing, revising and updating to create appropriate mental models of the information in the text (Pérez et al., 2018). It is possible that the increased difficulty of L2 language processing might interfere with this deeper level of understanding. This might have resulted in participants having a reduced capacity to allocate cognitive resources to compensate for the language difficulty. The idea that simpler materials may allow individuals to compensate for language difficulty more effectively compared to more complex and cognitively demanding materials is rooted in cognitive load theory and theories of second language processing (e.g., Bannert, 2002; Boekaerts, 2017; Seufert, 2020; Wirth et al., 2020). There is previous evidence regarding the challenges associated with compensating for language difficulty in various learning tasks, highlighting how task complexity can impact learners' ability to allocate cognitive resources effectively (e.g., Skehan, 1998; Zirnstein et al., 2019).

Compensatory effects for within language difficulties. Compensatory effects for within language difficulties (i.e., abstract words, unrelated lists or low cohesion texts) were not evident across experiments. Thus, compensation did not appear for the lexical-semantic and semantic-organizational manipulations: concrete words, related lists, and high-cohesion texts still showed higher accuracy in the memory tests compared to abstract words, unrelated lists, and low-cohesion texts. Thus, although participants' JOLs were sensitive to the objective difficulty of the materials, they did not spontaneously use this knowledge to compensate for abstract and unrelated words, and low-cohesion texts, and achieve the same level of learning as with the easier materials.

The cohesion effect, though, showed an intriguing pattern whereby results differed across languages. In Experiment 4 people did not compensate for the difficulty detected in the study phase for low-cohesion (in contrast to high-cohesion) texts. Similarly, in Experiment 5, lower-proficiency individuals did not show compensation effects despite they also detected the difficulty of the low-cohesion texts. The only condition where learning for low-cohesion and high-cohesion texts was comparable was when higher-proficiency individuals studied in L1. However, the effect of cohesion was also absent in JOLs for these individuals, suggesting that for higher-proficiency individuals, the lack of cohesion effects in learning outcomes was not due to compensation but to the fact that low- and high-cohesion texts produced similar levels of difficulty for them. As mentioned, this might be due to the greater engagement of control processes when L1 and L2 are presented in a mixed format which may induce better learning in L1 (e.g., Costa & Santesteban, 2004; Green & Wei, 2014). That is, the increased engagement of control processes in the language mixing format may have facilitated the monitoring and learning of L1 texts. Consequently, the cohesion effect might not have been apparent. Moreover, low-cohesion texts often require readers to make inferences due to the lack of explicit connections between sentences or paragraphs. These texts contain more ambiguous or vague content that may induce readers to employ higher cognitive functions, such as reasoning, anticipating information and inferencing, to discern the implied meaning. This inference-making process demands greater mental effort and engagement from the reader, as they actively construct connections to comprehend the text effectively. The development of this process might not have been possible in L2, as L2 processing recruit a significant number of cognitive resources and people have less available to devote to inference-making. This view is supported by the work developed by Pérez et al. (2018), who explored the ability to revise inferential information in L1 and L2. Participants were presented with short narrative texts whose first sentence facilitated a predictive inference. Then, either a consistent, neutral or inconsistent message appeared so that participants needed to revise their inference, and update the information. They concluded that although participants carried out some level of revision when reading in their L2, this process seemed quantitatively less efficient than in L1.

The lack of evidence for compensation in many conditions of our experiment is consistent with previous studies indicating that people do not fully compensate for item difficulty effects through self-regulation. Although, one would expect that if learners detect difficulties in the learning materials, they would compensate for this by allocating more time or by selecting a different strategy to better learn this information, there is strong evidence that neither self-paced study (i.e., how people allocate their study time), item selection for re-study, or the use of strategies (e.g., distributed practice, retrieval practice) completely compensate for difficulty (Cull & Zechmeister, 1994; Koriat, 2008; Koriat et al., 2006; Koriat & Ma'ayan, 2005; Le Ny et al., 1972; Mazzoni et al., 1990; Mazzoni & Cornoldi, 1993; Nelson & Leonesio, 1988; Pelegrina et al., 2000; see Tekin, 2022 for a review).

Koriat et al. (2006) proposed that the relationship between monitoring and control processes arises from the fact that metacognitive judgments are based on feedback from the outcome of control operations. Monitoring does not occur prior to the controlled action, but rather, it takes place afterwards. According to this hypothesis, the difficulty of an item is monitor *ad hoc*: learners allocate the appropriate resources to an item based on its demands, and they recognize that a specific item will be challenging to remember when they realize that it requires a relatively higher level of effort to commit to memory. Thus, although the initial assessment of a situation provides valuable information for executing control actions, the feedback obtained from these actions can subsequently be used as a foundation for monitoring. This monitoring process, in turn, can guide future control operations, creating a cyclical relationship between monitoring and control. In other words, subjective experience informs the initiation and self-regulation of control operation that may in turn change subjective experience. And this hypothesis may account for the lack of compensation in our data.

L2 Proficiency

The level of proficiency appears to introduce subtleties in the learning processes that could affect the resources accessible for metacognitive processing during the execution of L2 tasks. In Experiment 5, people with higher-proficiency level judged L1 and L2 texts equally easy to learn, and indeed did not show any sign of L2 cost in the learning assessment test. The lower-proficiency group, though, considered L2 texts as more difficult during study and actually showed a disadvantage in L2 learning. Hence, lower-proficiency participants can still metacognitively monitor the difficulty of the texts and detect difficulties, yet they cannot compensate for such difficulty. Nevertheless, they still show an L2 cost in terms of memory and learning. The poorer learning outcomes in L2 for lower-proficiency compared to higher-proficiency participants align with previous research findings. Vander Beken et al. (2018) reported a recall cost in L2 when participants were tested with essay-type questions, presumably attributed to a lack of writing skills. They suggested several reasons why writing skills could affect the L2 recall cost.

Firstly, L2 proficiency, being lower than L1 proficiency, significantly impacts both text comprehension and composition. This disparity in linguistic knowledge, especially in grammar, not only affects the understanding of text but also poses challenges in generating text, hampering sentence construction. Additionally, the proficiency level in L2 influences the time required for the "formulation", a sub-process in essay-type writing that involves organizing thoughts, structuring sentences, selecting appropriate vocabulary, and arranging these elements coherently to convey the intended message effectively (De Larios et al., 2001). Furthermore, interference from L1 is evident in L2 writing, as indicated by errors typical of the writer's L1 background (Watcharapunyawong & Usaha, 2013). Working memory capacity also emerges as a more critical factor in L2 writing quality (Bergsleithner, 2010). Lastly, motivation tends to be lower in a non-native language, potentially leading to writing anxiety due to L2 writing challenges, thereby hindering performance (Mat Daud et al., 2005).

We intended to overcome this issue by including open-ended questions that did not require much elaboration, which would prevent the production deficit. Moreover, our rubric accepted answers if they included the key words, and did not punish language mistakes such as grammatical, syntactic, or orthographic errors as long as they do not obscure meaning. One might claim that both proficiency groups differed in their reading and writing skills, that the lower-proficiency group is composed of unskilled or inexperienced readers in general. However, proficiency groups did not differ in their accuracy for texts in L1, which proves that the differences in the learning assessment test cannot be explained by people in the higher-proficiency group being better comprehenders. Hence, several factors could explain our lower-proficiency group recall cost in L2: impaired encoding, difficulty in integrating the information or simply retrieving it from memory. Connor (1984) also noted a recall disadvantage in L2, where native English speakers were able to remember significantly more subordinate sentences compared to second language learners. Dhaene & Woumans (2022) replicated the L2 recall cost and could not attenuate the effect with the use of advance organizers[3] along the to-be-studied materials. They found that L2 test scores in the advance organizers condition were comparable to those of L1 in the non-advance organizers condition. Thus, they concluded that utilizing advance organizers does not completely eliminate the disadvantages experienced by L2 learners, but it helps bring L2 students to the same starting level as L1 students. Future research should incorporate different testing format and allow self-paced reading, for example, so as to address these issues. In fact, some have claimed that providing sufficient time might be a crucial factor in effectively learning content in L2 (Vander Beken et al., 2020). It may occasionally turn L2 studying into a "desirably difficulty" (e.g., Metcalfe, 2011), where the more challenging learning conditions contribute to the formation of stronger memory traces.

Learning strategies

An additional point has to do with the learning strategies used by participants in the study phase. We found some commonalities and differences across experiment. Grouping words by their semantic meaning (86.1%), creating mental images (68.8%) and words rehearsal (76%) were the most prevalent strategies in Experiments 1, 2 and 3 respectively. Interestingly, all these strategies reflect prevalence of semantic processing across the three experiments, although we observed subtle differences among them. The

[3] Advance organizers are designed to provide a framework or structure for upcoming information, making it easier for students to comprehend and organize new content. They can take various forms such as visual representations (charts, diagrams, or concept maps), summaries or outlines presented before the main content, comparisons or symbolic representations, and pre-questions or thought-provoking prompts. The purpose of advance organizers is to activate prior knowledge, provide a structure for understanding, and facilitate the integration of new information by linking it to what students already know (Dhaene & Woumans, 2022).

underlying reason that modulates such differences is not evident and further research should also address this issue. More importantly, this research should also address whether language related differences modulate the use of this strategies.

In Experiment 4 and 5, results from the customized metacognitive questionnaire suggested that participants engaged different learning strategies when studying in L1 and L2. This might support the idea that they confront L2 learning with different strategies. Overall, people relied on deep-level strategies more in L1 than in L2. Deep-level strategies are considered to have more impact on long-term learning (Deekens et al., 2018; Lonka et al., 2004; Vermunt & Vermetten, 2004). For example, Deekens et al. (2018) discovered that students who practiced more frequent learning monitoring also employed deeper learning strategies more regularly than those who monitored less, resulting in superior academic performance. This observation indicates a close relationship between metacognitive monitoring and the application of deep learning strategies, contributing significantly to successful academic outcomes. Although our participants monitored correctly in both languages, it appears that L2 processing is notably more demanding and occurs within a potentially overloaded cognitive system (Adesope et al., 2010; Hessel & Schroeder, 2020, 2022; Pérez et al., 2018), allowing for fewer opportunities to employ deep-learning strategies.

However, the increased frequency of metacognitive self-regulation usage in L2 compared to L1 in both proficiency groups in Experiment 5 contradicts this hypothesis. This implies that individuals possibly had sufficient cognitive resources at their disposal and were able to utilize them to adopt efficient learning strategies even when studying in L2. Then, it could potentially be a matter of time allocation. The adoption of certain strategies might necessitate additional study time, which was not feasible within the time constraints of our experiments. Consequently, individuals might have employed deep-level strategies more often in L1, as they might have required extended study time for their application in L2 learning (Stoff & Eagle, 1971). However, the qualitative difference in language-specific use of metacognitive strategies is an aspect worth further exploration in future research.

Conclusion

In conclusion, this doctoral dissertation has shed light on several aspects of metacognitive learning strategies in second language learners. The findings suggest that monitoring strategies are preserved in both L1 and L2 contexts, demonstrating that individuals can effectively monitor study materials irrespective of the language in which they are presented. In sum, self-regulated learning is a complex process that appears not to be hindered by L2 processing.

As the self-regulated learning unfold, monitoring and control processes arise so as to evaluate and adjust one's attention, understanding, and behavior. People estimate task difficulty, make learning judgments and allocate resources in accordance, adjusting the pace of learning or regulating strategies use, for instance (Panadero, 2017). Furthermore, results from Experiment 5 raise a note of caution for L2 learning. Proficiency level seem to introduce some nuances in the learning processes, which may impact the resources available for metacognitive processing while performing L2 tasks. Future research should identify the direct and indirect effects of language proficiency and self-regulated learning strategies on learning in bilingual contexts.

Moreover, there are other crucial factors that might also play a role in the self-regulation of learning (Lonka et al., 2004). The difficulty of the materials and language proficiency might not be the only aspects that guide self-regulated learning: executive functions, motivation, self-efficacy and other task or situational demands can also influence the learning process. A framework that incorporates all these metacognitive and motivational components of self-regulation into account would be more suitable to explain the effects of learning in L2 on monitoring and retention.

So far, we have explored the monitoring process and the short-term consequences in memory and learning derived from studying in L2. Nevertheless, the data presented here have also limitations because they are based on behavioral assessments, which do not allow us to dissociate underlying cognitive mechanisms. Other areas remain unexplored, notably regarding control strategies. Future research should delve deeper into this domain, employing techniques like *eye-tracking* to investigate reading patterns related to, for example, text integration. Additionally, the study highlights the intriguing effects of language mixing within study materials, suggesting a need for further exploration into how the interplay between L1 and L2 influences metacognitive processes in the learning context. These pending issues offers promising directions for future investigations that would enrich our understanding of metacognitive strategies in L2 learning contexts.

Resumen y conclusiones.

En las últimas décadas, el uso de un segundo idioma (L2) se ha convertido en algo fundamental de la vida diaria para una gran proporción de la población (Grosjean, 2010). Este fenómeno lingüístico se manifiesta en una variedad de contextos donde individuos bilingües, con diferentes experiencias y habilidades lingüísticas, procesan información, adquieren nuevos conocimientos y toman decisiones a través de su L2.

El bilingüismo puede entenderse como un espectro que abarca una amplia gama de experiencias lingüísticas y culturales. Es un término con una definición amplia y dinámica, que varía según las características únicas de cada individuo (Luk y Bialystok, 2013). Al hablar de bilingüismo, es necesario considerar variables como la edad de adquisición, el porcentaje de tiempo que se dedica a cada idioma en la vida diaria y medidas objetivas de competencia lingüística. Todo ello nos ayuda a crear una imagen más precisa de las habilidades y de las experiencias lingüísticas de cada persona.

Por tanto, no existe una definición única y esto nos permite apreciar la diversidad de individuos y situaciones bilingües. Desde esta perspectiva, el bilingüismo incluye a toda persona que utiliza dos o más idiomas en su vida diaria, independientemente de su nivel de competencia (Cenoz, 2003). En la sociedad actual, podemos observar multitud de situaciones en las que el uso de L2 es indispensable: ámbito profesional, negocios y corporaciones multinacionales; ámbito político, diplomacia y relaciones internacionales; redes sociales, entretenimiento y medios de comunicación, por ejemplo. Además, y con gran importancia, el inglés es a menudo el idioma principal de enseñanza y comunicación en el ámbito de la investigación y en la educación superior, incluso para hablantes no nativos. De hecho, actualmente, la educación bilingüe es más la norma que la excepción. Existen numerosos contextos, y a diferentes niveles académicos, en los que la instrucción se realiza en un segundo idioma en el que, muy frecuentemente, no se tiene el mismo nivel de dominio que en nuestra lengua materna. Por ello, estudiar en L2 es todo un desafío para un gran número de personas.

A pesar de que el uso de un segundo idioma en diversos contextos es una práctica extendida, la investigación sobre cómo cambian las estrategias metacognitivas y de metamemoria cuando se aprende en un segundo idioma sigue siendo limitada (véase Buehler y cols., 2021, para un intento de explorar los mecanismos subyacentes al bajo rendimiento de estudiantes de primaria no nativos). Se han realizado numerosos estudios sobre el impacto de llevar a cabo tareas específicas en L1 o L2 en actividades como la toma de decisiones (Brouwer, 2021), la atención visual (Chabal y Marian, 2015), la percepción de emociones (Chen y cols., 2022), la memoria a largo plazo (Arndt y Beato, 2017; Marian y cols., 2021), la comprensión lectora y las inferencias (Pérez et al., 2018), e incluso la memoria prospectiva (López-Rojas y cols., 2022).

No obstante, todavía se sabe poco acerca de si los procesos metacognitivos involucrados en el aprendizaje experimentan cambios significativos cuando éste se realiza en L2. Comprender estos mecanismos es fundamental en entornos educativos, para detectar el origen del posible peor rendimiento de estudiantes no nativos (Buehler y cols., 2021). A medida que nuestro mundo continúa volviéndose más interconectado, es probable que se amplíe el papel de L2 en diversos aspectos de la vida. Esto hace cada vez más importante explorar el impacto del L2 en la cognición y el aprendizaje autorregulado. El objetivo de esta tesis es investigar los factores que median en los procesos metacognitivos de aprendizaje en contextos dominados por un L2.

Incluso para personas bilingües con alta competencia en el idioma, trabajar en L2 puede ser un desafío cognitivo (Ma y cols., 2014; Moreno y cols., 2010; Pérez y cols., 2018). Un gran número de estudios han demostrado el fenómeno de coactivación, por el que los dos idiomas se mantienen activos dentro del cerebro bilingüe en procesos de producción y comprensión del lenguaje, tanto en la modalidad escrita como hablada, e incluso en contextos en los que solo se está usando uno de los idiomas (Bialystok y cols., 2012; Bialystok, 2017; Chen y cols., 2017; Giezen y cols., 2015; Hatzidaki y cols., 2011; Iniesta y cols., 2021; Kroll y cols., 2014; Shook y Marian, 2012).

Básicamente, cuando una persona bilingüe habla o procesa información en uno de los idiomas que conoce, el otro(s) idioma(s) está(n) activo(s). Imagina a una persona bilingüe conversando en inglés-L2. Cuando aparece la palabra *candle* –vela– en la conversación, su cerebro automáticamente coactivará palabras en inglés con sonidos similares (p.ej., *candy* –caramelo–), pero también palabras que comparten fonología en su otro idioma (p.ej., *candado* en español-L1). Y no solo eso, sino que también activará las traducciones de esas palabras en otros idiomas (Shook y Marian, 2019), lo que puede dar lugar a interferencias léxicas y sintácticas momentáneas. Del mismo modo, al leer una palabra en un idioma, se activan las representaciones escritas y habladas de los demás idiomas conocidos. Estudios con palabras homógrafas (comparten grafía pero tienen distinto significado entre dos idiomas diferentes) y cognadas (comparten grafía y significado entre dos idiomas diferentes) en procesos de lectura (Dijkstra y Kroll, 2005; Libben y Titone, 2009; Schwartz y Kroll, 2006), escritura (Iniesta y cols., 2021) y comprensión oral (Marian y Spivey, 2003a), además de estudios con *eye-tracking* (Marian y Spivey, 2003b; Shook y Marian, 2019) aportan evidencia acerca de este fenómeno de co-activación.

