

TRABAJO FIN DE MÁSTER

Revisión sistemática de la “Retrieval Practice” como herramienta de aprendizaje en Educación Secundaria.



Estudiante: Loubna Er-rahmouny Mouzouri

Especialidad: Formación y Orientación Laboral

Tutor/a: Adolfo Aracil Marco

Curso académico: 2023-24

ÍNDICE

1. Resumen y palabras clave.....	2
2. Introducción.....	3
3. Método.....	4
4. Resultados.....	6
5. Discusión y conclusiones.....	18
6. Propuesta de intervención.....	23
7. Referencias.....	25

Tablas:

Tabla 1	7
Tabla 2	9
Tabla 3	14

Figuras

Figura 1.....	5
---------------	---



1. Resumen y palabras clave:

Resumen:

La *Retrieval Practice* (RP) es una estrategia de aprendizaje activo que surgió hace más de cien años y consiste en evocar de la memoria la información previamente aprendida. Esta revisión sistemática, basada en la metodología PRISMA, tiene como finalidad definir la RP, evaluar su eficacia en alumnado de ESO y Bachillerato en entornos educativos reales, y analizar las diferencias según el material impartido. Para ello, se utilizaron las bases de datos de ERIC, Pubmed, PsycINFO y Web of Science, identificando un total de 126 artículos publicados entre 2019 y 2024. De estos, solo 13 cumplían los criterios de inclusión establecidos.

Los resultados indican que la RP generalmente implica estudiar un contenido, realizar pruebas sobre él, y tras un periodo de tiempo, volverse a examinar. Además, la RP mejora el rendimiento académico en materias de Ciencias Naturales y de Aprendizaje de Vocabulario Extranjero, no obstante, en Ciencias Sociales los resultados son inconsistentes.

La principal limitación de esta revisión radica en la escasez de revisiones que analizan solamente alumnado de ESO y Bachillerato, lo que dificulta su comparación. Asimismo, otra limitación es la falta de consenso en la terminología de la RP y que hay estudios que comparan dos condiciones de RP en lugar de compararla con otras estrategias. No obstante, la RP es una estrategia de aprendizaje prometedora para estudiantes de Secundaria y Bachillerato, aunque es necesario más investigaciones para comprender completamente su eficacia y optimizar su uso en el aula.

Palabras clave: estrategia de aprendizaje; proceso de enseñanza-aprendizaje; prueba de evocación; rendimiento académico; estudiante de secundaria; *testing effect*

Abstract:

Retrieval Practice (RP) is an active learning strategy that originated over one hundred years ago. It involves extracting information from memory that was previously learned. This systematic review, based on the PRISMA methodology, aims to define the RP, assess its efficacy among secondary students in the classroom and evaluate differences according to the material used. To do so, it included ERIC, Pubmed, PsycINFO and Web of Science databases, identifying 126 articles published between 2019 and 2024. Of these, 13 articles fulfilled the inclusion requirements.

The results indicate that RP involves studying the material, engaging in tests, and taking a test after a delay. Moreover, the RP improves academic performance in natural sciences and foreign vocabulary learning, however, the results in social science are inconsistent.

The major limitation of this review was the limited number of papers focused on only secondary students, which makes it difficult to make a comparison with this review. Another limitation was the lack of consensus regarding the concept of RP, and some studies compare two RP conditions rather than comparing with another learning strategy. However, the RP is a promising strategy for secondary students, although further research is needed to better understand its efficacy and to optimize its use in the classroom.

Keywords: learning strategy, teaching-learning process, academic performance, high school, testing effect

2. Introducción

La investigación en Educación permite difundir conocimiento sobre esta área, lo que mejora su calidad, dotando al profesorado de herramientas eficaces para proveer al alumnado de estrategias que propicien su aprendizaje competencial, autónomo, significativo y reflexivo. A lo largo de los años, se han investigado un gran número de estrategias, entre ellas la *retrieval practice (RP)*, técnica de aprendizaje activo que consiste en recuperar o evocar la información previamente aprendida de la memoria, fortaleciendo así la retención de la información a largo plazo y reduciendo la probabilidad de olvido (Agarwal et al., 2018). Cuando la persona logra evocar esa información con éxito se produce el efecto conocido como *testing effect*. De hecho, el estudio de Ye et al. (2020) muestra que la RP facilitaría la actualización de la memoria, ya que se vio que estructuras de la corteza prefrontal medial están involucradas en el proceso de integración, diferenciación y consolidación de la memoria, en comparación con el reestudio. Igualmente, en el estudio llevado a cabo por Catherine-Noémie et al. (2022), encontraron que la RP produce cambios rápidos en las representaciones neocorticales de la información, tanto de adultos jóvenes como mayores. Asimismo, vieron activaciones generalmente en la corteza prefrontal, la corteza parietal y la corteza temporal.

Aunque el estudio de la RP se remonta a más de un siglo atrás, esta ha recibido especial atención en la última década (Yang et al., 2021), demostrando ser una estrategia que aumenta el rendimiento académico en diversos niveles educativos y áreas de contenido (Agarwal et al., 2021). En un estudio reciente, Langbroek, et al. (2024) realizaron una *umbrella review* y encontraron que la RP mejoraba el rendimiento académico cuando se trataba del aprendizaje factual y conceptual.

No obstante, a pesar de los beneficios demostrados de la RP, existen estudios que indican que el estudiantado prefiere utilizar estrategias como el reestudio antes que la RP a la hora de estudiar el contenido del que se va a examinar. Cuando se ha visto que la estrategia de reestudio no mejora el rendimiento académico en el aula (Callender y McDaniel, 2009). Además, aquel alumnado que emplea la RP lo hace tras haber estudiado y principalmente para evaluar su aprendizaje en lugar de mejorarlo (Dirkx et al., 2019; Tullis y Maddox, 2020).