Para controlar la interferencia y seleccionar activamente el idioma deseado según el contexto, es necesario involucrar recursos cognitivos adicionales de control y metacontrol (Beatty-Martínez y cols., 2020; Kroll y cols., 2015; Macizo y cols., 2010; Soares y cols., 2019). Estos mecanismos cognitivos se emplean para gestionar la coactivación de los idiomas y elegir el idioma apropiado para un contexto dado, lo que implica suprimir la interferencia del idioma no deseado y asegurarse de que el idioma adecuado se utilice con precisión y fluidez (Declerck y Koch, 2023). Por ejemplo, al hablar en inglés-L2, una persona bilingüe debe inhibir la tendencia a insertar, de manera no intencionada, palabras o estructuras de oraciones del español-L1. En este sentido, los estudios de imágenes cerebral han revelado que las bases neurales del control del idioma en personas bilingües comparten redes cerebrales con procesos de control cognitivo de dominio general (p.ej., Calabria y cols., 2018).

Estos procesos de control de idiomas podrían ser especialmente costosos para personas bilingües que muestran un desequilibrio en el dominio de ambas lenguas, ya que la interferencia de L1 (dominante) a L2 (menos dominante) es mayor que viceversa (p. ej., Contemori y Dussias, 2016; Meuter y Allport, 1999; Soares y cols., 2019). Este tipo de bilingües *no-balanceados* y de adquisición tardía del L2 dependen principalmente de la traducción del L1 al L2 ya que, en su caso, las representaciones semánticas de los conceptos son más débiles en L2 (Kroll y Stewart, 1994). Además, algunos estudios han demostrado que el procesamiento en L2 lleva un 20% más de tiempo, el reconocimiento de palabras es más lento y somos capaces de procesar una cantidad menor de información simultáneamente, en comparación con L1 (Dirix y cols., 2020).

Estos mecanismos de control de idiomas adquieren especial relevancia durante procesos de comprensión lectora en L1 y L2. Cuando leemos, numerosos procesos cognitivos y metacognitivos se ponen en marcha para facilitar la comprensión. En su nivel más básico, la lectura implica el reconocimiento visual de palabras escritas. Pero, la comprensión va más allá del mero reconocimiento de palabras e incluye procesos de nivel superior, como la actualización de información, las inferencias, la integración con el conocimiento previo así como la capacidad para monitorizar y evaluar la dificultad del texto (Castles y cols., 2018). Por lo tanto, el procesamiento del texto requiere la construcción de una representación mental (es decir, un modelo situacional, Van Dijk y Kintsch, 1983) basado en la interacción entre procesos de diferente complejidad: reconocimiento visual de palabras, es decir, decodificar la conexione entre palabra-ortografía-significado, e inferencias de significados a partir del texto circundante (van den Broek y Helder, 2017), por ejemplo.

Para lectores expertos, estos procesos operan de manera fluida y sin esfuerzo en L1, facilitados por un profundo conocimiento de la sintaxis, semántica y pragmática del idioma. En L1, estos mecanismos se desarrollan y perfeccionan a lo largo de años de exposición y práctica, lo que resulta en un reconocimiento rápido y preciso de palabras, una correcta actualización de información y buenas inferencias. Sin embargo, cuando las personas leen en L2, especialmente cuando su competencia en L2 no alcanza un nivel alto, el panorama psicolingüístico experimenta cambios significativos.

Investigaciones previas muestran una peor comprensión en L2 comparada con L1 (para una revisión, véase Melby-Lervåg y Lervåg, 2014). Por ejemplo, las palabras y las oraciones en L2 suelen procesarse a un ritmo más lento en comparación con sus equivalentes en L1 (Cop y cols., 2015; Whitford y Titone, 2012), y se suele retener un menor número de palabras en L2 en comparación con conceptos homólogos en L1 (Gablasova, 2014). El proceso de reconocimiento de palabras en L2 puede requerir más esfuerzo y atención consciente, lo que puede afectar no solo a la fluidez de lectura, sino también a la comprensión (Plat y cols., 2018; Schwartz y Kroll, 2006).

Sin embargo, no se puede pasar por alto el impacto que el nivel de dominio en L2 puede ejercer sobre el aprendizaje de textos en L2. La razón subyacente aquí es que la capacidad para construir las inferencias en L2 está limitada por la competencia lectora en L2 y el nivel de dominio del idioma. Cuando las habilidades de lectura en L2 son limitadas, las personas bilingües se ven obligadas a priorizar procesos de lectura fundamentales como la decodificación de palabras y el análisis sintáctico, sobre el procesamiento inferencial, asignando sus recursos cognitivos en consecuencia (Horiba, 1996; Hosoda, 2014). Dados los desafíos en el procesamiento inferencial, es probable que las personas menos competentes en L2 encuentren difícil construir modelos situacionales comprensivos, mientras que las personas altamente competentes pueden emplear su conocimiento previo con mayor efectividad y mejorar su comprensión del texto (Hosoda, 2017).

En resumen, toda la evidencia presentada hasta ahora sugiere que el procesamiento en L2 es más costoso que en L1 y podría ocurrir dentro de un sistema cognitivo presumiblemente sobrecargado. La pregunta es si esta sobrecarga cognitiva derivada de trabajar en L2 tiene consecuencias en las estrategias de aprendizaje y los recursos cognitivos dirigidos a procesar, estudiar y adquirir nueva información en L2. Es decir, ¿somos capaces de evaluar la dificultad del material y ajustar nuestras estrategias en concordancia de forma igualmente efectiva en L1 y L2? O, ¿estudiar en L2 supone un coste añadido que debemos asumir?

Desde la perspectiva del aprendizaje, la metacognición es una función clave que sirve para fines autorregulatorios, en la que el cerebro monitoriza el proceso de aprendizaje y regula los recursos dedicados al mismo. Está involucrada en el desarrollo de estrategias de aprendizaje exitosas relacionadas con el rendimiento académico. Según el modelo clásico de Nelson y Narens (1990), existen dos mecanismos subyacentes en las estrategias metacognitivas del proceso de aprendizaje: *monitorización* y *control*. La *monitorización* se refiere a la supervisión y evaluación de la efectividad de los recursos cognitivos. Estos procesos de monitorización desencadenan los procesos de *control* necesarios para seleccionar, regular y ajustar los recursos cognitivos dedicados a la tarea. Por ejemplo, detectar las partes de mayor dificultad de un texto –proceso de monitorización– nos permite regular nuestros recursos atencionales y desplegar estrategias de relectura –proceso de control–, que nos harán aumentar la comprensión del texto en su conjunto.

Estas dos funciones metacognitivas –monitorización y control– están conectadas entre sí y tienen consecuencias en la memoria. Algunas teorías sugieren que una correcta monitorización conduce a una regulación adecuada, en beneficio del aprendizaje y del rendimiento de la memoria (Pieger y cols., 2016). En este sentido, Deekens y cols., (2018) investigaron la relación entre la frecuencia de monitorización metacognitiva y la utilización de estrategias de codificación a nivel superficial y profundo. Las estrategias superficiales implican invertir un tiempo y un esfuerzo mínimos para cumplir con los requerimientos de la tarea de aprendizaje (p.ej., memorización de conceptos clave), mientras que las estrategias a nivel profundo involucran prestar atención al significado del material, relacionar ideas e integrarlas con conocimientos previos para maximizar la comprensión. Se considera que las estrategias a nivel profundo son más efectivas y generan un aprendizaje más duradero. En su estudio, Deekens y cols., (2018) encontraron que los estudiantes que monitorizaban su aprendizaje de manera más frecuente también empleaban con mayor frecuencia estrategias a nivel profundo frente a los estudiantes con una monitorización más baja, lo que resultaba en un mejor desempeño en evaluaciones académicas. Este patrón sugiere que la combinación de monitorización metacognitiva y estrategias de aprendizaje a nivel profundo están intrínsecamente relacionados con el éxito académico.

La interacción entre los procesos de monitorización y control es exigente y demandante en términos cognitivos (Efklides, 2014). Para que estos dos procesos ocurran y se pueda integrar el flujo de información simultáneo entre ambos, debe haber suficientes recursos cognitivos disponibles y un control cognitivo activo. Algunas investigaciones apoyan que estos procesos metacognitivos reclutan recursos cognitivos generales. Por ejemplo, Stine-Morrow y cols. (2006) demostraron que el aprendizaje en un grupo de personas mayores se redujo en comparación con un grupo de participantes más jóvenes cuando se requería monitorización, lo que sugiere que la monitorización metacognitiva podría comprometer el rendimiento en el grupo de edad más avanzada, cuyas habilidades de control ejecutivo están en declive. De manera similar, Tauber y Witherby (2019) mostraron que, a diferencia de los adultos más jóvenes, las instrucciones para utilizar estrategias metacognitivas no hicieron mejorar el rendimiento en adultos mayores, lo que sugiere que las deficiencias de control cognitivo relacionadas con la edad podrían dificultar la implementación de estas estrategias en participantes mayores. Del mismo modo, los estudios neuropsicológicos proporcionan evidencia de que los correlatos neurales de la metacognición están relacionados con los lóbulos frontales (Pannu y Kaszniak, 2005). Finalmente, una revisión de estudios de imagen cerebral revela que las áreas frontomediales y frontoparietales están involucradas en la metacognición (Do Lam y cols., 2012; Vaccaro y Fleming, 2018), lo que sugiere que la función ejecutiva se ve involucrada cuando desplegamos estrategias metacognitivas (Fernandez-Duque y cols., 2000).

Tradicionalmente, los juicios de aprendizaje (JOL, por sus siglas en inglés *judgment of learning*) son uno de los procedimientos utilizados en investigación para evaluar el proceso de monitorización, pues se consideran el resultado de las estrategias metacognitivas involucradas en la monitorización del aprendizaje. En una tarea típica, los participantes deben estimar –generalmente con un porcentaje– la probabilidad de recordar en el futuro el material que acaban de estudiar (Koriat, 1997; Nelson y Leonesio, 1988; Nelson y Narens, 1990).

Los JOLs son de naturaleza inferencial y pueden estar basados en una evaluación específica del ítem (p.ej., palabras individuales, imágenes, etc.) o sobre una lista completa de ítems (p.ej., listas de palabras, textos) (Rhodes, 2015). Según la teoría de la utilización de claves de Koriat (1997), las personas combinamos información de diferentes fuentes –características del material y contexto de estudio– para emitir nuestros juicios. Por ejemplo, se ha demostrado que los JOLs son sensibles a variaciones en la dificultad del material. Algunos factores que pueden afectar a la magnitud de los JOLs son características perceptivas como el tipo de letra (e.g., Magreehan et al., 2016; Thompson y cols., 2013); características léxicas como el grado de concreción (e.g., Hertzog y cols., 2003; Tauber y Rhodes, 2012); y características semánticas como el grado de asociación de palabras (e.g., Matvey y cols., 2006; Soderstrom y McCabe, 2011); la emocionalidad (e.g., Groninger, 1976; Tauber y Dunlosky, 2012; Zimmerman y Kelley, 2010); o el grado de coherencia y elaboración de textos (p.ej., Ackerman y Goldsmith, 2011; Ariel y cols., 2020; Lefèvre y Lories, 2004; Nguyen y McDaniel, 2016; Pieger y cols., 2016; véase Prinz y cols., 2020 para un metanálisis).

Generalmente, los juicios de aprendizaje suelen ser bastante precisos en predecir el posterior recuerdo. Así, por ejemplo, las palabras concretas y relacionadas semánticamente reciben JOLs más altos, y son de hecho mejor recordadas, que las palabras abstractas y no relacionadas (Hertzog y cols., 2003; Undorf y Erdfelder, 2015;

Witherby y Tauber, 2017). No obstante, en ocasiones, dado que los JOLs son producto de un proceso inferencial, se pueden producir errores si las claves en las que nos basamos para emitir los juicios no son realmente informativas del posterior desempeño en memoria (véase Kühl y Eitel, 2016, para una discusión sobre los resultados mixtos de un efecto perceptivo tanto en JOLs como en memoria).

En general, la investigación hasta el momento parece indicar que las personas son sensibles a diferentes características al juzgar la probabilidad de recordar el material estudiado. Sin embargo, dado que la monitorización requiere esfuerzo cognitivo, es posible que al estudiar en L2, las personas tengan menos recursos cognitivos disponibles para dedicar a dichos procesos metacognitivos y de aprendizaje. Realmente, aún desconocemos si el idioma de estudio juega un papel en el proceso de monitorización y si interactúa con otras señales disponibles en el contexto de aprendizaje.

A lo largo de los cinco experimentos que componen esta tesis, investigamos las consecuencias de estudiar en L1 o L2 en la interacción entre la monitorización y el control de la memoria, así como en el posterior aprendizaje. Por tanto, nuestro objetivo es, por un lado, investigar si la manipulación de características perceptivas, léxicas, semánticas y organizativas tiene efecto en los JOLs tanto en L1 como en L2. Es decir, queremos comprender cómo se manifiestan los efectos del tipo de fuente, de la concreción, de la relación semántica y de la cohesión en el proceso de monitorización durante el estudio en ambos idiomas. Al mismo tiempo, queremos saber si los participantes adaptan su percepción general del contexto de aprendizaje en función del idioma utilizado. Es decir, si sus JOLs varían en función del idioma de estudio. Finalmente, examinamos la interacción entre estos dos efectos: el idioma y la manipulación del material que va a ser estudiado.

Los participantes estudiaban diferentes tipos de material y emitían JOLs para cada una de las palabras de estudio (Experimentos 1 y 2), listas cortas de palabras (Experimento 3) o textos breves (Experimentos 4 y 5). En los Experimentos 1-4, los participantes eran bilingües *no balancelados* y de adquisición tardía con un nivel medio de dominio de inglés-L2. En el Experimento 5, tenemos dos grupos de participantes: un grupo de mayor y otro de menor competencia en L2. Todos los participantes realizaron las tareas tanto en L1 como en L2, siendo el idioma una variable bloqueada y contrabalanceada en los Experimentos 1-4, y una variable mixta y aleatoria en el Experimento 5. A lo largo de los experimentos, manipulamos diferentes características de los materiales que podrían servir de claves para la monitorización metacognitiva: el tipo de letra (Experimento 1), la concreción (Experimento 2) y el grado de relación semántica (Experimento 3) en palabras; y el nivel de cohesión en textos (Experimento 4 y 5). Exploramos si estas manipulaciones afectaban a los JOLs y al rendimiento posterior en memoria tanto en L1 como en L2, con el objetivo de comprobar hasta qué punto las personas bilingües no balanceadas y con diferente nivel de dominio de inglés-L2, pueden monitorizar y controlar su memoria en ambos idiomas.

En general, nuestra hipótesis era que la demanda cognitiva derivada del procesamiento en L2, en comparación con L1, podrían reducir la capacidad para desplegar estrategias metacognitivas, ya que ambas tareas –procesamiento en L2 y monitorización– requieren control cognitivo. Es decir, al estudiar en L2, las personas tienen menos recursos cognitivos disponibles para dedicar a los procesos metacognitivos de aprendizaje. Así, esperamos observar un uso menos preciso de las claves disponibles para monitorizar en L2 frente a L1. Esto significa que las manipulaciones para aumentar o disminuir la dificultad intrínseca del material –concreción, relación semántica o cohesión–, podrían no detectarse cuando se estudia en L2.

Además, evaluamos si los participantes ajustaban su percepción general del aprendizaje al contexto proporcionado por el idioma. Dado que cada bloque de los Experimentos 1-4 definía un contexto de idioma (L1 vs. L2), es posible evaluar si los participantes percibían el éxito del aprendizaje de manera diferente según el idioma y si ajustaban sus JOLs en consecuencia. En general, esperábamos que los participantes consideraran que el aprendizaje en L1 fuera más fácil y exitoso, emitiendo JOLs más altos, que el aprendizaje en L2.

Los resultados parecen indicar que aprender en L2 no compromete la monitorización del aprendizaje. Los participantes, independientemente de su nivel de dominio de L2, juzgaron con precisión los materiales tanto en L1 como en L2, utilizando el idioma y otras características como claves distintivas de la dificultad. A nivel de palabras, no encontramos diferencias en los JOLs debido a la clave perceptiva. En el Experimento 1, las personas no consideraron que la tipografía *difícil de leer* fuera menos probable de recordar que la tipografía fácil de leer. Las investigaciones previas coinciden en que el efecto del tipo de fuente en JOLs se observa con menos consistencia también en L1 (Magreehan y cols., 2016), y no siempre aparece en pruebas de reconocimiento (Oppenheimer y Alter, 2014; Rummer y cols., 2016; Xie y cols., 2018). Aunque se ha afirmado que la clave perceptiva podría tener un efecto directo en los JOLs y la memoria al afectar la facilidad de procesamiento (Yue y cols., 2013), el impacto positivo del efecto perceptivo del tipo de fuente es limitado. La dificultad perceptiva podría verse moderada por diferentes factores y puede ser relevante en situaciones específicas y entre ciertos individuos. Por ejemplo, presentar materiales degradados perceptualmente mejoró la retención y la comprensión de textos expositivos para individuos con altos niveles de memoria de trabajo (como sugieren Lehmann et al., 2016; aunque Strukelj y cols., 2016 tienen puntos de vista diferentes). Además, se ha encontrado que esta degradación perceptiva tiene impacto en los JOLs (como observaron Magreehan y cols., 2016, Experimentos 4 y 5) y en el tiempo de estudio (por ejemplo, Rummer y cols., 2016). El tipo de prueba también podría tener un papel crítico en que el efecto aparezca: alterar la apariencia visual de un elemento dentro de una lista puede interferir con el procesamiento conceptual y la elaboración del ítem, efectos que tienden a observarse cuando se utiliza una prueba de recuerdo libre, pero no en una prueba de reconocimiento (Nairne, 1988). Además, parece que el tipo de fuente difícil de leer disminuye los JOL en comparación con la fuente fácil de leer, pero solo cuando ambos tipos de fuente se presentan en listas mixtas (Maxwell y cols., 2021), y cuando se solicitan JOL inmediatos en lugar de demorados (Luna y cols., 2018). Aunque la literatura presenta resultados mixtos, parece que las variaciones en el tipo de fuente tienen una influencia limitada tanto en los juicios subjetivos como en el almacenamiento real o en la retención de información en la memoria. En nuestro caso, empleamos una manipulación intra-sujeto y solicitamos JOLs inmediatos. Sin embargo, esto no evitó que obtuviéramos resultados nulos en el Experimento 1. Lo más interesante es la consistencia de los resultados observados en L2 con respecto a la clave perceptiva, reflejando el mismo patrón encontrado en L1. Por lo tanto, nuestros resultados replican los efectos nulos reportados en estudios anteriores (por ejemplo, Magreehan y cols., 2016; Oppenheimer y Alter, 2014) tanto en L1 como en L2.