Actualmente, las únicas revisiones realizadas sobre la RP en un contexto educacional real son las revisiones de Yang et al. (2021), Agarwal et al. (2021) y Moreira et al. (2019). No obstante, estas abarcan una gran variedad de alumnado de diversas etapas. En cuanto a las demás revisiones se centran en aquellos estudios realizados en el laboratorio y/o entorno educativo como es el caso de Adesope et al. (2017), que mostró que la RP conllevaba un mejor rendimiento en estudiantes de primaria, secundaria y estudiantes de post-secundaria. Cabe destacar, como menciona Moreira et al. (2019), que los experimentos llevados a cabo en laboratorios difieren de los que se realizan en un entorno educativo real, como la cantidad de información que el estudiantado necesita aprender, la motivación, los estímulos distractores en ambos contextos, entre otros aspectos. Por ello, es necesario conocer la eficacia de la RP en contextos educativos actuales reales.

La búsqueda inicial muestra que la perspectiva específica del tema aún no se ha abordado en la literatura actual, es decir, la RP centrada en alumnado de Educación Secundaria y

Bachillerato en entornos educativos reales. Por lo tanto, el objetivo de esta revisión sistemática es identificar si la RP es eficaz en alumnado de educación secundaria y bachillerato, es decir, -estudiantado con un rango comprendido entre 12 a 18 años-, por medio de estudios realizados en un contexto educativo real, es decir, estudios llevados a cabo en el centro educativo donde el alumnado estudia, no en laboratorio. Las preguntas que se esperan responder con dicha investigación son las siguientes: a) ¿En qué consiste la RP?; b) ¿Es eficaz la RP en alumnado de Educación Secundaria y Bachillerato?; y, c) ¿Existen diferencias en función del contenido impartido?

3. Método

Con el propósito de llevar a cabo esta revisión, se aplicó el método PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses, Page et al., 2021). Se establecieron los siguientes criterios para determinar la inclusión y exclusión de artículos:

Criterios de inclusión:

- Artículos con una muestra de estudiantes de Secundaria, es decir, aquellos con edades entre 12 y 18 años.
- Artículos publicados entre 2019 y 2024.
- Artículos relacionados con el objetivo de la revisión, que permitan responder a las preguntas sobre en qué consiste la RP, su estado actual y su eficacia en entornos educativos.
- Artículos desarrollados en un entorno educativo, es decir, en el centro educativo donde se da el proceso de enseñanza-aprendizaje del alumnado, no en laboratorio.

Criterios de exclusión:

- Libros, revisiones sistemáticas, metaanálisis, *umbrella reviews* y tesis.
- Artículos sin acceso al texto completo.
- Población que presente algún problema de aprendizaje.

La Figura 1 muestra el diagrama de flujo de la revisión. La búsqueda exhaustiva se realizó el 3 de abril de 2024 en cuatro bases de datos: ERIC (n=35), PsycInfo (n=25), Web of Science (n=31) y Pubmed (n=54). Se emplearon diferentes estrategias de búsqueda según la base de datos. En ERIC se utilizaron varias fórmulas con los filtros "últimos cinco años" y "Secondary School", así como combinaciones de términos como "*Retrieval Practice*" y "*Testing Effect*" con el año 2019. En las demás bases de datos, se aplicaron fórmulas similares centradas en el período de 2019 a 2024. Esto resultó en 126 artículos. Cabe destacar que este estudio es de tipo académico, por lo tanto, como estudiante se puede acceder a un gran número de artículos en las bases de datos por medio de los recursos que ofrece la universidad. También, en ocasiones, cuando no se podía acceder al artículo en la base de datos utilizada, se ha utilizado PsycArticles.

Una vez recopilados los artículos, se procedió a eliminar los duplicados (n=57) y, mediante la revisión de títulos y resúmenes, se excluyeron aquellos que no cumplían con el objetivo de la revisión o no estaban relacionados con la RP (n=30), revisiones (n=4) y aquellos que no

cumplían con los criterios de población. Posteriormente, se realizó un segundo cribado, que consistió en primer lugar en un análisis por medio de la lectura del título y el resumen de los 30 artículos con la finalidad de tomar una decisión consensuada con el tutor sobre qué artículos incluir, lo que conllevó al descarte de 17 artículos debido a las siguientes razones: no se alineaban con el objetivo de la revisión ($n=9$), eran tesis doctorales ($n=6$) y hubo dos artículos que no se podía acceder al texto completo. El segundo análisis consistió en la lectura completa de los artículos. A partir de esta lectura, se descartó uno debido al hecho de que no se alineaba con el objetivo de esta revisión. Cabe destacar que se intentó contactar con el autor de uno de los artículos (Barenberg y Dutke, 2019) a través de *ResearchGate*, pero no se recibió respuesta.

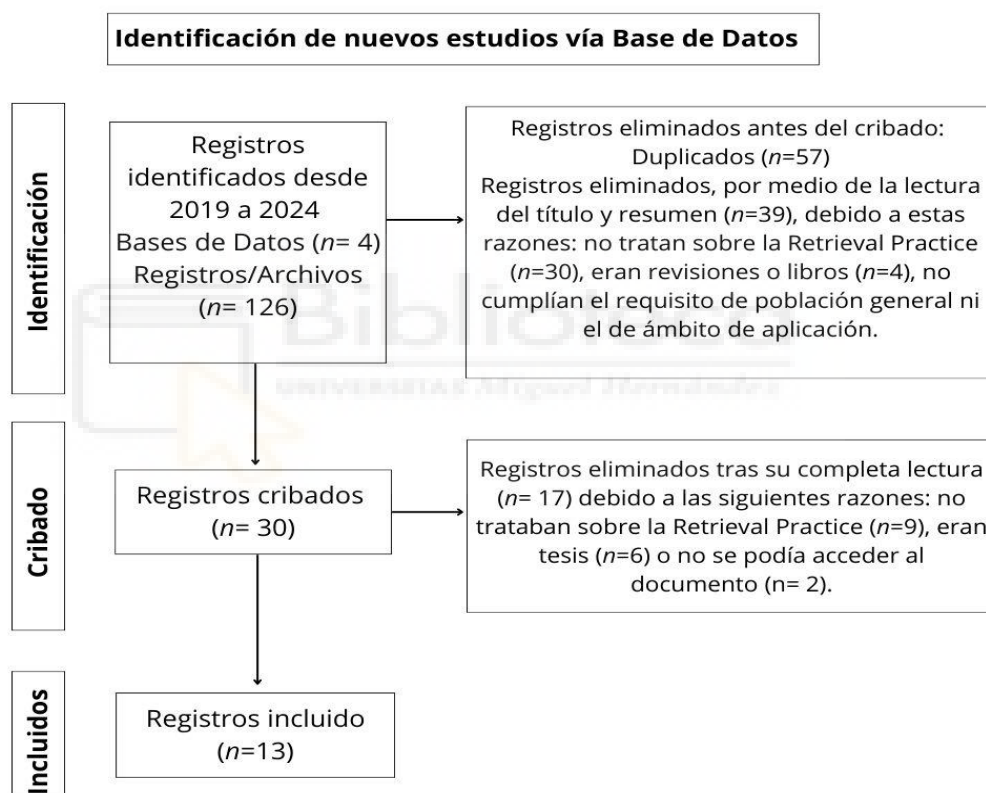


Figura 1. Diagrama de flujo que presenta el proceso de cribado de los artículos.

4. Resultados

Tras la búsqueda realizada se obtuvieron 13 artículos en total. De ellos, dos presentan 3 experimentos (van den Broek et al., 2019; van den Broek et al., 2021) y uno presenta 2 (Jonsson et al., 2021). Por lo tanto, al final, este estudio lo forman 18 estudios en total.

El 38.89% de los estudios se han realizado en Países Bajos, el 27.78% en Suecia, el 11.11% en Alemania, y el resto en Hungría, Estados Unidos, Australia y Nueva Zelanda. Respecto al sexo, todos los estudios están constituidos por una muestra mixta, a excepción del estudio llevado a cabo por Leggett et al. (2019) debido a que el centro educativo solo era para hombres. Asimismo, cabe comentar que hay un estudio en el que no se menciona el sexo de la muestra (Szeibert *et al.*, 2023).

Respecto al diseño de los estudios, la mayoría de ellos se basan en un diseño intra-sujeto (77.78%), y el resto son estudios experimentales (exploratorio= 5.56%; aleatorio= 5.56%; experimental= 5.56%; y cuasiexperimental= 5.56%).

Para analizar los artículos de esta revisión sistemática se agruparon según el contenido estudiado: estudios en Ciencias Naturales (n = 2, Tabla 1), Ciencias Sociales (n = 4, Tabla 2) y aprendizaje de vocabulario extranjero (n = 7, Tabla 3).

Estudios de Ciencias Naturales:

Respecto a los estudios de Ciencias Naturales, se ha observado que el procedimiento para llevar a cabo la RP consistió en aprender el material, realizar una prueba al final de la sesión para practicar el contenido aprendido, así como poner en práctica la RP, y seguidamente les han proporcionado *feedback*. También se utilizaron evaluaciones formativas en ambos estudios. No obstante, el formato fue diferente, ya que Molin et al. (2021) utilizaron preguntas de opción múltiple y el sistema de *clickers*, mientras que Szeibert et al. (2023), las pruebas consistían en ejercicios teóricos y prácticos. Asimismo, en ambos estudios se ha visto que la RP mejora el rendimiento académico del alumnado en comparación con la enseñanza tradicional (Molin et al., 2021) o sin RP (Szeibert et al., 2023).

Las diferencias entre estos dos estudios se encuentran en que Molin et al. (2021) investiga también el papel de la ansiedad, obteniendo como resultado que las evaluaciones formativas reducían el nivel de ansiedad y lo que podría actuar como mediador, mientras que en el otro estudio se pretendía reducir la brecha en el rendimiento académico debido al nivel socioeconómico, mostrando por medio de su estudio que aplicando la RP a alumnado con un nivel socioeconómico puede rendir mejor que alumnado con el mismo nivel socioeconómico e igual que el estudiantado con un nivel socioeconómico alto.

Tabla 1. Resumen de los estudios de Ciencias Naturales.