Los participantes reportaron diferentes JOLs ante claves léxicas y semánticas, otorgando JOLs más bajos a palabras abstractas y no relacionadas que a palabras concretas y relacionadas respectivamente (Experimentos 2 y 3). Estos resultados están en línea con investigaciones previas: los efectos de concreción y relación semántica aparecen consistentemente en la literatura de JOLs (Matvey y cols., 2006; Witherby y Tauber, 2017). Hay varios mecanismos cognitivos que contribuyen al fenómeno por el que las palabras concretas y las palabras semánticamente relacionadas suelen recibir JOLs más elevados y, de hecho, se recuerdan mejor en las pruebas de memoria. Las palabras concretas evocan imágenes mentales, representaciones y experiencias sensoriales más vívidas en comparación con las palabras abstractas (según la Teoría de la Codificación Dual; Paivio, 1991) y tienen más asociaciones contextuales con experiencias del mundo real, lo que las hace más fáciles de relacionar y de integrar dentro de estructuras de conocimiento existentes (según la Teoría de Disponibilidad de Contexto, Schwanenflugel y cols., 1992). Las palabras semánticamente relacionadas, por otro lado, comparten conexiones en significado o contexto, lo que lleva a una red de asociaciones (Collins y Loftus, 1975). Esta interconexión ayuda en los procesos de memoria, ya que la activación de una palabra o concepto puede desencadenar la activación de información relacionada, facilitando la recuperación en memoria (por ejemplo, Buchler y cols., 2008; Desaunay y cols., 2017; Neely y Tse, 2011). Los resultados de nuestros estudios confirman que esto también es cierto para el procesamiento en L2: los participantes fueron sensibles a estas diferencias en ambos idiomas, L1 y L2. Esto tiene implicaciones significativas, ya que se había hipotetizado que la activación semántica en L2 de personas bilingües podría ser más débil y que el acceso léxico en L2 está mediado por L1 y, por lo tanto, es más indirecto (Kroll y Stewart, 1994). Investigaciones anteriores indicaron un retraso en el acceso léxico durante el procesamiento en L2 (Opitz & Degner, 2012) o una activación más débil del área cerebral relacionada con la integración semántica, lo que indica un patrón de activación diferente de las regiones cerebrales durante el procesamiento en L2 (Zhang y cols., 2020). A pesar de esto, nuestros participantes monitorizaron correctamente los materiales en L2, identificando las claves de concreción y relación semántica.

A nivel de textos, las personas juzgaron los textos de alta cohesión como más fáciles de aprender que los textos de baja cohesión. El efecto de cohesión en textos también va en línea con estudios previos sobre evaluación y comprensión de textos en monolingües (Carroll y Korukina, 1999; Lefèvre y Lories, 2004; Rawson y Dunlosky, 2002). Replicamos estos hallazgos con una población bilingüe, lo que resalta la solidez del efecto de cohesión en diversos contextos lingüísticos y la buena resolución metacognitiva de los participantes. Investigaciones previas han señalado que los procesos de lectura son diferentes en L1 y L2 (por ejemplo, Cop y cols., 2015; Dirix y cols., 2020; Pérez y cols., 2018; Whitford y Titone, 2012). Por ejemplo, Whitford y Titone (2012) observaron mayores tiempos de fijación y mayores tiempos de lectura en oraciones en L2; y Cop y cols. (2015) sugirieron que los tiempos de lectura más lentos en L2 se debían a un mayor número de fijaciones, que eran más largas y más cercanas entre sí, y al hecho de que se omitían menos palabras en la lectura en L2 en comparación con la lectura en L1. Además, Pérez y cols. (2018) demostraron que los procesos cognitivos de alto nivel, como la revisión inferencial, eran menos eficientes en L2 en comparación con L1. Por lo tanto, podríamos haber esperado que la monitorización metacognitiva durante la lectura estuviera afectada en L2, ya que el proceso de lectura es menos automático y más exigente cuando se produce en un segundo idioma. Sin embargo, parece que la forma en que los individuos evalúan y perciben la coherencia del texto y la dificultad de aprendizaje es independiente del idioma. Los resultados también respaldan la idea de que la influencia de la cohesión en la evaluación del aprendizaje trasciende las diferencias lingüísticas y se mantiene consistente entre diversos contextos lingüísticos. Desde el marco del aprendizaje autorregulado, se asume que los procesos metacognitivos son exigentes y cognitivamente demandantes (Efklides, 2014; Stine-Morrow y cols., 2006; Tauber y Witherby, 2019). No obstante, nuestros datos sugieren que estos procesos no están afectados en un contexto de aprendizaje de L2. De hecho, la monitorización metacognitiva fue tan efectiva en L2 como en L1. Es posible que el control cognitivo dedicado a manejar las demandas de L2 y facilitar los procesos metacognitivos se apoyaran mutuamente en lugar de competir por recursos. Y así, las claves diagnósticas fueron monitorizadas con éxito. Sin embargo, debemos tener cautela con nuestras suposiciones ya que presentamos solo datos conductuales, los cuales tienen limitaciones evidentes en términos de proporcionar una visión de los mecanismos cognitivos subyacentes.

Con respecto al idioma de estudio, los participantes consideraron que estudiar en L2 era más difícil (es decir, dieron JOLs más bajos) que estudiar en L1. Sin embargo, este efecto varió en función de los experimentos. En los Experimentos 1 y 2, donde los JOLs involucraban palabras individuales, este patrón solo fue evidente cuando el bloque de L2 aparecía en primer lugar. En cambio, en el Experimento 3, donde los JOLs implicaban la evaluación de listas cortas de palabras, el efecto del idioma apareció independientemente del orden del bloque. Es posible que al considerar listas completas (Experimento 3), los estudiantes no solo hayan activado representaciones mentales de las palabras, sino que también hayan accedido a representaciones de palabras asociadas. Se ha demostrado que los vínculos de memoria entre palabras y representaciones conceptuales son más fuertes en L1 que en L2 (Kroll y Stewart, 1994), lo que a su vez podría manifestarse en diferentes juicios para L1 y L2. Las representaciones semántico-relacionales para ambos idiomas serían independientes de si el bloque de L2 se presentó primero o segundo, y, por lo tanto, el efecto del idioma en el Experimento 3 no interactúa con el orden de los bloques. En cambio, cuando la tarea involucraba palabras individuales (Experimentos 1-2) y no se requería un procesamiento semántico-relacional adicional, los JOLs dependían de la comparación entre los dos bloques; por lo tanto, los participantes consideraron que L1 era más fácil (JOLs más altos) después de haber estudiado primero la condición de idioma L2, que les resultó más difícil. En ese caso, parece que los participantes usaron el primer bloque como punto de referencia para la comparación y consideraron que L1 era más fácil cuando se contrastaba con el bloque de L2 más difícil. En línea con muchos estudios previos sobre bilingüismo (Beatty-Martínez y cols., 2020, 2021; Beatty-Martínez y Dussias, 2017), estos resultados proporcionan evidencia de que los efectos de L2 dependen del contexto en el que se realiza el aprendizaje.

Asimismo, las personas juzgaron los textos en L2 como más difíciles que los textos en L1 cuando el idioma estaba bloqueado (Experimento 4) y también cuando el idioma se alternaba en la fase de estudio (Experimento 5). Este efecto de idioma en JOLs era independiente del efecto de cohesión en el Experimento 4, pero estaba modulado por la interacción en el Experimento 5, de tal manera que no hubo diferencia en L1 y L2 para textos de alta cohesión, pero para textos de baja cohesión, la condición de L2 recibió JOLs significativamente más bajos. Esto sugiere que los participantes consideraron los textos de baja cohesión en L2 como la condición más difícil entre las cuatro. Esta interacción no apareció cuando el idioma estaba bloqueado (Experimento 4) frente a cuando era una variable mixta y aleatoria (Experimento 5). Estudios previos respaldan la idea de que reconocer e integrar un código lingüístico diferente implica un coste de procesamiento y que el *code-mixing* (alternancia de idiomas) requiere un alto nivel de competencia y control cognitivo en ambos idiomas (por ejemplo, Costa y Santesteban, 2004; Grainger y Beauvillain, 1988; Grainger y O'Regan, 1992; Green y Wei, 2014; Johns y cols., 2019; Moreno y cols., 2002). Los costes del *code-switching* (cambio de idioma) generalmente surgen debido a las demandas cognitivas requeridas para navegar entre los dos idiomas. Cuando los individuos bilingües se ven obligados a cambiar entre idiomas dentro de un contexto, necesitan gestionar diferentes sistemas lingüísticos, seleccionar palabras o estructuras apropiadas e inhibir la interferencia del idioma no relevante. Este proceso demanda recursos cognitivos y la necesidad de manejar y controlar simultáneamente dos sistemas lingüísticos puede tener un impacto significativo en la eficiencia del uso del lenguaje (Costa y Santesteban, 2004; Green y Wei, 2014). En el Experimento 5, la alternancia de idiomas en la presentación de textos pudo haber aumentado la relevancia del idioma como una clave distintiva, debido a la necesidad de control lingüístico derivado del cambio, a expensas de otras claves diagnósticas dentro del texto, como la cohesión. Es posible que los participantes dedicaran más recursos cognitivos a la tarea debido a la necesidad de control del lenguaje, y como consecuencia, la clave de cohesión fuera más destacada bajo la condición de idioma más exigente (L2), lo que explicaría la interacción entre los dos factores (cohesión x lenguaje).

En general, los resultados en memoria corroboran el patrón encontrado en JOLs para los cinco experimentos. Efectivamente, los participantes recordaron las palabras de forma similar independientemente del tipo de fuente (Experimento 1). Sin embargo, recordaron mejor las palabras concretas (frente a las abstractas, Experimento 2), las palabras relacionadas semánticamente (frente a las palabras no relacionadas, Experimento 3) y los textos de alta cohesión (frente a los de baja cohesión, Experimento 4 y 5), tal y como habían estimado en la fase de estudio con los JOLs. Por lo tanto, el patrón de resultados más destacado es que el idioma no obstaculizó la monitorización en ninguna de nuestras manipulaciones; la monitorización del aprendizaje se realizó de manera similar en L1 y en L2. Incluso cuando los participantes consideraron que estudiar en L2 era más difícil, esta dificultad no impidió el uso de procesos de monitorización que les permitieron evaluar y juzgar el material con precisión. El hecho de que las personas bilingües pudieran evaluar igualmente su grado de aprendizaje para características como la concreción, la relación semántica y la cohesión en L1 y L2 tiene claras implicaciones prácticas, ya que sugiere que, al menos para bilingües con un nivel de competencia intermedio, la monitorización y el control del aprendizaje no se ven afectados como consecuencia del contexto de L2.

Curiosamente, se observó un mecanismo compensatorio en los Experimentos 2 y 3, donde los materiales en L2 se percibieron como más difíciles (es decir, recibieron JOLs más bajos) pero se recordaron mejor. Parece que, con materiales sencillos, los participantes fueron capaces de compensar la dificultad de L2 y las palabras en L2 obtuvieron mejor índice de recuerdo que en L1. A pesar de que esta compensación en L2 apareció para materiales simples (palabras), los participantes no fueron capaces de compensar para conseguir un aprendizaje similar a L1 cuando la complejidad del material aumentó: los textos en L2 se recordaron peor que los textos en L1, especialmente cuando el nivel de competencia del idioma no era alto. La compensación tampoco apareció para las manipulaciones léxico-semánticas y semántico-organizativas: las palabras concretas, las listas relacionadas y los textos de alta cohesión mostraron un mejor recuerdo en comparación con las palabras abstractas, las listas no relacionadas y los textos de baja cohesión. Por lo tanto, aunque los JOLs de los participantes fueron sensibles a la dificultad objetiva de los materiales, no utilizaron espontáneamente este conocimiento para compensar las palabras abstractas y no relacionadas, y así lograr el mismo nivel de aprendizaje que con los materiales más fáciles.

El efecto de la cohesión, sin embargo, mostró un patrón interesante en el que los resultados diferían entre idiomas. Las personas en el Experimento 4 no compensaron la dificultad detectada en la fase de estudio para lograr un aprendizaje comparable en textos de baja cohesión (frente a los de alta cohesión). De manera similar, en el Experimento 5, las personas con menor competencia no mostraron efectos de compensación a pesar de que también detectaron la dificultad de los textos de baja cohesión. La única condición en la que el aprendizaje para los textos de baja y alta cohesión fue similar fue cuando las personas con mayor competencia estudiaron en L1. Sin embargo, el efecto de cohesión también estaba ausente en los JOLs para estas personas. Esto sugiere que, para individuos con mayor competencia en L2, la falta de efectos de cohesión en la prueba de recuerdo no se debió a la compensación, sino al hecho de que todos los textos, independientemente de su grado de cohesión, produjeron niveles similares de dificultad para ellos. Como ya se mencionó, esto podría deberse a la alta demanda de procesos de control cuando L1 y L2 se presentan en un formato mixto, lo que pudo inducir un mejor aprendizaje en L1 (por ejemplo, Costa y Santesteban, 2004; Green y Wei, 2014). Además, los textos de baja cohesión a menudo requieren que los lectores hagan inferencias debido a la falta de conexiones explícitas entre oraciones o párrafos. Estos textos contienen contenido más ambiguo que requiere que los lectores empleen funciones cognitivas superiores, como el razonamiento, la anticipación de información y la inferencia, para discernir el significado implícito. Estos procesos requieren un mayor esfuerzo mental por parte del lector, ya que se deben construir conexiones activamente para comprender el texto de manera efectiva. Esto podría no haber sido posible en L2, ya que el procesamiento en un segundo idioma recluta una cantidad significativa de recursos cognitivos, reduciendo su disponibilidad para hacer las inferencias necesarias en la comprensión de textos de baja cohesión. Esta perspectiva es respaldada por el trabajo desarrollado por Pérez y cols. (2018), quienes demostraron que la revisión inferencial fue menos eficiente en L2 en comparación con L1.

La ausencia de evidencia de compensación en las condiciones de mayor complejidad de nuestra investigación, concuerda con estudios previos que indican que las personas no compensan los efectos de la dificultad a través de la autorregulación. Se podría esperar que, si los estudiantes detectan dificultades en los materiales de aprendizaje, lleven a cabo mecanismos compensatorios asignando más tiempo o seleccionando una estrategia diferente para aprender mejor la información. Sin embargo, existen evidencias sólidas de que ni el estudio autorregulado, ni la selección de ítems para volver a estudiarlos, ni el uso de estrategias (por ejemplo, estudio distribuido) compensan completamente la dificultad (Cull y Zechmeister, 1994; Koriat, 2008; Koriat y cols., 2006; Koriat y Ma'ayan, 2005; Le Ny y cols., 1972; Mazzoni y cols., 1990; Mazzoni y Cornoldi, 1993; Nelson y Leonesio, 1988; Pelegrina y cols., 2000; ver Tekin, 2022 para una revisión).

Algo destacable y de gran importancia es que el nivel de competencia en L2 parece introducir matices en los procesos de aprendizaje que podrían afectar a los recursos disponibles para el procesamiento metacognitivo. En nuestros estudios encontramos diferencias en la prueba de compresión entre los grupos de nivel de L2. Las personas con un nivel de competencia más alto consideraron que los textos en L1 y L2 eran igualmente fáciles de aprender y, de hecho, no mostraron ninguna desventaja de comprensión en L2. Sin embargo, el grupo de competencia más baja consideró que los textos en L2 eran más difíciles durante el estudio y, de hecho, mostraron una desventaja en el aprendizaje. Aunque parece que estos participantes eran capaces de emplear los procesos metacognitivos necesarios para identificar la dificultad de los textos, ya que evaluaron correctamente su dificultad, no lograron compensar dichas dificultades. El hecho de que los participantes de menor competencia L2 obtuvieran peores resultados de aprendizaje en comparación con los de mayor competencia va en línea con hallazgos de investigaciones anteriores. Vander Beken y cols. (2018) informaron sobre un coste en L2 cuando los participantes debían responder con preguntas de tipo ensayo, atribuido presumiblemente a la falta de habilidades de escritura. En nuestro caso, quisimos evitar este problema al incluir preguntas abiertas que no requerían mucha elaboración, lo que podría evitar el déficit de producción. Se podría discutir que ambos grupos de nivel de L2 diferían en sus habilidades de lectura y escritura, es decir, que el grupo de menor competencia estaba compuesto por lectores no menos avanzados o inexpertos. Sin embargo, ambos grupos mostraron una precisión similar para los textos en L1, lo que demuestra que las diferencias en la prueba de evaluación del aprendizaje no pueden ser explicadas porque el grupo de mayor competencia fuera mejor lector o “comprendedor”. Varios factores podrían explicar el coste de recuerdo en L2 en el grupo de menor competencia: peor codificación o mayor dificultad para integrar la información o simplemente para recuperarla de la memoria. Las investigaciones futuras deberían incorporar diferentes formatos de pruebas y permitir la lectura autorregulada para abordar estos problemas. De hecho, algunos estudios han afirmado que proporcionar el suficiente tiempo de estudio podría ser un factor crucial para aprender contenido en L2 de manera efectiva (Vander Beken y cols., 2020). Eventualmente, esto podría convertir el estudio en L2 en una "dificultad deseable" (por ejemplo, Metcalfe, 2011), donde las condiciones de aprendizaje que suponen un mayor desafío contribuyen a la formación de trazos de memoria más fuertes.