Estudio	Población	Intervención	Comparación	Resultados
Molin et al. (2021).	<p>$n = 139$ estudiantes de Física.</p> <p>Edad: 16.2 ± 0.69 años.</p> <p>Sexo: 72 hombres; 67 mujeres.</p> <p>Nivel socioeconómico: no indicado.</p> <p>País: Países Bajos.</p> <p>Grupo control: $n = 66$.</p> <p>Grupo intervención: $n = 73$.</p>	<p>Experimento aleatorio.</p> <p>Grupo control: clases tradicionales.</p> <p>Grupo intervención: evaluación formativa repetida en el aula y <i>feedback</i> inmediato.</p> <p>Duración: 17 semanas, tres veces a la semana.</p>	<p>Comparación entre grupos, al final de la intervención:</p> <p>Resultados académicos en una prueba final.</p> <p>Nivel de ansiedad.</p>	<p>Las evaluaciones formativas mejoran el rendimiento en Física en comparación con la enseñanza tradicional.</p> <p>Las evaluaciones formativas reducen de manera significativa la ansiedad en Física en comparación con la enseñanza tradicional.</p> <p>La ansiedad es una variable mediadora entre las evaluaciones formativas y el rendimiento académico.</p>
Szeibert et al. (2023)	<p>$n = 57$ estudiantes de Matemáticas.</p> <p>Edad: no indicada (9º grado).</p> <p>Sexo: no indicado.</p> <p>Nivel socioeconómico:</p> <ul style="list-style-type: none"> Grupo experimental, se encuentra en el puesto 506 del ranking de mejor centro a nivel nacional. Grupo élite: se encuentra en el puesto 14 del ranking de mejor centro a nivel nacional. <p>País: Hungría.</p> <p>Grupo control: n (formación profesional) = 9; n (grupo de élite) = 25.</p> <p>Grupo intervención: $n = 23$.</p>	<p>Estudio de caso exploratorio.</p> <p>Grupo control: RP con <i>feedback</i>.</p> <p>Grupo intervención: lección normal (corregir deberes, aprender algo nuevo y practicar el nuevo material por medio de tareas y problemas).</p> <p>Duración: Formación profesional: 4 semanas. Escuela de élite: 6 semanas.</p>	<p>Comparación de grupos, al final de la intervención.</p> <p>Resultados académicos en una prueba final.</p>	<p>La RP en el grupo experimental de formación profesional tuvo éxito en la unidad de Geometría, ya que obtuvieron puntuaciones significativamente más altas en cada tarea, comprendieron el contenido y cometieron menos errores en comparación con el grupo control de formación profesional.</p> <p>El grupo experimental obtuvo una puntuación media tan alta como la del grupo élite en la prueba final.</p>

Estudios en Ciencias Sociales:

Respecto al grupo de Estudios Sociales, el procedimiento para realizar la RP es el mismo que el grupo anterior: se proporciona material sobre el contenido que se pretende enseñar, luego se evalúa dicho contenido mediante una prueba y, tras un periodo, se vuelve a evaluar con una prueba final.

En cuanto a las similitudes se puede observar que el alumnado recibe *feedback* tras las pruebas en los estudios de Agarwal (2019) y Legget et al. (2019), aunque cabe destacar que en el primer caso se proporciona solamente la respuesta correcta, mientras que, en el segundo, el estudiantado debe corregir las respuestas erróneas, escribiendo la respuesta correcta que se le indica. Asimismo, la RP mejoró el rendimiento académico en comparación con la condición de ninguna prueba (Agarwal, 2019), con lectura del material (Legget et al., 2019) y con copiar el material (Rowley y McCrudden, 2020).

Por otro lado, hubo ciertas diferencias entre estos estudios como que en el estudio de Agarwal (2019) se llevó a cabo un pre-test antes de iniciar la intervención y se utilizó un formato de preguntas de opción múltiple con *clickers*. Además, su objetivo fue analizar si el aprendizaje factual debe preceder a la habilidad basándose en la teoría de la Taxonomía de Bloom y sus resultados fueron los siguientes: el rendimiento aumentaba a lo largo de la intervención (pre-test, post-test y test de revisión); las pruebas mixtas (preguntas de orden inferior y superior) y de orden superior tuvieron los mismos efectos; y la RP con preguntas de orden superior conllevaba significativamente mayor rendimiento en comparación con la condición de ninguna prueba.

Respecto al estudio llevado a cabo por Jägerskog et al. (2019), la prueba final se llevó a cabo inmediatamente tras la intervención y a la semana. No obstante, en este caso la RP no mejoró el rendimiento y al combinar la presentación visuoverbal con esta estrategia tampoco mejoró el rendimiento, más allá del efecto del aprendizaje visuoverbal.

Finalmente, el estudio realizado por Leggett et al. (2019) encontró que la RP con *feedback* mejoró la retención de los hechos en comparación con la lectura, independientemente del rendimiento de las pruebas de revisión, la diversión de la actividad y sus creencias de efectividad.

Tabla 2. Resumen de los estudios de Ciencias Sociales.

Estudio	Población	Intervención	Comparación	Resultados
Agarwal (2019).	<i>n</i> = 142 estudiantes – 88 estudiantes de Historia mundial. Edad: 11.58 años Sexo: 40 hombres; 48 mujeres. Nivel socioeconómico: no indicado. País: La región central de Estados Unidos.	Diseño intra-sujeto. 2 grupos intervención: RP pruebas de orden superior RP pruebas de orden superior e inferior (mixto). Duración: alrededor de 10 días.	Comparación entre grupos, al final de la intervención, y entre pretest, post-test y revisión. Resultados académicos.	El rendimiento aumentó en las tres pruebas tanto para la condición de cuestionarios mixtos como para los de orden superior. Por lo tanto, la RP aumentó el rendimiento en comparación con la condición de sin cuestionarios. La RP llevada a cabo mediante cuestionarios de orden superior llevó a una mejora significativa en el rendimiento de las pruebas de orden superior en comparación con la condición sin cuestionarios. La prueba mixta produjo el mayor nivel de rendimiento tanto en las pruebas finales de orden inferior como de orden superior.
Jägerskog, et al. (2019).	<i>n</i> = 133 estudiantes de Ciencias Sociales (Modelo de memoria de Atkinson-Shiffrin). Edad: 17.3 ± 0.54 años. Sexo: 64 hombres; 69 mujeres. Nivel socioeconómico: no indicado. País: Suecia.	Diseño intra-sujeto. 2 condiciones de formato de presentación (presentación visuoverbal y verbal). 2 estrategias de estudio: Reestudio y RP. 2 intervalos de intervención: Inmediato y tras una semana. Duración: 2 semanas	Comparación entre grupos, inmediatamente y al final de la intervención: Resultados académicos en dos pruebas.	La presentación visuoverbal conlleva a un recuerdo notablemente mayor, independientemente del intervalo de retención. Se observó un beneficio de la RP en términos de menor olvido. No obstante, no superó al reestudio en ningún intervalo de retención. La presentación visuoverbal mostró puntuaciones significativamente más altas en la prueba de transferencia en comparación con la presentación verbal. No se encontró una interacción significativa entre las dos estrategias para los datos de transferencia. Los datos mostraron que no existe una interacción estadísticamente significativa entre la estrategia de estudios y el formato de presentación visual.
Leggett, Burt y Carroll (2019).	<i>n</i> = 89 estudiantes – 80 estudiantes de Geografía (seguridad alimentaria). Edad: 14/15 años. Sexo: hombres. Nivel socioeconómico: no indicado. País: Australia.	Diseño intra-sujeto. Grupos de intervención: RP con pistas con <i>feedback</i> RP sin pistas con <i>feedback</i> Lectura con pistas Lectura sin pistas Duración: 2 sesiones, una semana entre la 1ª y 2ª sesión.	Comparación entre las condiciones: –RP con pistas –RP sin pistas –Lectura con pistas –Lectura sin pistas Resultados en una prueba final.	RP con <i>feedback</i> mejora la retención de los hechos más que la lectura, independientemente del rendimiento de la práctica, el disfrute de la actividad o la creencia de su efectividad. No se observó ningún efecto consistente al proporcionar pistas durante la revisión.



MASTERPROF UMH

UNIVERSITAS Miguel Hernández

MÁSTER UNIVERSITARIO EN FORMACIÓN DEL PROFESORADO ESO Y BACHILLERATO, FP Y ENSEÑANZAS DE IDIOMAS

Rowley &
McCrudden
(2020).

$n = 39$ estudiantes de Ciencias (sobrepesca).

Edad: 11 a 13 años.

Sexo: 20 hombres; 19 mujeres.

Nivel socioeconómico: medio bajo (7 en una escala donde 1 es un nivel socioeconómico alto y 10 un nivel socioeconómico bajo).

País: Nueva Zelanda.

Grupo control: $n = 20$.; Grupo intervención: $n = 19$.

Diseño experimental.

Grupo control:

copiar sus apuntes

Grupo intervención:

RP

Duración: 5 días, tres sesiones (1° sesión: tomar notas; 2° sesión: RP o copiar notas; 3° sesión: prueba final).

Comparación entre grupos, al final de la intervención:

Resultados académicos en una prueba final.

El grupo de RP recordó más unidades e ideas que el grupo de copiar.



Estudios de Aprendizaje de Vocabulario Extranjero:

Respecto al grupo de estudios del Aprendizaje de Vocabulario Extranjero, hay un total de 7 artículos. A continuación, se procederá a mencionar las similitudes y diferencias entre ellos.

En cuanto a las similitudes, se observó que en la mayoría de los estudios de este bloque el procedimiento se inició con la familiarización del contenido (fase de codificación), posteriormente se realizaron varios ciclos de ensayo con las palabras a aprender (condición RP, fase de intervención) y, finalmente, se llevó a cabo una prueba final del rendimiento. Todos proporcionaron *feedback* durante la fase de intervención. Asimismo, se dividieron el conjunto de pares de palabras por condición, por lo tanto, cada condición poseía un conjunto de palabras, lo que permitía analizarla independientemente.

En cuanto a las diferencias, se podrían agrupar los estudios en función de si se ha manipulado alguna variable en el proceso, estudio de transferencia y estudios que hayan evaluado características de personalidad y capacidad de memoria de los sujetos.

A continuación, se procederá a mencionar las diferencias y similitudes entre el grupo en el que se ha manipulado alguna variable dentro del proceso de RP.

En cuanto al estudio llevado a cabo por van den Broek et al. (2019), se analizó en 3 experimentos el efecto del *feedback* con pistas ortográficas (se les proporcionaba la primera y última grafema de la palabra), mnemotécnicas (uso de claves para recordar la palabra) y *cognates* (palabras de otros idiomas que son similares fonéticamente o ortográficamente) en comparación con la condición de RP con *feedback* con la respuesta correcta, obteniendo como resultado que el *feedback* con pistas ortográficas, mnemotécnicas y *cognates* no llevaron a un mayor recuerdo de las palabras estudiadas o rendimiento en comparación con el *feedback* con la respuesta correcta.

En cuanto al estudio llevado a cabo por van den Broek et al. (2021), se realizaron tres experimentos en los que se analizaron la introducción de imágenes durante y tras la RP (se proporciona la imagen cuando se lanza la pregunta y en el *feedback*), así como tras la RP (se proporciona la imagen solamente junto al *feedback*) y sin imagen. Por medio de estos experimentos encontraron que la introducción de imágenes durante y tras la RP no mejoraba el rendimiento del alumnado.

En cuanto al estudio llevado a cabo por van den Broek et al. (2023), el procedimiento de proporcionar la prueba fue diferente, ya que se dejaron unos segundos entre la presentación de la pregunta y de las opciones de respuesta, de manera que el/la estudiante pudiese evocar la respuesta. Este tipo de prueba se le llamó preguntas con respuesta de opciones múltiples escalonada (MCQs escalonada). El estudio se basó en 4 experimentos, no obstante, debido a las características de esta revisión se ha tenido en cuenta únicamente el experimento 1 que compara las MCQs escalonada con MCQs estándar en alumnado de

Secundaria. Los resultados que se obtuvieron fueron que, en la fase práctica, el alumnado fue significativamente más rápido a la hora de responder en la condición MCQs escalonada en comparación con la MCQs estándar. Respecto al post-test, las palabras practicadas en la MCQs fueron significativamente mayores en comparación con la otra condición. Además, se observó una relación significativamente superior en el post-test en la MCQs escalonada en comparación con MCQs estándar, pero solo en aquel alumnado que utilizó la RP. Por último, se vio una relación significativa entre la condición y la RP, lo que sugiere que el *testing effect* dependía del uso de la RP.