Finalmente, los resultados del cuestionario metacognitivo sugirieron que los participantes emplearon diferentes estrategias de aprendizaje al estudiar en L1 y en L2. Esto podría respaldar la idea de que afrontan el aprendizaje en L2 de forma distinta al aprendizaje en L1. En general, al estudiar en L1, las personas usaban más estrategias de aprendizaje a un nivel profundo que al estudiar en L2. Se considera que las estrategias a un nivel profundo tienen un mayor impacto en el aprendizaje a largo plazo (Deekens y cols., 2018; Lonka y cols., 2004; Vermunt y Vermetten, 2004). Aunque nuestros participantes monitorizaron correctamente en ambos idiomas, parece que el procesamiento en L2 es notablemente más exigente y ocurre bajo un sistema cognitivo potencialmente sobrecargado (Adesope y cols., 2010; Hessel y Schroeder, 2020, 2022; Pérez y cols., 2018), lo que permite menos oportunidades para emplear estrategias de aprendizaje profundo. Sin embargo, en el Experimento 5, ambos grupos de nivel de L2 mostraron una mayor frecuencia de uso de autorregulación metacognitiva en L2 en comparación con L1, lo que contradice esta hipótesis. Esto sugiere que las personas posiblemente tenían recursos cognitivos suficientes y pudieron utilizarlos para adoptar estrategias de aprendizaje eficientes incluso al estudiar en L2. Es posible que la adopción de ciertas estrategias pudiera requerir tiempo adicional de estudio que los participantes no podrían emplear dadas las limitaciones de tiempo de estudio de nuestros experimentos. En consecuencia, puede que las personas hayan empleado estrategias a un nivel profundo con más frecuencia en L1, ya que podrían haber necesitado más tiempo de estudio para aplicarlas en el aprendizaje en L2 (Stoff & Eagle, 1971). Sin embargo, la diferencia cualitativa en el uso de estrategias metacognitivas específicas del idioma es un aspecto que se debe explorar en futuras investigaciones.

En resumen, la monitorización metacognitiva es un proceso complejo del aprendizaje autorregulado que al parecer no se ve obstaculizado en un contexto de L2. Hasta ahora, hemos explorado el proceso de monitorización y las consecuencias a corto plazo en memoria derivadas del estudio en L2. A medida que se desarrolla el aprendizaje autorregulado, surgen procesos de monitorización y control para evaluar y ajustar la atención, la comprensión y el comportamiento. Las personas estiman la dificultad de la tarea, hacen juicios de aprendizaje y asignan recursos en consecuencia, ajustando por ejemplo, el ritmo de aprendizaje o regulando el uso de estrategias, por ejemplo (Panadero, 2017). En general, en relación con nuestra pregunta principal de si estudiar en L2 compromete la monitorización metacognitiva, los resultados parecen sugerir lo contrario. Aparentemente, los participantes monitorizaron correctamente el material de estudio. Sin embargo, las personas no pudieron compensar la dificultad auto-percibida para lograr un aprendizaje exitoso en ninguna de las manipulaciones (perceptiva, léxica, semántica y organizativa), tal y como lo demuestra el rendimiento en las pruebas de memoria. Tampoco pudieron sobreponerse a la dificultad del idioma con materiales más complejos (textos). Sí pudieron hacerlo con materiales relativamente sencillos – palabras individuales y listas cortas–, y L2 obtuvo un rendimiento superior que L1 para este tipo de material.

Nuestros resultados sugieren que la complejidad y las características del material de estudio podrían ser un aspecto central de los procesos metacognitivos de aprendizaje en L2. Las manipulaciones perceptivas pueden no tener un impacto en el acceso semántico de la palabra ni en su representación mental, ni interferir en cómo las personas monitorizan su aprendizaje. Las listas de palabras concretas y abstractas, las listas de palabras relacionadas y no relacionadas, así como los textos, requirieron un procesamiento más profundo, y el idioma juega un papel tanto en la monitorización como en el aprendizaje. La evidencia presentada aquí sugiere la idea de que el aprendizaje en L2 conduce a expectativas de desempeño en la tarea más bajas, pero que finalmente resulta en un rendimiento ajustado al nivel de dominio en L2. Por lo tanto, esto parece indicar que la instrucción en L2 no compromete los procesos de monitorización metacognitiva. A pesar de ello, los resultados del Experimento 5 plantean una advertencia relevante sobre el aprendizaje en L2 en entornos académicos. El nivel de competencia en L2 parece introducir algunas sutilezas en los procesos de aprendizaje que pueden afectar los recursos disponibles para el procesamiento metacognitivo durante las actividades de estudio.

No debemos olvidar que los datos presentados aquí son puramente conductuales, y que las asunciones que se puedan derivar de ellos son limitadas. Investigaciones futuras deberían incorporar técnicas electrofisiológicas como *eye-tracking* para explorar los mecanismos subyacentes. Además, sería interesante identificar los efectos directos e indirectos de la competencia lingüística y las estrategias de aprendizaje en los resultados académicos de la educación bilingüe. También, existen otros factores cruciales que podrían desempeñar un papel en la autorregulación del aprendizaje (Lonka y cols., 2004). La dificultad del material y la competencia lingüística podrían no ser los únicos aspectos que guían el aprendizaje autorregulado: las funciones ejecutivas, la motivación, la autoeficacia y otras demandas de la tarea o del contexto también pueden influir en el proceso de aprendizaje. En resumen, desarrollar un marco que incorpore estos componentes metacognitivos y motivacionales del aprendizaje autorregulado resulta crucial para explicar los efectos del aprendizaje en L2 sobre la monitorización y la memoria. Esto abre nuevas direcciones para guiar investigaciones futuras, enriqueciendo así nuestra comprensión de las estrategias metacognitivas en entornos de aprendizaje en L2.

References

Ackerman, R., & Goldsmith, M. (2011). Metacognitive Regulation of Text Learning: On Screen Versus on Paper. *Journal of Experimental Psychology: Applied*, *17*(1), 18–32. https://doi.org/10.1037/a0022086

Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A Systematic Review and Meta-Analysis of the Cognitive Correlates of Bilingualism. *Review of Educational Research*, *80*(2), 207–245. https://doi.org/10.3102/0034654310368803

Adi-Japha, E., Berberich-Artzi, J., & Libnawi, A. (2010). Cognitive flexibility in drawings of bilingual children. *Child Development*, *81*(5), 1356–1366.

Alptekin, C., & Erçetin, G. (2010). The role of L1 and L2 working memory in literal and inferential comprehension in L2 reading. *Journal of Research in Reading*, *33*(2), 206–219. https://doi.org/10.1111/j.1467-9817.2009.01412.x

Alptekin, C., & Erçetin, G. (2011). Effects of working memory capacity and content familiarity on literal and inferential comprehension in l2 reading. *TESOL Quarterly*, *45*(2), 235–266. https://doi.org/10.5054/tq.2011.247705

Altarriba, J., & Bauer, L. M. (2004). The distinctiveness of emotion concepts: A comparison between emotion, abstract, and concrete words. *American Journal of Psychology*, *117*, 389–410.

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology*, *136*(4), 569–576.

Anwyl-Irvine, A., Dalmaijer, E. S., Hodges, N., & Evershed, J. K. (2021). Realistic precision and accuracy of online experiment platforms, web browsers, and devices. *Behavior Research Methods*, *53*(4), 1407–1425. https://doi.org/10.3758/s13428-020-01501-5

Anwyl-Irvine, A., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Ariel, R., Karpicke, J. D., Witherby, A. E., & Tauber, S. K. (2020). Do Judgments of Learning Directly Enhance Learning of Educational Materials? *Educational Psychology Review*. https://doi.org/10.1007/s10648-020-09556-8

Arndt, J., & Beato, M. S. (2017). The role of language proficiency in producing false memories. *Journal of Memory and Language*, *95*, 146–158. https://doi.org/10.1016/j.jml.2017.03.004

Association, W. M. (2013). World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. *JAMA*, *310*(20), 2191–2194. https://doi.org/10.1001/jama.2013.281053

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database (CD-ROM).*

Bannert, M. (2002). Managing cognitive load - Recent trends in cognitive load theory. *Learning and Instruction*, *12*(1), 139–146.

Beatty-Martínez, A. L., & Dussias, P. E. (2017). Bilingual experience shapes language processing: Evidence from codeswitching. *Journal of Memory and Language*, *95*, 173–189. https://doi.org/10.1016/j.jml.2017.04.002

Beatty-Martínez, A. L., Navarro-Torres, C. A., Dussias, P. E., Bajo, M. T., Guzzardo Tamargo, R. E., & Kroll, J. F. (2020). Interactional context mediates the consequences of bilingualism for language and cognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*, *46*(6), 1022–1047. https://doi.org/https://doi.org/10.1037/xlm0000770

Beatty-Martínez, A. L., Tamargo, R. E. G., & Dussias, P. E. (2021). Phasic pupillary responses reveal differential engagement of attentional control in bilingual spoken language processing.

Scientific Reports, 1–12. https://doi.org/10.1038/s41598-021-03008-1

Begg, I., Duft, S., Lalonde, P., Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory and Language*, *28*(5), 610–632. https://www.unhcr.org/publications/manuals/4d9352319/unhcr-protection-training-manual-european-border-entry-officials-2-legal.html?query=excom 1989

Bergsleithner, J. M. (2010). Working memory capacity and L2 writing performance. *Ciências & Cognição*, *15*(2), 2–20.

Best, R. M., Rowe, M., Ozuru, Y., & McNamara, D. S. (2005). Deep-Level Comprehension of Science Texts: The Role of the Reader and the Text. *Topics in Language Disorders*, *25*(1), 65–83.

Bialystok, E. (2001). *Bilingualism in development: Language, literacy, and cognition.* Cambridge University Press. https://doi.org/https://doi.org/10.1017/CBO9780511605963

Bialystok, E. (2017). The bilingual adaptation: How minds accommodate experience. *Psychological Bulletin*, *143*, 233–262. https://doi.org/doi:10.1037/bul0000099

Bialystok, E., Abutalebi, J., Bak, T. H., Burke, D. M., & Kroll, J. F. (2016). Aging in two languages: Implications for public health. *Ageing Research Reviews*, *27*, 56–60. https://doi.org/10.1016/j.arr.2016.03.003

Bialystok, E., Craik, F. I., & Freedman, M. (2007). Bilingualism as a protection against the onset of symptoms of dementia. *Neuropsychologia*, *45*(2), 459–464. https://doi.org/10.1016/j.neuropsychologia.2006.10.009.

Bialystok, E., & Craik, F. I. M. (2022). How does bilingualism modify cognitive function? Attention to the mechanism. *Psychonomic Bulletin & Review*, *0123456789*. https://doi.org/10.3758/s13423-022-02057-5

Bialystok, E., Craik, F. I. M., & Luk, G. (2012). Bilingualism: Consequences for mind and brain. *Trends in Cognitive Sciences*, *16*(4), 240–249. https://doi.org/10.1016/j.tics.2012.03.001

Bialystok, E., & Shapero, D. (2005). Ambiguous benefits: the effect of bilingualism on reversing ambiguous figures. *Developmental Science*, *8*(6), 595–604. https://doi.org/https://doi.org/10.1111/j.1467-7687.2005.00451.x

Bjork, R. A., & Bjork, E. L. (2020). Desirable difficulties in theory and practice. *Journal of Applied Research in Memory and Cognition*, *9*(4), 475–479.

Blake, A. B., & Castel, A. D. (2018). On belief and fluency in the construction of judgments of learning: Assessing and altering the direct effects of belief. *Acta Psychologica*, *186*, 27–38. https://doi.org/10.1016/j.actpsy.2018.04.004

Blanco-Elorrieta, E., & Pylkkänen, L. (2017). Bilingual language switching in the laboratory versus in the wild: The spatiotemporal dynamics of adaptive language control. *Journal of Neuroscience*, *37*(37), 9022–9036. https://doi.org/10.1523/JNEUROSCI.0553-17.2017

Boekaerts, M. (2017). Cognitive load and self-regulation: Attempts to build a bridge. *Learning and Instruction*, *51*, 90–97. https://doi.org/10.1016/j.learninstruc.2017.07.001

Bousfield, W. A., & Cohen, B. H. (1956). Clustering in recall as a function of the number of word-categories in stimulus-word lists. *The Journal of General Psychology*, *54*(1), 95–106.

Broekkamp, H., & Van Hout-Wolters, B. H. A. M. (2007). Students' adaptation of study strategies when preparing for classroom tests. *Educational Psychology Review*, *19*(4), 401–428. https://doi.org/10.1007 /s10648-006-9025-0.

Brookshear, J. G. (2005). *Computer Science. An overview.* (9th ed.). Pearson Internation Edition.

Brouwer, S. (2021). The interplay between emotion and modality in the Foreign-Language effect on moral decision making. *Bilingualism*, *24*(2), 223–230. https://doi.org/10.1017/S136672892000022X

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904–911. https://doi.org/10.3758/s13428-013-0403-5

Buchler, N. G., Light, L. L., & Reder, L. M. (2008). Memory for items and associations: Distinct representations and processes in associative recognition. *Journal of Memory and Language*, *59*(2), 183–199.

Buehler, F. J., van Loon, M. H., Bayard, N. S., Steiner, M., & Roebers, C. M. (2021). Comparing metacognitive monitoring between native and non-native speaking primary school students. *Metacognition and Learning*, *16*(3), 749–768. https://doi.org/10.1007/s11409-021-09261-z

Byun, K., Chu, H., Kim, M., Park, I., Kim, S., & Jung, J. (2011). English-medium teaching in Korean higher education: policy debates and reality. *Higher Education*, *62*(4), 431–449. https://doi.org/10.1007/s10734-010-9397-4

Calabria, M., Costa, A., Green, D. W., & Abutalebi, J. (2018). Neural basis of bilingual language control. *Annals of the New York Academy of Sciences*, *1426*(1), 221–235.

Campbell, R., & Sais, E. (1995). Accelerated Metalinguistic (Phonological) Awareness In Bilingual-Children. *British Journal Of Developmental Psychology*, *13*, 61–68. https://doi.org/10.1111/j.2044-835X.1995.tb00664.x

Cano, F. (2007). Approaches to learning and study orchestrations in high school students. *European Journal of Psychology of Education*, *22*(2), 131–151.

Carroll, M., & Korukina, S. (1999). The Effect of Text Coherence and Modality on Metamemory Judgements. *Memory*, *7*(3), 309–322. https://doi.org/10.1080/096582199387940

Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest, 19*(1), 5–51. https://doi.org/https:// doi. org/ 10. 1177/ 15291 00618 772271

Cenoz, J. (2003). The Additive Effect of Bilingualism on Third Language Acquisition: A Review. *International Journal of Bilingualism - INT J BILING*, *7*, 71–87. https://doi.org/10.1177/13670069030070010501

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*, 354–380.

Chabal, S., & Marian, V. (2015). Speakers of different languages process the visual world differently. *Journal of Experimental Psychology: General*, *144*(3), 539–550. https://doi.org/https://doi.org/10.1037/xge0000075

Chen, P., Bobb, S. C., Hoshino, N., & Marian, V. (2017). Neural signatures of language co-activation and control in bilingual spoken word comprehension. *Brain Research*, *1665*, 50–64. https://doi.org/10.1016/j.brainres.2017.03.023.Neural

Chen, P., Chung-Fat-yim, A., & Marian, V. (2022). Cultural Experience Influences Multisensory Emotion Perception in Bilinguals. *Languages*, *7*(1), 1–17. https://doi.org/10.3390/languages7010012

Christoffels, I. K., Firk, C., & Schiller, N. O. (2007). Bilingual language control: An event-related brain potential study. *Brain Research*, *1147*(1), 192–208. https://doi.org/10.1016/j.brainres.2007.01.137

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428. https://doi.org/10.1037/0033-295X.82.6.407

Collins, V. L., Dickson, S. V, Simmons, D. C., & Kameenui, E. J. (1996). Metacognition and its relation to reading comprehension: A synthesis of the research. *Retrieved from: Idea. Uoregon. Edu/~ Ncite/Documents/Techrep/Tech23. Html*.

Connor, U. (1984). Recall of text: Differences between first and second language readers. *TESOL Quarterly*, *18*(2), 239– 256. https://doi.org/doi.org/10.2307/3586692

Contemori, C., & Dussias, P. E. (2016). Referential choice in a second language: evidence for a listener-oriented approach. *Language, Cognition and Neuroscience*, *31*(10), 1257–1272. https://doi.org/10.1080/23273798.2016.1220604

Conway, A. R. A., Kane, M. J., & Al, C. E. T. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*(5), 769–786. https://doi.org/10.1021/cen-v020n021.p1399

Cop, U., Drieghe, D., & Duyck, W. (2015). Eye movement patterns in natural reading: A comparison of monolingual and bilingual reading of a novel. *PLoS One*, *10*(8), e0134008. https://doi.org/10.1371/journ al.pone.0134008

Costa, A., & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language*, *50*(4), 491–511. https://doi.org/10.1016/j.jml.2004.02.002

Crossley, S. A., Kyle, K., & McNamara, D. S. (2016). The development and use of cohesive devices in L2 writing and their relations to judgments of essay quality. *Journal of Second Language Writing*, *32*, 1–16. https://doi.org/10.1016/j.jslw.2016.01.003

Crossley, S. A., Yang, H. S., & McNamara, D. S. (2014). What's so simple about simplified texts? A computational and psycholinguistic investigation of text comprehension and text processing. *Reading in a Foreign Language*, *26*(1), 92–113.