Otro estudio se centró en investigar la RP en la transferencia (Barenberg et al., 2021). Para ello, la prueba final estaba constituida por una tarea de no transferencia de pregunta corta (palabra alemana → palabra inglesa) y tareas de no transferencia: tarea de pregunta corta invertida (palabra inglesa → palabra alemana), tarea de pregunta corta y fraseo, y tarea de recuerdo por medio de pistas contextuales (rellenar huecos). Los resultados obtenidos fueron que los estudiantes de RP superaron a la condición no-RP en las cuatro tareas y se observó un efecto significativo del *testing effect* en la prueba de rendimiento final, concretamente en la tarea de pregunta corta, de pregunta corta revertida y pregunta corta y fraseo.

Por último, en los siguientes estudios se analizó, además, la relación de las características de personalidad y la capacidad de memoria con la RP. En los tres estudios, se realizó el post-test inmediatamente tras la fase de prueba, a la semana y a las 4 semanas, a excepción de Wiklund-Hörnqvist et al. (2022) que fue inmediatamente y una semana tras la fase de intervención. Asimismo, se observó el *testing effect*, independientemente de la capacidad de memoria, la necesidad de cognición (característica que explica las diferencias en el nivel de motivación y el esfuerzo) y la perseverancia (Bertilsson et al., 2021), independientemente del grupo de capacidad cognitiva (alta, intermedia y baja) y el intervalo de retención (Jonsson et al., 2021) y del grupo de necesidad de cognición (NFC, que está relacionada con su nivel de motivación, alto y bajo) (Wiklund-Hörnqvist et al., 2022).

Además, respecto al estudio de Bertilsson et al. (2021), los pares de palabras de RP fueron significativamente más recordados en comparación con los pares de palabras del reestudio en las pruebas de retención (tras cinco minutos, una semana y 4 semanas),

En cuanto al trabajo de Jonsson et al. (2021), el grupo de capacidad cognitiva baja y RP rindió igual o más en comparación con el grupo de capacidad cognitiva intermedia y de reestudio. Esto mismo ocurrió en el grupo de capacidad cognitiva intermedia y RP en comparación con el grupo de capacidad cognitiva alta y reestudio. También cabe destacar que este estudio presentaba 2 experimentos. El segundo consistió en realizar la prueba final tras una semana en un escáner de resonancia magnética, al igual que en el estudio realizado por Wiklund-Hörnqvist et al. (2022). En ambos, la sesión de post-test consistió en responder, en primer lugar, si sabían, creían que sabían o no sabían la respuesta de la

pregunta. Posteriormente, debían elegir entre varias opciones la letra que correspondía a la segunda letra de la palabra correcta. Esta sesión acababa con una imagen cerebral estructural. En lo que se refiere a resultados de neuroimagen, en ambos estudios, se observó una actividad mayor en las regiones corticales (Jonsson et al., 2021) y subcorticales (Wiklund-Hörnqvist et al., 2022), sobre todo en el hemisferio izquierdo en comparación con la condición de reestudio. Además, no hubo diferencias significativas entre los diferentes grupos de capacidad cognitiva (Jonsson et al., 2021) y entre los diferentes grupos de NFC (Wiklund-Hörnqvist et al., 2022).



Tabla 3. Resumen de estudios sobre el Aprendizaje de Vocabulario Extranjero.

Estudio	Población	Intervención	Comparación	Resultados
van den Broek et al. (2019).	Experimento 1: <i>n</i> = 108 estudiantes → 85 estudiantes para el aprendizaje de vocabulario extranjero. Edad: 14.21; ± 0.77 años. Sexo: 30 hombres; 55 mujeres. Nivel socioeconómico: no indicado. País: Países Bajos.	Diseño intra-sujeto. 2 condiciones: RP con <i>feedback</i> con pistas ortográficas RP con <i>feedback</i> con la respuesta. Duración: dos semanas; dos sesiones.	Comparación entre grupos, al final de la intervención: Resultados académicos en una prueba final.	En la prueba realizada 7 días tras la intervención no hubo diferencias significativas entre las dos condiciones.
	Experimento 2: <i>n</i> = 120 → 90 estudiantes para el aprendizaje de vocabulario extranjero. Edad: 13.96; ±0.76 años. Sexo: 39 hombres; 51 mujeres. Nivel socioeconómico: no indicado. País: Países Bajos.	Diseño intra-sujeto. 2 condiciones: RP con <i>feedback</i> con pistas mnemotécnicas RP con <i>feedback</i> con la respuesta. Duración: dos semanas; dos sesiones.	Comparación entre grupos, al final de la intervención: Resultados académicos en una prueba final.	No existieron diferencias significativas en el número total de palabras recordadas entre las dos condiciones.
	Experimento 3: <i>n</i> = 88 →74 estudiantes para el aprendizaje de vocabulario extranjero. Edad: 13.2; ±0.6 años. Sexo: 30 hombres; 44 mujeres. Nivel socioeconómico: No indicado. País: Alemania.	Diseño intra-sujeto. 2 condiciones: RP con <i>feedback</i> con pistas <i>cognate</i> RP con <i>feedback</i> con la respuesta. Duración: 4 días; 2 sesiones.	Comparación entre grupos, al final de la intervención: Resultados académicos en una prueba final	No hubo diferencias significativas en el número total de palabras recordadas entre las condiciones.
Barenberg et al. (2021).	<i>n</i> = 200 estudiantes →182 estudiantes de inglés.	Diseño cuasiexperimental.	Comparación entre grupos, al final de la intervención:	Los resultados indican un efecto beneficioso de la RP en el aprendizaje de vocabulario en estudiantado de octavo curso.