Cull, W. L., & Zechmeister, E. B. (1994). The learning ability paradox in adult metamemory research: Where are the metamemory differences between good and poor learners? *Memory & Cognition*, *22*(2), 249–257.

Dafouz, E., & Camacho-Miñano, M. M. (2016). Exploring the impact of English-medium instruction on university student academic achievement: The case of accounting. *English for Specific Purposes*, *44*, 57–67. https://doi.org/10.1016/j.esp.2016.06.001

Davis, C. J. (2005). N-watch: A program for deriving neighborhood size and other psycholinguistic statistics. *Behavior Research Methods*, *37*(1), 65–70. https://doi.org/10.3758/BF03206399

Davis, C. J., & Perea, M. (2005). BuscaPalabras: A program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in Spanish. *Behavior Research Methods*, *37*(1), 65–70. https://doi.org/10.3758/BF03192738

De Bruin, A., Roelofs, A., Dijkstra, T., & FitzPatrick, I. (2014). Domain-general inhibition areas of the brain are involved in language switching: FMRI evidence from trilingual speakers. *NeuroImage*, *90*, 348–359. https://doi.org/10.1016/j.neuroimage.2013.12.049

De Larios, J. R., Marín, J., & Murphy, L. (2001). A temporal analysis of formulation processes in L1 and L2 writing. *Language Learning*, *51*(3), 497–538. https://doi.org/10.1111/0023-8333.00163

Declerck, M., Grainger, J., Koch, I., & Philipp, A. M. (2017). Is language control just a form of executive control? Evidence for overlapping processes in language switching and task switching. *Journal of Memory and Language*, *95*, 138–145. https://doi.org/10.1016/j.jml.2017.03.005

Declerck, M., & Koch, I. (2023). The concept of inhibition in bilingual control. *Psychological Review*, *130*(4), 953–976. https://doi.org/https://doi.org/10.1037/rev0000367

Declerck, M., & Philipp, A. M. (2015). A review of control processes and their locus in language switching. *Psychonomic Bulletin and Review*, *22*(6), 1630–1645. https://doi.org/10.3758/s13423-015-0836-1

Deekens, V. M., Greene, J. A., & Lobczowski, N. G. (2018). Monitoring and depth of strategy use in computer-based learning environments for science and history. *British Journal of Educational Psychology*, *88*(1), 63–79. https://doi.org/10.1111/bjep.12174

Denney, N. W. (1974). Clustering in middle and old age. *Developmental Psychology*, *10*(4), 471–475. https://doi.org/https://doi.org/10.1037/h0036604

Desaunay, P., Clochon, P., Doidy, F., Lambrechts, A., Bowler, D. M., Gérardin, P., Baleyte, J.-M., Eustache, F., & Guillery-Girard, B. (2017). Impact of Semantic Relatedness on Associative Memory: An ERP Study. *Frontiers in Human Neuroscience*, *11*. https://doi.org/10.3389/fnhum.2017.00335

Dhaene, S., & Woumans, E. (2022). Text recall and use of advance organisers in first and second language. *Studies in Second Language Acquisition*, 1–12. https://doi.org/DOI: 10.1017/S0272263122000109

Diemand-Yauman, C., Oppenheimer, D. M., & Vaughan, E. B. (2011). Fortune favors the bold and the italicized: Effects of disfluency on educational outcomes. *Cognition*, *118*(1), 111–115. https://doi.org/10.1016/j.cognition.2010.09.012

Dijkstra, T., & Kroll, J. F. (2005). Bilingual visual word recognition and lexical access. In *Handbook of bilingualism: Psycholinguistic* *approachesPsycholinguistic Approaches* (pp. 178–201). Oxford University Press.

Dijkstra, T., & van Heuven, W. J. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, *5*(3), 175–197. https://doi.org/10.1017/S1366 72890 2003012

Dinsmore, D. L., & Zoellner, B. P. (2018). The relation between cognitive and metacognitive strategic processing during a science simulation. *British Journal of Educational Psychology*, *88*(1), 95–117. https://doi.org/10.1111/bjep.12177

Dirix, N., Vander Beken, H., De Bruyne, E., Brysbaert, M., & Duyck, W. (2020). Reading Text When Studying in a Second Language: An Eye-Tracking Study. *Reading Research Quarterly*, *55*(0), 371–397. https://doi.org/10.1002/rrq.277

Do Lam, A. T. A., Axmacher, N., Fell, J., Staresina, B. P., Gauggel, S., Wagner, T., Olligs, J., & Weis, S. (2012). Monitoring the mind: The neurocognitive correlates of metamemory. *PLoS ONE*, *7*(1), 1–10. https://doi.org/10.1371/journal.pone.0030009

Dufort, B., Erickson, S., Hamilton, M., Soderquist, D., & Zigray, S. (2018). *World Geography*. (version 1.). Michigan Open Book Project. https://doi.org/10.4135/9781412972024.n1099

Duncan, S. E. (2005). Child bilingualism and cognitive functioning: A study of four Hispanic groups. *Dissertation Abstracts International, A:The Humanities and Social Sciences*, *65*(8), 2895--A.

Dunlosky, J., & Ariel, R. (2011). Self-regulated learning and the allocation of study time. In B. Ross (Ed.), *Psychology of learning and motivation* (pp. 103–140). Elsevier.

Dunlosky, J., & Matvey, G. (2001). Empirical analysis of the intrinsic-extrinsic distinction of judgements of learning (JOLs): Effects of relatedness and serial position on JOLs. *Journal of Experimental Psychology: Learning Memory and Cognition*, *27*(5), 1180–1191. https://doi.org/10.1037/0278-7393.27.5.1180

Dunlosky, J., & Nelson, T. O. (1994). Does the sensitivity of judgments of learning (JOLs) to the effects of various study activities depend on when the JOLs occur? *Journal of Memory and Language*, *33*, 545–565.

Dunlosky, J., & Tauber, S. K. (2015). A Brief History of Metamemory Research and Handbook Overview. *The Oxford Handbook of Metamemory, August 2018*, 1–27. https://doi.org/10.1093/oxfordhb/9780199336746.013.13

Dussias, P. E., & Piñar, P. (2010). Effects of reading span and plausibility in the reanalysis ofwh-gaps byChinese-English second language speakers. *Second Language Research*, *26*, 443–472.

Efklides, A. (2014). How does metacognition contribute to the regulation of learning? An integrative approach. *Psychological Topics*, *23*(1), 1–30.

Farley, A. P., Ramonda, K., & Liu, X. (2012). The concreteness effect and the bilingual lexicon: The impact of visual stimuli attachment on meaning recall of abstract L2 words. *Language Teaching Research*, *16*(4), 449–466. https://doi.org/10.1177/1362168812436910

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fernandez-Duque, D., Baird, J. A., & Posner, M. I. (2000). Executive Attention and Metacognitive Regulation. *Consciousness and Cognition*, *9*(2), 288–307. https://doi.org/10.1006/ccog.2000.0447

Flavell, J. H. (1979). Metacognition and Cognitive Monitoring: A New Area of Cognitive-Developmental Inquiry. *American Psychologist*, *34*(10), 906–911.

Follmer, D. J., & Sperling, R. A. (2018). Interactions between reader and text: Contributions of cognitive processes, strategy use, and text cohesion to comprehension of expository science text. *Learning and Individual Differences*, *67*(September 2017), 177–187. https://doi.org/10.1016/j.lindif.2018.08.005

Foucart, A., Romero-Rivas, C., Gort, B. L., & Costa, A. (2016). Discourse comprehension in L2: Making sense of what is not explicitly said. *Brain and Language*, *163*, 32–41. https://doi.org/10.1016/j.bandl.2016.09.001

Fox, E. (2009). The role of reader characteristics in processing and learning from informational text. *Review of Educational Research*, *79*, 202–300.

Francis, W. S., & Gutiérrez, M. (2012a). Bilingual recognition memory: Stronger performance but weaker levels-of-processing effects in the less fluent language. *Memory and Cognition*, *40*(3), 496–503. https://doi.org/https://doi.org/10.3758/s13421-011-0163-3

Francis, W. S., & Gutiérrez, M. (2012b). Bilingual recognition memory: Stronger performance but weaker levels-of-processing effects in the less fluent language. *Memory and Cognition*, *40*(3), 496–503. https://doi.org/10.3758/s13421-011-0163-3

Francis, W. S., & Strobach, E. N. (2013). The bilingual L2 advantage in recognition memory. *Psychonomic Bulletin & Review*, *20*(6), 1296–1303. https://doi.org/10.3758/s13423-013-0427-y

Frank, D. J., & Kuhlmann, B. G. (2017). More than just beliefs: Experience and beliefs jointly contribute to volume effects on metacognitive judgments. *Journal of Experimental Psychology: Learning Memory and Cognition*, *43*(5), 680–693. https://doi.org/10.1037/xlm0000332

French, M. M. J., Blood, A., Bright, N. D., Futak, D., Grohmann, M. J., Hasthorpe, A., Heritage, J., Poland, R. L., Reece, S., & Tabor, J. (2013). Changing Fonts in Education: How the Benefits Vary with Ability and Dyslexia. *The Journal of Educational Research*, *106*(4), 301–304. https://doi.org/10.1080/00220671.2012.736430

Fukaya, T. (2013). Explanation generation, not explanation expectancy, improves metacomprehension accuracy. *Metacognition and Learning*, *8*(1), 1–18. https://doi.org/10.1007/s11409-012-9093-0

Gablasova, D. (2014). Learning and retaining specialized vocabulary from textbook reading: Comparison of learning outcomes through L1 and L2. *Modern Language Journal*, *98*(4), 976–991. https://doi.org/10.1111/modl.12150

Gagné, N., & Franzen, L. (2023). How to Run Behavioural Experiments Online: Best Practice Suggestions for Cognitive Psychology and Neuroscience. *Swiss Psychology Open*, *3*(1), 1. https://doi.org/10.5334/spo.34

Galambos, S. J., & Goldin-Meadow, S. (1990). The effects of learning two languages on levels of metalinguistic awareness. *Cognition*, *34*, 1–56.

Galambos, S. J., & Hakuta, K. (1988). *Subject-specific and task-specific characteristics of metal i ng u ist ic awareness in bilingual children*. 141–162.

Gasparinatou, A., & Grigoriadou, M. (2013). Exploring the effect of background knowledge and text cohesion on learning from texts in computer science. *Educational Psychology*, *33*(6), 645–670. https://doi.org/10.1080/01443410.2013.790309

Gathercole, V. C. M., Thomas, E. M., Kennedy, I., Prys, C., Young, N., Guasch, N. V., Roberts, E. J., Hughes, E. K., & Jones, L. (2014). Does language dominance affect cognitive performance in bilinguals? Lifespan evidence from preschoolers through older adults on card sorting, Simon, and metalinguistic tasks. *Frontiers in Psychology*, *5*(FEB), 1–14. https://doi.org/10.3389/fpsyg.2014.00011

Giezen, M. R., Blumenfeld, H. K., Shook, A., Marian, V., & Emmorey, K. (2015). Parallel language activation and inhibitory control in bimodal bilinguals. *Cognition*, *141*, 9–25. https://doi.org/10.1016/j.cognition.2015.04.009

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition*, *13*, 8–20.

Glanzer, M., Adams, J. K., Iverson, G. J., & Kim, K. (1993). The regularities of recognition memory. *Psychological Review*, *100*(3), 546.

Gollan, T., Montoya, R. I., Cera, C., & Sandoval, T. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, *58*(3), 787–814. https://doi.org/10.1016/j.jml.2007.07.001

Grainger, J., & Beauvillain, C. (1988). Associative priming in bilinguals: Some limits of interlingual facilitation effects. *Canadian Journal of Psychology*, *42*, 261–273.

Grainger, J., & O'Regan, J. K. (1992). A psychophysical investigation of language priming effects in two English–French bilinguals. *European Journal of Cognitive Psychology*, *4*, 323–339.

Grant, R., Parker, P., Coward, F., Cussans, T., Regan, S., Levy, J., & Wilkinson, P. (2016). *The History Book. Big ideas simply explained.* (A. Beeden & S. Kennedy (eds.)). DK Publishing.

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, *25*(5), 515–530. https://doi.org/10.1080/20445911.2013.796377

Green, D. W., & Wei, L. (2014). A control process model of code-switching. *Language, Cognition and Neuroscience*, *29*(4), 499–511. https://doi.org/10.1080/23273798.2014.882515

Groninger, L. D. (1976). Predicting recognition during storage: The capacity of the memory system to evaluate itself. *Bulletin of the Psychonomic Society*, *7*, 425–428.

Grosjean, F. (2010). *Bilingual: Life and reality*. Harvard University Press.

Halamish, V. (2018). Can very small font size enhance memory? *Memory and Cognition*, *46*(6), 979–993. https://doi.org/10.3758/s13421-018-0816-6

Hall, S. S., Maltby, J., Filik, R., & Paterson, K. B. (2016). Key skills for science learning: the importance of text cohesion and reading ability. *Educational Psychology*, *36*(2), 191–215. https://doi.org/10.1080/01443410.2014.926313

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English.* Longman.

Harris, K. R., & Graham, S. (2009). Self-regulated strategy development in writing: Premises, evolution, and the future. *British Journal of Educational Psychology*, *2*(6), 113–135. https://doi.org/10.1348/978185409x422542

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). Elvesier Science Publisher B. V. https://doi.org/10.1007/s10749-010-0111-6

Hatzidaki, A., Branigan, H. P., & Pickering, M. J. (2011). Co-activation of syntax in bilingual language production. *Cognitive Psychology*, *62*(2), 123–150. https://doi.org/10.1016/j.cogpsych.2010.10.002

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d′. *Behavior Research Methods, Instruments, & Computers*, *27*(1), 46–51. https://doi.org/10.3758/BF03203619

Hayakawa, S., Costa, A., Foucart, A., & Keysar, B. (2016). Using a Foreign Language Changes Our Choices. *Trends in Cognitive Sciences*, *20*(11), 791–793. https://doi.org/https://doi.org/10.1016/j.tics.2016.08.004

Hertzog, C., Dunlosky, J., Emanuel Robinson, A., & Kidder, D. P. (2003). Encoding Fluency Is a Cue Used for Judgments About Learning. *Journal of Experimental Psychology: Learning Memory and Cognition*, *29*(1), 22–34. https://doi.org/10.1037/0278-7393.29.1.22

Hessel, A. K., & Schroeder, S. (2020). Interactions Between Lower- and Higher-Level Processing When Reading in a Second Language: An Eye-Tracking Study. *Discourse Processes*, *57*(10), 940–964. https://doi.org/10.1080/0163853X.2020.1833673

Hessel, A. K., & Schroeder, S. (2022). Word processing difficulty and executive control interactively shape comprehension monitoring in a second language: an eye-tracking study. *Reading and Writing*. https://doi.org/10.1007/s11145-022-10269-3

Hirshman, E., & Arndt, J. (1997). Discriminating alternative conceptions of false recognition: The cases of word concreteness and word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(1), 306–1323.

Holmes, V. M., & Langford, J. (1976). Comprehension and recall of abstract and concrete sentences. *Journal of Verbal Learning and Verbal Behavior*, *15*, 559–566.

Horiba, Y. (1996). Comprehension processes in L2 reading: Language competence, textual coherence, and inferences. *Studies in Second Language Acquisition*, *18*(4), 433–473. https://doi.org/10.1017/S02 72 26310 0015370

Horiba, Y. (2000). Reader control in reading: Effects of language competence, text type, and task. *Discourse Processes*, *29*, 223–267.

Hosoda, M. (2014). Generating causal bridging inferences in EFL expository reading: Combining on-line and off-line processing.

ARELE (Annual Review of English Language Education in Japan), *25*, 239–254.

Hosoda, M. (2015). Causal inferences during EFL reading of expository texts: Effects of two kinds of familiarity and L2 reading proficiency. *ARELE*, *26*, 205–220.

Hosoda, M. (2017). Learning from Expository Text in L2 Reading: Memory for Causal Relations and L2 Reading Proficiency. *Reading in a Foreign Language*, *29*(2), 245–263.

Hourihan, K. L., Fraundorf, S. H., & Benjamin, A. S. (2017). The influences of valence and arousal on judgments of learning and on recall. *Memory and Cognition*, *45*(1), 121–136. https://doi.org/10.3758/s13421-016-0646-3

Howard, D. V, McAndrews, M. P., & Lasaga, M. I. (1981). Semantic Priming of Lexical Decisions in Young and Old Adults1. *Journal of Gerontology*, *36*(6), 707–714. https://doi.org/10.1093/geronj/36.6.707

Hu, X., Li, T., Zheng, J., Su, N., Liu, Z., & Luo, L. (2015). How much do metamemory beliefs contribute to the font-size effect in judgments of learning? *PLoS ONE*, *10*(11), 1–11. https://doi.org/10.1371/journal.pone.0142351

Hu, X., Liu, Z., Li, T., & Luo, L. (2016). Influence of cue word perceptual information on metamemory accuracy in judgement of learning. *Memory*, *24*(3), 383–398. https://doi.org/10.1080/09658211.2015.1009470

Huff, J. D., & Nietfeld, J. L. (2009). Using strategy instruction and confidence judgments to improve metacognitive monitoring. *Metacognition and Learning*, *4*(2), 161–176. https://doi.org/10.1007/s11409-009-9042-8

Iniesta, A., Paolieri, D., Serrano, F., & Bajo, M. T. (2021). Bilingual writing coactivation: lexical and sublexical processing in a word dictation task. *Bilingualism: Language and Cognition*, 1–16. https://doi.org/doi: 10.1017/S1366728921000274

Janes, J. L., Rivers, M. L., & Dunlosky, J. (2018). The influence of making judgments of learning on memory performance: Positive, negative, or both? *Psychonomic Bulletin and Review*, *25*(6), 2356–2364. https://doi.org/10.3758/s13423-018-1463-4

Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language Learning*, *64*(1), 160–212. https://doi.org/10.1111/lang.12034

Johns, M. A., Valdés Kroff, J. R., & Dussias, P. E. (2019). Mixing things up: How blocking and mixing affect the processing of codemixed sentences. *International Journal of Bilingualism*, *23*(2), 584–611. https://doi.org/10.1177/1367006917752570

Jordano, M. L., & Touron, D. R. (2018). How often are thoughts metacognitive? Findings from research on self-regulated learning, think-aloud protocols, and mind-wandering. *Psychonomic Bulletin and Review*, *25*(4), 1269–1286. https://doi.org/10.3758/s13423-018-1490-1

Jung, J. (2018). Effects of task complexity and working memory capacity on L2 reading comprehension. *System*, *74*, 21–37. https://doi.org/10.1016/j.system.2018.02.005

Kaan, E., Kheder, S., Kreidler, A., Tomíc, A., & Valdés Kroff, J. R. (2020). Processing code-switches in the presence of others: An ERP study. *Frontiers in Psychology, 11*, 1288. https://doi.org/https://doi.org/10.3389/fpsyg.2020.01288

Kaan, E., Kirkham, J., & Wijnen, F. (2016). Prediction and integration in native and second-language processing of elliptical structures. *Bilingualism: Language and Cognition*, *19*(1), 1–18. https://doi.org/10.1017/S1366728914000844

Kaushanskaya, M., & Marian, V. (2009). The bilingual advantage in novel word learning. *Psychonomic Bulletin & Review*, *16*(4), 705–710. https://doi.org/10.3758/PBR.16.4.705

Kaushanskaya, M., & Rechtzigel, K. (2012). Concreteness effects in bilingual and monolingual word learning. *Psychonomic Bulletin and Review*, *19*(5), 935–941. https://doi.org/10.3758/s13423-012-0271-5

Kim, M. (2023). Exploring literal and inferential reading comprehension among L2 adolescent learners: the roles of working memory capacity, syllogistic inference, and L2 linguistic knowledge. *Reading and Writing*, *36*(5), 1085–1110. https://doi.org/10.1007/s11145-022-10320-3

King, J. F., Zechmeister, E. B., & Shaughnessy, J. J. (1980). Judgments of knowing: The influence of retrieval practice. *The American Journal of Psychology*, *93*, 329–343.

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review*, *100*(4), 609–639. https://doi.org/10.1037/0033-295X.100.4.609

Koriat, A. (1997). Monitoring One's Own Knowledge during Study: A Cue-Utilization Approach to Judgments of Learning. *Journal of Experimental Psychology: General*, *126*(4), 349–370. https://doi.org/10.1037/0096-3445.126.4.349

Koriat, A. (2008). Easy comes, easy goes? The link between learning and remembering and its exploitation in metacognition. *Memory & Cognition*, *36*(2), 416–428.

Koriat, A. (2015). Metacognition: Decision making Processes in Self-monitoring and Self-regulation. In G. Keren & G. Wu (Eds.), *The Wiley Blackwell Handbook of Judgment and Decision Making.: Vol. I*. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118468333.ch5

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, *103*(3), 490–517. https://doi.org/10.1037/0033-295X.103.3.490

Koriat, A., & Ma'ayan, H. (2005). The effects of encoding fluency and retrieval fluency on judgments of learning. *Journal of Memory and Language*, *52*(4), 478–492.

Koriat, A., Ma'ayan, H., & Nussinson, R. (2006). The intricate relationships between monitoring and control in metacognition: Lessons for the cause-and-effect relation between subjective experience and behavior. *Journal of Experimental Psychology: General*, *135*(1), 36–69. https://doi.org/10.1037/0096-3445.135.1.36

Kornell, N., & Rhodes, M. G. (2013). Feedback reduces the metacognitive benefit of tests. *Journal of Experimental Psychology: Applied*, *19*, 1–13.

Kornell, N., Rhodes, M. G., Castel, A. D., & Tauber, S. K. (2011). The ease-of-processing heuristic and the stability bias: Dissociating memory, memory beliefs, and memory judgments. *Psychological*

Science, *22*(6), 787–794. https://doi.org/10.1177/0956797611407929

Krebs, S. S., & Roebers, C. M. (2012). The impact of retrieval processes, age, general achievement level, and test scoring scheme for children's metacognitive monitoring and controlling. *Metacognition and Learning*, *7*(2), 75–90. https://doi.org/10.1007/s11409-011-9079-3

Kroll, J. F., Bobb, S. C., & Hoshino, N. (2014). Two Languages in Mind: Bilingualism as a Tool to Investigate Language, Cognition, and the Brain. *Current Directions in Psychological Science*, *23*(3), 159–163. https://doi.org/10.1177/0963721414528511

Kroll, J. F., Dussias, P. E., Bice, K., & Perrotti, L. (2015). Bilingualism, mind, and brain. *Annual Review of Linguistics.*, *1*(1), 377–394. https://doi.org/10.1016/j.physbeh.2017.03.040

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: evidence for asymmetric connections betwwen bilingual memory representations. *Journal of Memory and Language*, *33*, 149–174. https://doi.org/10.1121/1.2934955

Kühl, T., & Eitel, A. (2016). Effects of disfluency on cognitive and metacognitive processes and outcomes. *Metacognition and Learning*, *11*(1), 1–13. https://doi.org/10.1007/s11409-016-9154-x

Kuperman, V., Kyröläinen, A.-J., Porretta, V., Brysbaert, M., & Yang, S. (2021). A lingering question addressed: Reading rate and most efficient listening rate are highly similar. *Journal of Experimental Psychology: Human Perception and Performance*, *47*(8), 1103–1112. https://doi.org/10.1037/xhp0000932

Lanska, M., Olds, J. M., & Westerman, D. L. (2014). Fluency effects in recognition memory: Are perceptual fluency and conceptual fluency interchangeable. *Journal of Experimental Psychology: Learning Memory and Cognition*, *40*(1), 1–11. https://doi.org/10.1037/a0034309

Le Ny, J. F., Denhiere, G., & Le Taillanter, D. (1972). Regulation of study-time and interstimulus similarty in self-paced learning conditions. *Acta Psychologica*, *36*(4), 280–289.

Lefèvre, N., & Lories, G. (2004). Text cohesion and metacomprehension: Immediate and delayed judgments. *Memory* *and Cognition*, *32*(8), 1238–1254. https://doi.org/10.3758/BF03206315

Lehmann, J., Goussios, C., & Seufert, T. (2016). Working memory capacity and disfluency effect: An aptitude-treatment-interaction study. *Metacognition and Learning.*, *11*(1), 89–105.

Lemhöfer, K., Spalek, K., & Schriefers, H. (2008). Cross-language effects of grammatical gender in bilingual word recognition and production. *Journal of Memory and Language*, *59*(3), 312–330. https://doi.org/10.1016/j.jml.2008.06.005

Li, X., Chen, G., & Yang, C. (2021). How cognitive conflict affects judgments of learning: Evaluating the contributions of processing fluency and metamemory beliefs. *Memory and Cognition*, 912–922. https://doi.org/10.3758/s13421-021-01143-8

Libben, M. R., & Titone, D. A. (2009). Bilingual lexical access in context: Evidence from eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(2), 381–390. https://doi.org/https://doi.org/10.1037/a0014875

Logan, J. M., Castel, A. D., Haber, S., & Viehman, E. J. (2012). Metacognition and the spacing effect: The role of repetition, feedback, and instruction on judgments of learning for massed and spaced rehearsal. *Metacognition and Learning*, *7*(3), 175–195. https://doi.org/10.1007/s11409-012-9090-3

Lonka, K., Olkinuora, E., & Mäkinen, J. (2004). Aspects and prospects of measuring studying and learning in higher education. *Educational Psychology Review*, *16*(4), 301–323. https://doi.org/10.1007/s10648-004-0002-1

López-Rojas, C., Rossi, E., Marful, A., & Bajo, M. T. (2022). Prospective memory in bilinguals and monolinguals: ERP and behavioural correlates of prospective processing in bilinguals. *Brain and Language*, *225*. https://doi.org/10.1016/j.bandl.2021.105059

Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology*, *25*(5), 605–621. https://doi.org/10.1080/20445911.2013.795574

Luk, G., De Sa, E., & Bialystok, E. (2011). Is there a relation between onset age of bilingualism and enhancement of cognitive control?

Bilingualism: Language and Cognition, *14*(4), 588–595. https://doi.org/10.1017/S1366728911000010

Luna, K., Albuquerque, P. B., & Martín-Luengo, B. (2019). Cognitive load eliminates the effect of perceptual information on judgments of learning with sentences. *Memory and Cognition*, *47*(1), 106–116. https://doi.org/10.3758/s13421-018-0853-1

Luna, K., Martín-Luengo, B., & Albuquerque, P. B. (2018). Do delayed judgements of learning reduce metamemory illusions? A meta-analysis. *Quarterly Journal of Experimental Psychology*, *71*(7), 1626–1636. https://doi.org/10.1080/17470218.2017.1343362

Ma, H., Hu, J., Xi, J., Shen, W., Ge, J., Geng, F., & Yao, D. (2014). Bilingual cognitive control in language switching: An fMRI study of English-Chinese late bilinguals. *PloS One*, *9*(9), e106468.

Macaro, E. (2018). *English medium instruction*. Oxford University Press.

Macizo, P., Bajo, M. T., & Cruz Martín, M. (2010). Inhibitory processes in bilingual language comprehension: Evidence from Spanish-English interlexical homographs. *Journal of Memory and Language*, *63*(2), 232–244. https://doi.org/10.1016/j.jml.2010.04.002

Macizo, P., Bajo, T., & Paolieri, D. (2012). Language switching and language competition. *Second Language Research*, *28*(2), 131–149. https://doi.org/10.1177/0267658311434893

Macmillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group data: estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin*, *98*(1), 185–199.

Magreehan, D. A., Serra, M. J., Schwartz, N. H., & Narciss, S. (2016). Further boundary conditions for the effects of perceptual disfluency on judgments of learning. *Metacognition and Learning*, *11*(1), 35–56. https://doi.org/10.1007/s11409-015-9147-1

Marcouse, I., Anderson, P., Black, A., Machin, D., & Watson, N. (2014). *The Business Book. Big ideas simply explained.* (S. Atkinson (ed.)).

Marful, A., Díez, E., & Fernandez, A. (2015). Normative data for the 56 categories of Battig and Montague (1969) in Spanish. *Behavior Research Methods*, *47*(3), 902–910.

https://doi.org/10.3758/s13428-014-0513-8

Marian, V., Bartolotti, J., van den Berg, A., & Hayakawa, S. (2021). Costs and Benefits of Native Language Similarity for Non-native Word Learning. *Frontiers in Psychology*, *12*(May). https://doi.org/10.3389/fpsyg.2021.651506

Marian, V., Blumenfeld, H. K., & Boukrina, O. V. (2008). Sensitivity to phonological similarity within and across languages. *Journal of Psycholinguistic Research*, *37*(3), 141–170. https://doi.org/10.1007/s10936-007-9064-9

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing Language Profiles in Bilinguals and Multilinguals. *Journal of Speech Language and Hearing Research*, *50*(4), 940. https://doi.org/10.1044/1092-4388(2007/067)

Marian, V., & Spivey, M. (2003a). Bilingual and monolingual processing of competing lexical items. *Applied Psycholinguistics*, *24*(2), 173–193. https://doi.org/10.1017/S0142716403000092

Marian, V., & Spivey, M. (2003b). Competing activation in bilingual language processing: Within- and between-language competition. *Bilingualism: Language and Cognition*, *6*(2), 97–115. https://doi.org/doi:10.1017/S1366728903001068

Mat Daud, N. S., Mat Daud, N., & Abu Kassim, N. L. (2005). Second language writing anxiety: Cause or effect? *Malaysian Journal of ELT Research*, *1*(1), 19.

Matvey, G., Dunlosky, J., & Schwartz, B. L. (2006). The effects of categorical relatedness on judgements of learning (JOLs). *Memory*, *14*(2), 253–261. https://doi.org/10.1080/09658210500216844

Maxwell, N. P., Perry, T., & Huff, M. J. (2021). Perceptually fluent features of study words do not inflate judgemnts of learning: evidence from font size, highlights, and Sans Forgetica font type. *Metacognition and Learning*, 293–319. https://doi.org/10.1007/s11409-021-09284-6

Mazzoni, G., & Cornoldi, C. (1993). Strategies in study item allocation: Why is study time sometimes not effective? *Journal of Experimental Psychology: General*, *122*, 47–60.

Mazzoni, G., Cornoldi, C., & Marchitelli, G. (1990). Do memorability ratings affect study-time allocation? *Memory & Cognition*, *18*(2), 196–204.

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, *14*(1), 1–43. https://doi.org/10.1207/s1532690xci1401_1

Melby-Lervåg, M., & Lervåg, A. (2014). Reading comprehension and its underlying components in second-language learners: A meta-analysis of studies comparing first- and second-language learners. *Psychological Bulletin*, *140*(2), 409–433. https://doi.org/10.1037/a0033890

Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current Directions in Psychological Science*, *18*, 159–163. https://doi.org/doi.org/10.1111/j.1467-8721.2009.01628.x

Metcalfe, J. (2011). Desirable difficulties and studying in the region of proximal learning. In *Successful remembering and successful forgetting: A festschrift in honor of Robert A. Bjork.* (pp. 259–276). Psychology Press.

Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review*, *15*(1), 174–179.

Meuter, R. F. I., & Allport, A. (1999). Bilingual Language Switching in Naming: Asymmetrical Costs of Language Selection. *Journal of Memory and Language*, *40*(1), 25–40. https://doi.org/10.1006/jmla.1998.2602

Morales, J., Gómez-Ariza, C. J., & Bajo, M. T. (2013). Dual mechanisms of cognitive control in bilinguals and monolinguals. *Journal of Cognitive Psychology*, *25*, 531–546.

Morales, L., Paolieri, D., Cubelli, R., & Bajo, M. T. (2014). Transfer of Spanish grammatical gender to English: Evidence from immersed and non-immersed bilinguals. *Bilingualism: Language and Cognition*, *17*(4), 700–708. https://doi.org/10.1017/S1366728914000017

Moreno, E. M., Federmeier, K. D., & Kutas, M. (2002). Switching languages, switching Palabras (words): An electrophysiological study of code switching. *Brain and Language*, *80*(2), 188–207. https://doi.org/10.1006/brln.2001.2588

Moreno, M. A., & van Orden, G. C. (2001). *Word Recognition, Cognitive Psychology of* (N. J. Smelser & P. B. B. T.-I. E. of the S. & B. S. Baltes (eds.); pp. 16556–16561). Pergamon. https://doi.org/https://doi.org/10.1016/B0-08-043076-7/01553-9

Moreno, S., Bialystok, E., Wodniecka, Z., & Alain, C. (2010). Conflict Resolution in Sentence Processing by Bilinguals. *Journal of Neurolinguistics*, *23*(6), 564–579. https://doi.org/10.1016/j.jneuroling.2010.05.002.Conflict

Mueller, M. L., & Dunlosky, J. (2017). How beliefs can impact judgments of learning: Evaluating analytic processing theory with beliefs about fluency. *Journal of Memory and Language*, *93*, 245–258. https://doi.org/10.1016/j.jml.2016.10.008

Mueller, M. L., Tauber, S. K., & Dunlosky, J. (2013). Contributions of beliefs and processing fluency to the effect of relatedness on judgments of learning. *Psychonomic Bulletin and Review*, *20*(2), 378–384. https://doi.org/10.3758/s13423-012-0343-6

Mummery, C. J., Shallice, T., & Price, C. J. (1999). Dual-Process Model in Semantic Priming: A Functional Imaging Perspective. *NeuroImage*, *9*(5), 516–525. https://doi.org/https://doi.org/10.1006/nimg.1999.0434

Nairne, J. S. (1988). The mnemonic value of perceptual identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(2), 248–255.

Neely, J. H., & Tse, C. S. (2011). Semantic relatedness effects on true and false memories in episodic recognition: A methodological and empirical review. In *The Foundations of Remembering* (pp. 313–352).

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, *95*, 109–133. https://doi.org/http://dx.doi.org/10.1037/0033-2909.95.1.109.

Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the “labor-in-vain effect.” *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(4), 676–686.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation.* (pp. 125–141). Academic Press.

Nguyen, K., & McDaniel, M. (2016). The JOIs of text comprehension: Supplementing retrieval practice to enhance inference performance. *Journal of Experimental Psychology: Applied*, *22*(1), 59–71. https://doi.org/10.1037/xap0000066

Ning, S., Hayakawa, S., Bartolotti, J., & Marian, V. (2020). On language and thought: Bilingual experience influences semantic associations. *Journal of Neurolinguistics*, *56*, 100932.

Opitz, B., & Degner, J. (2012). Emotionality in a second language: It’s a matter of time. *Neuropsychologia*, *50*(8), 1961–1967. https://doi.org/10.1016/j.neuropsychologia.2012.04.021

Oppenheimer, D. M., & Alter, A. L. (2014). The search for moderators in disfluency research. *Applied Cognitive Psychology*, *28*(4), 502–504. https://doi.org/10.1002/acp.3023

Oswald, F. L., McAbee, S. T., Redick, T. S., & Hambrick, D. Z. (2014). The development of a short domain-general measure of working memory capacity. *Behavior Research Methods*, *47*(4), 1343–1355. https://doi.org/10.3758/s13428-014-0543-2

Ozuru, Y., Kurby, C. A., & McNamara, D. S. (2012). The effect of metacomprehension judgment task on comprehension monitoring and metacognitive accuracy. *Metacognition and Learning*, *7*(2), 113–131. https://doi.org/10.1007/s11409-012-9087-y

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology*, *45*(255–287).