	<p>Grupo intervención: RP (formato de pregunta abierta) con <i>feedback</i>.</p> <p>Grupo control: reestudio.</p> <p>Prueba final: cuatro formatos diferentes (recuerdo con clave, recuerdo con clave invertida, tarea de recuerdo sugerido y formulación y tarea de recuerdo mediante claves de contexto).</p> <p>Duración: 3 sesiones (pre-experimental; intervención: test final). Una semana entre cada sesión.</p>	<p>Resultados académicos en una prueba final.</p> <p>2 tipos de prueba final: contexto sin transferencia (recuerdo con clave) y con transferencia (recuerdo con clave invertida; tarea de recuerdo sugerido y formulación, y tarea de recuerdo mediante claves de contexto).</p> <p>Dificultad.</p>	<p>En la prueba final, el estudiantado perteneciente al grupo de RP obtuvo más respuestas correctas en todas las tareas que el grupo control.</p> <p>La RP promueve la transferencia entre formatos de pruebas.</p> <p>El formato de prueba mediante señales contextuales no presentó ningún efecto de prueba significativo en los dos grupos, de hecho, no respondieron a todas las preguntas (al parecer la tarea resultó muy difícil).</p> <p>La comparación de la RP en contextos de transferencia y no transferencia mostró efectos de la RP más pequeños en el contexto de transferencia en comparación con el contexto de no transferencia.</p> <p>Dos formatos de prueba del contexto de transferencia (recuerdo con claves y fraseo y recurso por claves del contexto) fueron más difíciles que el formato de prueba del contexto de no transferencia.</p>
<p>Bertilsson et al. (2021).</p> <p>$n = 196$ → 151 estudiantes de 2 programas (Ciencias Naturales y ciencias sociales) para aprender Suajili.</p> <p>Edad: 17.1 ± 0.62 años.</p> <p>Sexo: 82 hombres; 69 mujeres.</p> <p>Nivel socioeconómico: no indicado.</p> <p>País: Suecia.</p>	<p>Diseño intra-sujeto.</p> <p>Dos condiciones: RP (con <i>feedback</i>) Reestudio.</p> <p>Duración: 4 semanas.</p>	<p>Comparación entre condiciones.</p> <p>Resultados académicos en una prueba a los 5 minutos, una y 4 semanas tras la intervención.</p> <p>Memoria de Trabajo.</p> <p>Determinación.</p> <p>Necesidad de cognición.</p>	<p>La retención de los pares de palabras practicados a través de la RP fue significativamente mejor que en el reestudio en las tres pruebas de retención.</p> <p>La RP es una estrategia de aprendizaje efectiva útil, independientemente de la capacidad de memoria de trabajo, necesidad de cognición y determinación.</p>
<p>Jonsson et al. (2021).</p> <p>Experimento 1: $n = 324$ participantes para el aprendizaje de una lengua extranjera.</p> <p>Edad: $17.55; \pm 0.73$ años.</p> <p>Sexo: 163 hombres; 161 mujeres.</p> <p>Nivel socioeconómico: no indicado.</p> <p>País: Suecia.</p>	<p>Diseño intra-sujeto.</p> <p>2 condiciones: RP reestudio</p> <p>3 grupos: capacidad cognitiva (baja intermedia y alta)</p> <p>Duración: 4 semanas; 3 sesiones.</p>	<p>Comparación entre grupos, al final de la intervención:</p> <p>Resultados académicos en una prueba final.</p>	<p>Los resultados mostraron efectos significativos de la RP en los intervalos de retención directos, a corto plazo (1 semana) y a largo plazo (4 semanas) en comparación con el reestudio.</p> <p>También se vio un efecto principal de la capacidad cognitiva, los estudiantes con mayor capacidad cognitiva tuvieron un mejor desempeño en comparación con aquellos con menor capacidad.</p> <p>No se encontró una interacción significativa entre la capacidad cognitiva y las condiciones de aprendizaje.</p>

Experimento 2:
 $n = 81$ participantes del grupo anterior.

Edad: 18.13; ± 0.46 años.

Sexo: 40 hombres; 41 mujeres.

Nivel socioeconómico: no indicado.

País: Suecia.

Este grupo no realizó la prueba junto a los demás grupos tras una y 4 semanas en el experimento 1.

Habilidad cognitiva baja: $n = 21$
 Habilidad cognitiva intermedia: $n = 24$
 Habilidad cognitiva alta: $n = 36$

van den Broek et al. (2021).

Experimento 1:
 $n = 80$ 1º sesión \rightarrow 74 estudiantes que están aprendiendo una lengua extranjera por medio de sustantivos.

Edad: 13.5; ± 1.04 años.

Sexo: 34 hombres; 46 mujeres.

Nivel socioeconómico: no indicado.

País: Países Bajos.

Experimento 2:
 $n = 135$ estudiantes que están aprendiendo una lengua extranjera por medio de *Idioms*.

Edad: 12.30; ± 0.59 años.

Sexo: 84 hombres; 51 mujeres.

Diseño intra-sujeto.

2 condiciones:

RP

reestudio.

3 grupos con diferentes habilidades cognitivas.

Duración: 1 semana; 2 sesiones (1 semana entre 1º y 2º sesión).

Se aplicó la RP con los grupos de capacidad cognitiva tras una semana, utilizando fMRI.

Diseño intra-sujeto.

2 condiciones:

Imágenes durante y tras la RP.

Imágenes tras la RP durante el *feedback*.

Duración: 5 días; 2 sesiones (4 días entre la 1º y 2º sesión).

Diseño intra-sujetos.

2 condiciones:

imágenes durante y tras la RP

sin imágenes

Duración: 3 días; 2 sesiones (2 días entre la 1º y 2º sesión).

Comparación entre grupos, al final de la intervención:

Resultados imágenes cerebrales.