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, *8*(422), 1–28. https://doi.org/10.3389/fpsyg.2017.00422

Panadero, E., Fraile, J., Fernández Ruiz, J., Castilla-Estévez, D., & Ruiz, M. A. (2019). Spanish university assessment practices: examination tradition with diversity by faculty. *Assessment &*

Evaluation in Higher Education, *44*(3), 379–397. https://doi.org/10.1080/02602938.2018.1512553

Pannu, J. K., & Kaszniak, A. W. (2005). Metamemory experiments in neurological populations: a review. *Neuropsychology Review*, *15*(3), 105–130.

Paolieri, D., Padilla, F., Koreneva, O., Morales, L., & Macizo, P. (2019). Gender congruency effects in Russian-Spanish and Italian-Spanish bilinguals: The role of language proximity and concreteness of words. *Bilingualism*, *22*(1), 112–129. https://doi.org/10.1017/S1366728917000591

Pelegrina, S., Bajo, M. T., & Justicia, F. (2000). Differential allocation of study time: Incomplete compensation for the difficulty of the materials. *Memory*, *8*(6), 377–392. https://doi.org/10.1080/09658210050156831

Pérez-Mata, M. N., Read, J. D., & Diges, M. (2002). Effects of divided attention and word concreteness on correct recall and false memory reports. *Memory*, *10*(3), 161–177. https://doi.org/10.1080/09658210143000308

Pérez, A. I., Hansen, L., & Bajo, M. T. (2018). The nature of first and second language processing: The role of cognitive control and L2 proficiency during text-level comprehension. *Bilingualism: Language and Cognition*, 1–19. https://doi.org/10.1017/S1366728918000846

Pérez, A. I., Schmidt, E., & Tsimpli, I. (2023). Inferential evaluation and revision in L1 and L2 text comprehension : An eye movement study. *Bilingualism: Language and Cognition*, 1–14.

Pessoa, S., Miller, R. T., & Kaufer, D. (2014). Students' challenges and development in the transition to academic writing at an English-medium university in Qatar. *IRAL, International Review of Applied Linguistics in Language Teaching*, *52*(2), 127–156. https://doi.org/https://doi.org/10.1515/iral-2014-0006

Pieger, E., Mengelkamp, C., & Bannert, M. (2016). Metacognitive judgments and disfluency - Does disfluency lead to more accurate judgments, better control, and better performance? *Learning and Instruction*, *44*, 31–40. https://doi.org/10.1016/j.learninstruc.2016.01.012

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of Self-Regulation* (pp. 452–50). Academic Press.

Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review*, *16*(385–407), 385–407. https://doi.org/doi: 10.1007/s10648-004-0006-x

Pintrich, P. R., & De Groot, E. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, *82*(1), 33–40.

Pintrich, P. R., Smith, D. A. F., García, T., & McKeachie, W. J. (1991). *A Manual for the Use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. Office of Educational Research and Improvement. National Center for Research to Improve Postsecondary Teaching and Learning. https://doi.org/10.5901/mjss.2015.v6n1p156

Pintrich, P. R., Smith, D. A. F., García, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the motivated strategies for learning questionnaire (MSLQ). *Educational and Psychological Measurement1*, *53*, 801–813.

Pintrich, P. R., & Zusho, A. (2007). Motivation and self-regulated learning in the college classroom. In R. Perry & J. Smart (Eds.), *Handbook on Teaching and Learning in Higher Education.* Springer Publishers.

Plat, R., Lowie, W., & de Bot, K. (2018). Word Naming in the L1 and L2: A Dynamic Perspective on Automatization and the Degree of Semantic Involvement in Naming. *Frontiers in Psychology*, *8*, 2256. https://doi.org/10.3389/fpsyg.2017.02256

Pressley, M., Ghatala, E. S., Pirie, J., & Woloshyn, V. E. (1990). Being really, really certain you know the main idea doesn’t mean you do. *National Reading Conference Yearbook*, *39*(249–256).

Prinz, A., Golke, S., & Wittwer, J. (2020). How accurately can learners discriminate their comprehension of texts? A comprehensive meta-analysis on relative metacomprehension accuracy and influencing factors. *Educational Research Review*, *31*(August). https://doi.org/10.1016/j.edurev.2020.100358

Ransdell, S., Arecco, M. R., & Levy, C. M. (2001). Bilingual long- term working memory: The effects of working memory loads on writing quality and fluency. *Applied PsychoLinguistics*, *22*, 113–128.

Rawson, K. A., & Dunlosky, J. (2002). Are performance predictions for text based on ease of processing? *Journal of Experimental Psychology: Learning, Memory, and Cognition.*, *28*(1), 69–80.

Reynolds, M. G., Schlöffel, S., & Peressotti, F. (2016). Asymmetric switch costs in numeral naming and number word reading: Implications for models of bilingual language production. *Frontiers in Psychology*, *6*(JAN), 1–15. https://doi.org/10.3389/fpsyg.2015.02011

Rhodes, M. G. (2015). Judgments of Learning: Methods, Data, and Theory. In *The Oxford Handbook of Metamemory*. https://doi.org/10.1093/oxfordhb/9780199336746.013.4

Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General*, *137*, 615–625.

Richardson, J. T. E. (2003). Dual coding versus relational processing in memory for concrete and abstract words. *European Journal of Cognitive Psychology*, *15*, 481–509.

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long- term retention. *Trends in Cognitive Science*, *15*, 20–27.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Roessel, J., Schoel, C., & Stahlberg, D. (2018). What's in an accent? General spontaneous biases against nonnative accents: An investigation with conceptual and auditory IATs. *European Journal of Social Psychology*, *48*, 535–550. https://doi.org/10.1002/ejsp.2339

Romani, C., McAlpine, S., & Martin, R. C. (2008). Concreteness effects in different tasks: Implications for models of short-term memory.

Quarterly Journal of Experimental Psychology, *61*, 292–323. https://doi.org/doi:10.1080/ 17470210601147747

Rosner, T. M., Davis, H., & Milliken, B. (2015). Perceptual blurring and recognition memory: A desirable difficulty effect revealed. *Acta Psychologica*, *160*, 11–22. https://doi.org/10.1016/j.actpsy.2015.06.006

Rummer, R., Schweppe, J., & Schwede, A. (2016). Fortune is fickle: null-effects of disfluency on learning outcomes. *Metacognition and Learning*, *11*(1), 57–70. https://doi.org/10.1007/s11409-015-9151-5

Sabourin, L., & Viñerte, S. (2019). Cognitive control among immersed bilinguals: Considering differences in linguistic and non-linguistic processing. *Bilingualism: Language and Cognition*, *22*(3), 590–605. https://doi.org/10.1017/S1366728918000524

Sandoval, T., Gollan, T., Ferreira, V., & Salmon, D. (2010). What causes the bilingual disadvantage in verbal fluency? The dual-task analogy. *Bilingualism: Language and Cognition*, *13*(2), 231–252. https://doi.org/10.1017/S1366728909990514

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime: User's guide.* (Psychology Software Incorporated.).

Schroeder, J., Harvey, D., Coyne, C., Jackson, T., Farndon, J., Singer, A., & Harris, T. (2019). *The Ecology Book. Big ideas simply explained.* (H. Fewster & C. Hallinan (eds.)). DK Publishing. https://doi.org/10.1080/00988157.1974.9977066

Schunk, D. H., & Greene, J. A. (2018). Historical, contemporary, and future perspectives on self-regulated learning and performance. In J. A. Greene & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 1–15). Routledge/Taylor & Francis Group. https://doi.org/10.4324/9781315697048-1

Schunk, D. H., & Zimmerman, B. J. (1994). *Self-regulation of learning and performance: Issues and educational applications.* Lawrence Erlbaum Associates.

Schwanenflugel, P. J., Akin, C., & Luh, W. (1992). Context availability and the recall of abstract and concrete words. *Memory and Cognition*, *20*(1), 96–104.

Schwanenflugel, P. J., Harnishfeger, K. K., & Stowe, R. W. (1988). Context availability and lexical decisions for abstract and concrete words. *Journal of Memory and Language*, *27*(5), 499–520.

Schwartz, A. I., & Kroll, J. F. (2006). Bilingual lexical activation in sentence context. *Journal of Memory and Language*, *55*(2), 197–212. https://doi.org/10.1016/j.jml.2006.03.004

Sebastián, N., Martí, M. A., Carreiras, M. F., & Cuetos, F. (2000). *LEXESP, léxico informatizado del español.* Ediciones de la Universitat de Barcelona.

Seufert, T. (2020). Building Bridges Between Self-Regulation and Cognitive Load—an Invitation for a Broad and Differentiated Attempt. *Educational Psychology Review*, *32*(4), 1151–1162. https://doi.org/10.1007/s10648-020-09574-6

Seufert, T., Wagner, F., & Westphal, J. (2017). The effects of different levels of disfluency on learning outcomes and cognitive load. *Instructional Science*, *45*(2), 221–238. https://doi.org/10.1007/s11251-016-9387-8

Shook, A., & Marian, V. (2012). Bimodal bilinguals co-activate both languages during spoken comprehension. *Cognition*, *124*(3), 314–324. https://doi.org/10.1016/j.cognition.2012.05.014

Shook, A., & Marian, V. (2013). The bilingual language interaction network for comprehension of speech. *Bilingualism*, *16*(2), 304–324. https://doi.org/10.1017/S1366728912000466

Shook, A., & Marian, V. (2019). Covert Co-Activation of Bilinguals' Non-Target Language: Phonological Competition from Translations. *Linguist Approaches Biling.*, *9*(2), 228–252. https://doi.org/10.1075/lab.17022.sho.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.

Soares, A. P., Oliveira, H., Ferreira, M., Comesaña, M., MacEdo, A. F., Ferré, P., Acuña-Fariña, C., Hernández-Cabrera, J., & Fraga, I. (2019). Lexico-syntactic interactions during the processing of temporally ambiguous L2 relative clauses: An eye-tracking study with intermediate and advanced Portuguese-English bilinguals. *PLoS ONE*, *14*(5), 1–27. https://doi.org/10.1371/journal.pone.0216779

Soderstrom, N. C., & McCabe, D. P. (2011). The interplay between value and relatedness as bases for metacognitive monitoring and control: Evidence for agenda-based monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1236– 1242. https://doi.org/doi:10.1037/a0023548

Soemer, A., & Schiefele, U. (2019). Text difficulty, topic interest, and mind wandering during reading. *Learning and Instruction*, *61*(December 2018), 12–22. https://doi.org/10.1016/j.learninstruc.2018.12.006

Stine-Morrow, E. A. L., Shake, M. C., Miles, J. R., & Noh, S. R. (2006). Adult age differences in the effects of goals on self-regulated sentence processing. *Psychology and Aging*, *21*(4), 790–803. https://doi.org/10.1037/0882-7974.21.4.790

Stoff, D. M., & Eagle, M. N. (1971). The relationship among reported strategies, presentation rate, and verbal ability and their effects on free recall learning. *Journal of Experimental Psychology*, *87*(3), 423–428.

Strukelj, A., Scheiter, K., Nyström, M., & Holmqvist, K. (2016). Exploring the lack of a disfluency effect: evidence from eye movements. *Metacognition and Learning*, *11*(1), 71–88. https://doi.org/10.1007/s11409-015-9146-2

Sungkhasettee, V. W., Friedman, M. C., & Castel, A. D. (2011). Memory and metamemory for inverted words: Illusions of competency and desirable difficulties. *Psychonomic Bulletin & Review*, *18*(5), 973. https://doi.org/10.3758/s13423-011-0114-9

Susser, J. A., Mulligan, N. W., & Besken, M. (2013). The effects of list composition and perceptual fluency on judgments of learning (JOLs). *Memory and Cognition*, *41*(7), 1000–1011. https://doi.org/10.3758/s13421-013-0323-8

Taconnat, L., Morel, S., Guerrero–Sastoque, L., Frasca, M., & Vibert, N. (2020). What eye movements reveal about strategy encoding of words in younger and older adults. *Memory*, *28*(4), 1–16. https://doi.org/10.1080/09658211.2020.1745848

Taconnat, L., Raz, N., Toczé, C., Bouazzaoui, B., Sauzéon, H., Fay, S., & Isingrini, M. (2009). Ageing and organisation strategies in free recall: The role of cognitive flexibility. *European Journal of*

Cognitive Psychology, *21*(2–3), 347–365. https://doi.org/10.1080/09541440802296413

Tauber, S. K., & Dunlosky, J. (2012). Can older adults judge their learning of emotional information? *Psychology and Aging*, *27*, 924–933.

Tauber, S. K., & Rhodes, M. G. (2012). Measuring memory monitoring with judgements of retention (JORs). *Quarterly Journal of Experimental Psychology*, *65*(7), 1376–1396. https://doi.org/10.1080/17470218.2012.656665

Tauber, S. K., & Witherby, A. E. (2019). Do Judgments of Learning Modify Older Adults' Actual Learning? *Psychology and Aging*, *34*(6), 836–847. https://doi.org/10.1037/pag0000376

Taylor, A., Sanson, M., Burnell, R., Wade, K. A., & Garry, M. (2020). Disfluent difficulties are not desirable difficulties: the (lack of) effect of Sans Forgetica on memory. *Memory*, 1–8. https://doi.org/10.1080/09658211.2020.1758726

Tekin, E. (2022). Can Learners Allocate Their Study Time Effectively? It Is Complicated. In *Educational Psychology Review* (Vol. 34, Issue 2). Springer US. https://doi.org/10.1007/s10648-021-09645-2

Thiede, K. W., Anderson, M. C. M., & Herriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology*, *95*, 66–73.

Thompson, V. A., Ackerman, R., Sidi, Y., Ball, L. J., Pennycook, G., & Prowse Turner, J. A. (2013). The role of answer fluency and perceptual fluency in the monitoring and control of reasoning: Reply to Alter, Oppenheimer, and Epley (2013). *Cognition*, *128*(2), 256–258. https://doi.org/10.1016/j.cognition.2013.03.003

Thompson, V. A., Turner, J. A. P., Pennycook, G., Ball, L. J., Brack, H., Ophir, Y., & Ackerman, R. (2013). The role of answer fluency and perceptual fluency as metacognitive cues for initiating analytic thinking. *Cognition*, *128*(2), 237–251. https://doi.org/10.1016/j.cognition.2012.09.012

Thorpe, C., Todd, M., Yuill, C., Tomley, S., Hobbs, M., & Weeks, M. (2015). *The Sociology Book. Big ideas simply explained.* (S. Atkinson (ed.)). DK Publishing.

Titone, R. (1997). Early second-language learning: Bilingualism and metalinguistic development. *Europe Plurilingue*, *12*, 138–149.

Tullis, J. G., & Benjamin, A. S. (2012). Consequences of restudy choices in younger and older learners. *Psychonomic Bulletin and Review*, *19*(4), 743–749. https://doi.org/10.3758/s13423-012-0266-2

Tullis, J. G., Finley, J. R., & Benjamin, A. S. (2013). Metacognition of the testing effect: Guiding learners to predict the benefits of retrieval. *Memory & Cognition*, *41*, 429–442.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*, 127–154. https://doi.org/doi: 10.1016/0749-596X(89)90040-5

Undorf, M., & Bröder, A. (2020). Cue integration in metamemory judgements is strategic. *Quarterly Journal of Experimental Psychology*, *73*(4), 629–642. https://doi.org/10.1177/1747021819882308

Undorf, M., & Erdfelder, E. (2015). The relatedness effect on judgments of learning: A closer look at the contribution of processing fluency. *Memory and Cognition*, *43*(4), 647–658. https://doi.org/10.3758/s13421-014-0479-x

Undorf, M., Söllner, A., & Bröder, A. (2018). Simultaneous utilization of multiple cues in judgments of learning. *Memory and Cognition*, *46*(4), 507–519. https://doi.org/10.3758/s13421-017-0780-6

Undorf, M., & Zander, T. (2017). Intuition and metacognition: The effect of semantic coherence on judgments of learning. *Psychonomic Bulletin and Review*, *24*(4), 1217–1224. https://doi.org/10.3758/s13423-016-1189-0

Undorf, M., Zimdahl, M. F., & Bernstein, D. M. (2017). Perceptual fluency contributes to effects of stimulus size on judgments of learning. *Journal of Memory and Language*, *92*, 293–304. https://doi.org/10.1016/j.jml.2016.07.003

Ushiro, Y., Hosoda, M., Komuro, R., Mori, Y., & Nishi, T. (2022). Quantitative and Qualitative Effects of the Reading Goal on the Monitoring of Global Causal Coherence in L2 Reading. *JACET Journal*, *66*, 113–131.

Vaccaro, A. G., & Fleming, S. M. (2018). Thinking about thinking: A coordinate-based meta-analysis of neuroimaging studies of metacognitive judgements. *Brain and Neuroscience Advances*, *2*, 1–14. https://doi.org/10.1177/2398212818810591

van den Broek, P., & Helder, A. (2017). Cognitive Processes in Discourse Comprehension: Passive Processes, Reader-Initiated Processes, and Evolving Mental Representations. *Discourse Processes*, *54*(5–6), 360–372. https://doi.org/10.1080/0163853X.2017.1306677

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. Academic Press.

van Hell, J. G., & de Groot, A. M. B. (1998). Disentangling context availability and concreteness in lexical decision and word translation. *The Quarterly Journal of Experimental Psychology*, *51A*, 41–63.

van Hell, J. G., & de Groot, A. M. B. (2008). Sentence context modulates visual word recognition and translation in bilinguals. *Acta Psychologica*, *128*, 431–451.

Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language*, *50*(3), 289–335. https://doi.org/10.1016/j.jml.2003.10.003

Vander Beken, H., & Brysbaert, M. (2018). Studying texts in a second language: The importance of test type. *Bilingualism: Language and Cognition*, *21*(5), 1062–1074. https://doi.org/10.1017/S1366728917000189

Vander Beken, H., De Bruyne, E., & Brysbaert, M. (2020). Studying texts in a non-native language: A further investigation of factors involved in the L2 recall cost. *Quarterly Journal of Experimental Psychology*, *73*(6), 891–907. https://doi.org/10.1177/1747021820910694

Vander Beken, H., Woumans, E., & Brysbaert, M. (2018). Studying texts in a second language: No disadvantage in long-term recognition memory. *Bilingualism: Language and Cognition*, *21*(4), 826–838. https://doi.org/10.1017/S1366728917000360

Verhoef, K., Roelofs, A., & Chwilla, D. J. (2009). Role of inhibition in language switching: Evidence from event-related brain potentials in overt picture naming. *Cognition*, *110*(1), 84–99. https://doi.org/10.1016/j.cognition.2008.10.013

Vermunt, J. D., & Vermetten, Y. J. (2004). Patterns in student learning: Relationships between learning strategies, conceptions of learning, and learning orientations. *Educational Psychology Review*, *16*(4), 359–384. https://doi.org/10.1007/s10648-004-0005-y

Wagner, A. D., & Koutstaal, W. (2002). *Priming* (V. S. B. T.-E. of the H. B. Ramachandran (ed.); pp. 27–46). Academic Press. https://doi.org/https://doi.org/10.1016/B0-12-227210-2/00286-7

Walker, I., & Hulme, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *25*, 1256–1271.

Wang, J., & Xing, Q. (2019). Metacognitive illusion in category learning: Contributions of processing fluency and beliefs. *Advances in Cognitive Psychology*, *15*(2), 100–110. https://doi.org/10.5709/acp-0260-3

Watcharapunyawong, S., & Usaha, S. (2013). Thai EFL students' writing errors in different text types: The interference of the first language. *English Language Teaching*, *6*(1), 67–78. https://doi.org/10.5539/elt.v6n1p67

Wehr, T., & Wippich, W. (2004). Typography and color: effects of salience and fluency on conscious recollective experience. *Psychological Research*, *69*(1–2), 138–146. https://doi.org/10.1007/s00426-003-0162-5

Weissgerber, S. C., & Reinhard, M. A. (2017). Is disfluency desirable for learning? *Learning and Instruction*, *49*, 199–217. https://doi.org/10.1016/j.learninstruc.2017.02.004

West, R. L., & Thorn, R. M. (2001). Goal-setting, self-efficacy, and memory performance in older and younger adults. *Experimental Aging Research*, *27*(1), 41–65. https://doi.org/https://doi.org/10.1080/ 036107301750046133

Whitford, V., & Titone, D. (2012). Second-language experience modulates first- and second-language word frequency effects:

Evidence from eye movement measures of natural paragraph reading. *Psychonomic Bulletin & Review*, *19*(1), 73–80. https://doi.org/10. 3758/s13423-011-0179-5

Wirth, J., Stebner, F., Trypke, M., Schuster, C., & Leutner, D. (2020). An Interactive Layers Model of Self-Regulated Learning and Cognitive Load. *Educational Psychology Review*, *32*(4), 1127–1149. https://doi.org/10.1007/s10648-020-09568-4

Witherby, A. E., & Tauber, S. K. (2017). The concreteness effect on judgments of learning: Evaluating the contributions of fluency and beliefs. *Memory and Cognition*, *45*(4), 639–650. https://doi.org/10.3758/s13421-016-0681-0

Woll, N. (2018). Investigating dimensions of metalinguistic awareness: what think-aloud protocols revealed about the cognitive processes involved in positive transfer from L2 to L3. *Language Awareness*, *27*(1–2), 167–185. https://doi.org/10.1080/09658416.2018.1432057

Xie, H., Zhou, Z., & Liu, Q. (2018). Null Effects of Perceptual Disfluency on Learning Outcomes in a Text-Based Educational Context: a Meta-analysis. *Educational Psychology Review*, *30*(3), 745–771. https://doi.org/10.1007/s10648-018-9442-x

Yahya, M., & Özkan Ceylan, A. (2022). Interactions between language and inhibitory control: Evidence from a combined language switching and Stroop paradigm. *International Journal of Bilingualism*, *26*(6), 675–694. https://doi.org/10.1177/13670069211062554

Yoshida, M. (2003). Working memory capacity and the use of inference in L2 reading. *JACET Bulliten*, *40*, 57–82.

Yue, C. L., Castel, A. D., & Bjork, R. A. (2013). When disfluency is - and is not- a desirable difficulty: The influence of typeface clarity on metacognitive judgments and memory. *Memory & Cognition*, *41*, 229–241.

Zechmeister, E. B., & Shaughnessy, J. J. (1980a). When you know that you know and when you think that you know but you don't. *Bulletin of the Psychonomic Society.*, *15*(41–44).

Zechmeister, E. B., & Shaughnessy, J. J. (1980b). When you know that you know and when you think that you know but you don't.

Bulletin of the Psychonomic Society, *15*, 41–44.

Zhang, X., Yang, J., Wang, R., & Li, P. (2020). A neuroimaging study of semantic representation in first and second languages. *Language, Cognition and Neuroscience*, *35*(10), 1223–1238. https://doi.org/10.1080/23273798.2020.1738509

Zimmerman, B. J. (1986). Development of self-regulated learning: Which are the key subprocesses? *Contemporary Educational Psychology*, *16*(307–313).

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). Academic Press. https://www.elsevier.com/books/%0Ahandbook-of-self-regulation/boekaerts/978-0-12-109890-2

Zimmerman, B. J. (2008). Investigating Self-Regulation and Motivation: Historical Background, Methodological Developments, and Future Prospects. *American Educational Research Journal*, *45*(1), 166–183. https://doi.org/doi:10.3102/0002831207312909.

Zimmerman, B. J., & Martinez-Pons, M. (1986). Development of a structured interview for assessing student use of self-regulated learning strategies. *American Educational Research Journal*, *23*, 614–628.

Zimmerman, B. J., & Martinez-Pons, M. (1988). Construct validation of a strategy model of student self-regulated learning. *Journal of Educational Psychology*, *80*(3), 284–290.

Zimmerman, B. J., & Schunk, D. H. (1989). *Self-regulated learning and academic achievement: Theory, research, and practice.* Springer-Verlag.

Zimmerman, B. J., & Schunk, D. H. (2011). Self-regulated learning and performance: An introduction and an overview. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 1–12, Chapter xiv, 484 Pages). Routledge/Taylor & Francis Group PP - New York, NY. https://www.proquest.com/books/self-regulated-learning-performance-introduction/docview/1074674955/se-2?accountid=14542

Zimmerman, C. A., & Kelley, C. M. (2010). “I'll remember this!” Effects of emotionality on memory predictions versus memory performance. *Journal of Memory and Language*, *62*, 240–253.

Zirnstein, M., van Hell, J. G., & Kroll, J. F. (2019). Cognitive control and language ability contribute to online reading comprehension: Implications for older adult bilinguals. *International Journal of Bilingualism*, *23*(5), 971–985. https://doi.org/10.1177/1367006918763143

Zivian, M. T., & Darjes, R. W. (1983). Free-recall by in-school and out-of- school adults – Performance and metamemory. *Developmental Psychology*, *19*(4), 513–520. https://doi.org/https://doi.org/10.1037/0012-1649.19.4.513

Zusho, A. (2017). Toward an Integrated Model of Student Learning in the College Classroom. *Educational Psychology Review*, *29*, 301–324. https://doi.org/10.1007/s10648-017-9408-4

Supporting Information

Below, you will find the following information:

S1 Table. Experiment 1: Estimated means (and standard deviations) for hits, false alarms, misses and correct rejections by language, font type and block order. Means of 0.00 is a matter of decimal places: a. 0.003125, b. 0.003378378.

S2 Table. Experiment 2: Estimated means (and standard deviations) for hits, false alarms, misses and correct rejections by language, concreteness and block order.

S3 Table. Experiment 3: Estimated means (and standard deviations) for hits, false alarms, misses and correct rejections by language, type of list and block order. Means of 0.00 is a matter of decimal places: a. 0.004166667.

S4 Table. Experiment 2: Concreteness ratings of each list. Mean (SD). Note that concreteness ratings for English words were based on Brysbaert et al. (2014) using a 5-point scale, whereas values for Spanish words were based on LEXESP (Sebastián et al., 2000) using a

7-point scale. Thus, the descriptive statistics are in different scales. However, the criteria to consider a word abstract or concrete was equivalent for both data set as explained in the manuscript.

S5 Table. Experiment 3: Frequency and number of letters of each study list. Mean (SD). Log10_FRQ = mean estimated frequency; LEN_L = mean estimated number of letters. Unrelated words were used both for unrelated-word studied lists and for new words in the recognition test. All *p* values were above 0.05. ($p>0.05$) showing there was no significant difference.

S6 Table. Experiment 4-5: Example of a high- and low-cohesion English version of a text. 1 = noun repetition, 2 = use of connectors, 3 = use of passive voice, 4 = pronoun replacement to create ambiguity.

S7 Table. Experiment 4-5: Texts and references they were extracted from.

S1 Table.

Experiment 1: Estimated means (and standard deviations) for hits, false alarms, misses and correct rejections by language, font type and block order.

			Block order			
			L1-first		L2-first	
			Language		Language	
			L1	L2	L1	L2
Condition	Easy-to-read font	Hits	0.90 (0.31)	0.89 (0.32)	0.86 (0.39)	0.87 (0.38)
		Misses	0.09 (0.27)	0.11 (0.29)	0.13 (0.38)	0.10 (0.35)
		No response	0.01 (0.33)	0.01 (0.32)	0.01 (0.24)	0.03 (0.28)
	Difficult-to-read font	Hits	0.89 (0.32)	0.88 (0.32)	0.88 (0.38)	0.85 (0.40)
		Misses	0.10 (0.29)	0.12 (0.30)	0.11 (0.36)	0.13 (0.38)
		No response	0.01 (0.32)	0.00[a] (0.32)	0.00[b] (0.24)	0.02 (0.26)
	New items	Correct rejections	0.91 (0.28)	0.90 (0.30)	0.79 (0.41)	0.80 (0.40)
		False alarms	0.08 (0.26)	0.09 (0.28)	0.19 (0.39)	0.17 (0.38)
		No response	0.11 (0.32)	0.12 (0.32)	0.07 (0.25)	0.08 (0.27)

Note: Means of 0.00 is a matter of decimal places: a. 0.003125, b. 0.003378378.

S2 Table.

Experiment 2: Estimated means (and standard deviations) for hits, false alarms, misses and correct rejections by language, concreteness and block order.

			Block order			
			L1-first		L2-first	
			Language		Language	
			L1	L2	L1	L2
Condition	Concrete	Hits	0.89 (0.31)	0.90 (0.30)	0.87 (0.33)	0.91 (0.29)
		Misses	0.09 (0.28)	0.09 (0.28)	0.11 (0.31)	0.07 (0.26)
		No response	0.02 (0.15)	0.01 (0.09)	0.01 (0.11)	0.02 (0.12)
	Abstract	Hits	0.82 (0.39)	0.85 (0.36)	0.79 (0.41)	0.86 (0.35)
		Misses	0.13 (0.34)	0.14 (0.34)	0.20 (0.39)	0.12 (0.32)
		No response	0.05 (0.22)	0.01 (0.12)	0.02 (0.15)	0.03 (0.16)
	New items	Correct rejections	0.85 (0.36)	0.82 (0.38)	0.84 (0.37)	0.87 (0.33)
		False alarms	0.12 (0.33)	0.16 (0.37)	0.15 (0.35)	0.10 (0.29)
		No response	0.03 (0.17)	0.01 (0.11)	0.02 (0.14)	0.03 (0.17)

S3 Table.

Experiment 2: Concreteness ratings of each list. Mean (SD).

		Spanish – L1	English – L2
Mean (SD) concreteness rating	Concrete	5.78 (0.5)	4.56 (0.4)
	Abstract	3.76 (0.69)	2.57 (0.67)
	Total	4.7 (1.18)	3.64 (1.13)
Min. concreteness rating	Concrete	4.8	3.64
	Abstract	2.22	1.25
Max. concreteness rating	Concrete	6.66	5
	Abstract	4.79	3.54

Note: concreteness ratings for English words were based on Brysbaert et al. (2014) using a 5-point scale, whereas values for Spanish words were based on LEXESP (Sebastián et al., 2000) using a 7-point scale. Thus, the descriptive statistics are in different scales. However, the criteria to consider a word abstract or concrete was equivalent for both data set as explained in the manuscript.

S4 Table.

Experiment 3: Estimated means (and standard deviations) for hits, false alarms, misses and correct rejections by language, type of list and block order.

			Block order			
			L1-first		L2-first	
			Language		Language	
			L1	L2	L1	L2
Condition	Words grouped into semantic categories	Hits	0.77 (0.42)	0.87 (0.34)	0.86 (0.35)	0.91 (0.29)
	Unrelated words	Misses	0.21 (0.41)	0.12 (0.33)	0.13 (0.34)	0.09 (0.28)
		No response	0.02 (0.14)	0.01 (0.10)	0.01 (0.09)	0.01 (0.10)
		Hits	0.68 (0.47)	0.73 (0.45)	0.70 (0.46)	0.86 (0.35)
		Misses	0.29 (0.45)	0.27 (0.44)	0.29 (0.45)	0.14 (0.34)
		No response	0.03 (0.17)	0.00[a](0.06)	0.01 (0.10)	0.01 (0.07)
	New items	Correct rejections	0.71 (0.45)	0.81 (0.39)	0.83 (0.38)	0.86 (0.35)
		False alarms	0.26 (0.44)	0.18 (0.38)	0.17 (0.37)	0.13 (0.33)
		No response	0.04 (0.18)	0.01 (0.10)	0.01 (0.10)	0.01 (0.11)

Note: Means of 0.00 is a matter of decimal places: a. 0.004166667.

S5 Table.

Experiment 3: Frequency and number of letters of each study list. Mean (SD).

		Log10_FREQ	LEN_L
English–L2	Semantic-category words	1.5 (0.7)	5.5 (2.0)
	Unrelated words	1.1 (0.5)	5.5 (1.2)
Spanish–L1	Semantic-category words	1.3 (0.7)	6.0 (1.7)
	Unrelated words	1.1 (0.5)	5.8 (1.1)

Note: Log10_FRQ = mean estimated frequency; LEN_L = mean estimated number of letters. Unrelated words were used both for unrelated-word studied lists and for new words in the recognition test. All *p* values were above 0.05. (p>0.05) showing there was no significant difference.

S6 Table.

Experiment 4-5: Example of a high- and low-cohesion English version of a text.

Title: Traits of mammals
High-cohesion version
Over many years[1], mammals have developed different kinds of specialized teeth, which has enabled them to be successful. These different teeth[2] allow mammals[2] to eat many different kinds of food. This trait also helps them to live in different kinds of environments. Basically[1], there are four types of teeth in mammals[2]: incisors, canines, premolars and molars. The number and shape of each of these types of teeth[1] are related to the kind of food the mammal eats. Meat-eating mammals, such as wolves and lions, have long, pointed canine teeth, that are used for cutting. Plant-eating mammals, such as horses and cows, have large, flat premolars and molars. These teeth are used for grinding plant materials. Mammals such as we, humans, have many different kinds of teeth, which help us eat the many different kinds of food in their diets.
Low-cohesion version
Mammals have very specialized teeth that have made[3] them be successful and allowed them to eat many different kinds of food in different kinds of environment. For the four types of teeth there are, the number and shape of each of them are related to the kind of food the mammal eats. Meat-eating mammals, such as wolves and lions, have long, pointed canine teeth that are used for tearing. Their[4] incisors are chisel-shaped and are used for cutting. Those[4] that are plant-eating, such as horses and cows, have large, flat premolars and molars. These teeth are used for grinding plant materials. We humans have many different kinds of teeth as we eat many different kinds of food in the diet.
Open-ended questions
What type of teeth do meat-eating mammals have? Long and pointed canine teeth. What type of teeth do plant-eating mammals have? Large and flat premolars and molars What does having different kind of specialized teeth allow mammals? These different teeth allow mammals to eat many different kinds of food and to live in different kinds of environments.

Note: 1 = noun repetition, 2 = use of connectors, 3 = use of passive voice, 4 = pronoun replacement to create ambiguity.

S7 Table.

Experiment 4-5: Texts and references they were extracted from.

	Text	Reference
1	Networking and the Internet	Extracted from Brookshear (2005), a book used by Gasparinatou & Grigoriadou (2013).
2	Security – Forms of attack	
3	Artificial Intelligence	
4	Geologic Processes	Adapted from Ariel, Karpicke, Witherby, & Tauber (2020).
5	Inorganic Substances	
6	Crystalline Solids	
7	Elements	
8	Compounds	
9	Traits of Mammals	Adapted from McNamara, Kintsch, Songer, & Kintsch (1996).
10	Heart Disease	
11	Heat Distribution in Animals	Adapted from Ozuru, Kurby, & McNamara (2012).
12	Africa	Extracted from the interactive book *World Geography. Michigan Open Project* (Dufort et al., 2018).
13	Apartheid	
14	Antarctica	
15	The global financial crisis	Extracted from *The History Book. Big ideas simply explained* (Grant et al., 2016)
16	Global issues, local perspectives	Extracted from *The Sociology Book. Big ideas simply explained* (Thorpe et al., 2015)
17	Climate change	
18	Environmental pollution	Extracted from *The Ecology Book. Big ideas simply explained* (Schroeder et al., 2019).
19	Effects of pollution on health	
20	Emotional Intelligence	Extracted from *The Business Book. Big ideas simply explained* (Marcouse et al., 2014).
E1	Supermarkets	Extracted from a FCE reading test.
E2	Charlie Chaplin	Extracted from the book *The History Book. Big ideas simply explained* (Grant et al., 2016)