Comparación entre condiciones, al final de la intervención:

Resultados académicos en una prueba final (tres tipos: español al neerlandés; neerlandés a español; neerlandés e imagen al español).

Resultados afectivo-motivacionales.

Resultados juicios sobre el aprendizaje.

Resultados académicos en una prueba final (4 tipos: Transferencia, Recuerdo del *idiom*, recuerdo del *idiom* junto a una imagen y *cloze* test).

Resultados afectivo-motivacionales.

Resultados juicios sobre el aprendizaje.

Se encontró que la interacción entre la capacidad cognitiva y la condición de aprendizaje no fue significativa, por lo tanto, el efecto significativo de la RP fue independiente de la capacidad cognitiva.

Se ha visto una mayor actividad en varias regiones corticales, sobre todo en el hemisferio izquierdo.

Las activaciones más comunes al comparar el aprendizaje entre las dos condiciones se vieron en la ínsula anterior, la circunvolución precentral, la corteza parietal superior y la circunvolución angular.

Se observó un efecto negativo al añadir imágenes durante y tras la RP, no obstante, este resultado es independiente del éxito de la RP.

Las imágenes durante y tras de la RP no mejoraron los resultados afectivo-motivacionales, a excepción de los sentimientos de competencia, y llevaron al estudiantado a sobreestimar su retención.

Los resultados del grupo sin imágenes fueron significativamente más altos que en el grupo de imágenes durante y después de la RP.

Las imágenes no mostraron efectos en la parte afectivo-motivacional, a excepción del sentimiento de competencia.

La mayoría prefieren la condición de imágenes durante y tras la RP y se sobreestiman su desempeño.

Nivel socioeconómico: no indicado.

País: Países Bajos.

Experimento 3:
 $n = 78$ estudiantes que están aprendiendo palabras abstractas de un idioma extranjero.
 Edad: 13.8; ± 0.71 años.
 Sexo: 39 hombres; 39 mujeres.
 Nivel socioeconómico: no indicado.
 País: Países Bajos.

Diseño intra-sujeto.
 3 condiciones:
 imágenes durante y tras la RP
 imágenes tras la RP en el *Feedback*
 sin imágenes

Duración: $\pm 3/4$ días; 2 sesiones (1 / 3 días entre 1º y 2º sesión).

Resultados académicos en una prueba final (2 tipos: francés – neerlandés; francés + imagen – neerlandés).

Los resultados mostraron que agregar imágenes a la RP tenía un efecto negativo en el recuerdo posterior solo cuando las imágenes eran visibles mientras el estudiantado intentaba recuperar las palabras de la memoria, pero no cuando las imágenes eran visibles después que el alumnado recuperase y enviara una respuesta durante la práctica.

También, que los efectos de las imágenes en la RP dependían de si la prueba final provocaba recuerdo con o sin imágenes de la práctica.

Wiklund-Hörnqvist et al. (2022)

$n = 274$ estudiantes que aprenden Suajili.
 Edad: 17.51 ± 0.74 años.
 Sexo: 140 hombres; 137 mujeres.
 Nivel socioeconómico: no indicado.
 País: Suecia.
 Grupo intervención: $n = 204$
 NFC alto: $n = 112$
 NFC bajo: $n = 108$
 Grupo fMRI: $n = 70$.
 Sexo: 32 hombres; 38 mujeres.

Diseño intra-sujeto.
 2 condiciones:
 RP
 Estudio.

2 grupos:
 NFC alto
 NFC bajo.

3ª condición: grupo fMRI (a la semana).

Duración: 1 semana; 2 sesiones (intervención y test de retención).

Comparación entre grupos, al final de la intervención por medio de prueba de retención y fMRI:

Resultados académicos en una prueba final.
 Resultados de la imagen cerebral.

La RP produce mejor retención a largo plazo en comparación con el reestudio, independientemente de NFC.

Niveles altos de NFC están relacionados con un rendimiento más alto, con independencia de la condición de aprendizaje.

Se observó activación de varias regiones cerebrales corticales y subcorticales lateralizadas a la izquierda, independientemente del desempeño.

Las diferencias en la actividad cerebral tras una semana de la intervención fueron independientes del nivel NFC.

van den Broek et al. (2023).

$n = 83$ alumnos --75 que aprenden esperanto.
 Edad: 13.4; ± 0.55 años.
 Sexo: 26 hombres; 49 mujeres.
 Nivel socioeconómico: no indicado.
 País: Países Bajos.

Diseño intra-sujeto.
 Dos condiciones:
 RP con preguntas de opción múltiple escalonada (MCQs escalonada).
 RP con MCQs estándar.
 Duración: 2 sesiones, dos días entre ellas.

Comparación entre las dos condiciones:

Resultados académicos en una prueba final con las 20 palabras de preguntas abiertas.
 Uso de la RP o no.

Los resultados de aprendizaje en el postest fueron significativamente más altos para las palabras practicadas con MCQ escalonada que para las MCQ estándar.

Los estudiantes que utilizaron la RP obtuvieron resultados más altos que aquellos que no lo hicieron en la MCQ escalonada.

5. Discusión y conclusiones

El propósito de esta revisión fue determinar si la RP es eficaz en alumnado de Educación Secundaria y Bachillerato, es decir, en estudiantado cuyas edades oscilan entre 12 y 18 años, con un enfoque exclusivo en estudios realizados en un entorno educativo práctico, excluyendo estudios de laboratorio. Asimismo, se buscaba definir la RP y examinar posibles diferencias en función del contenido impartido. Buscando en diferentes bases de datos se pudo localizar varios trabajos que han permitido abordar estos temas.

En cuanto al procedimiento de la RP, todos los estudios incluyen una fase de estudio, la realización de tareas de RP y un post-test tras un periodo de tiempo que varía de 5 minutos tras la finalización de las tareas de RP, hasta 4 semanas después. Por otro lado, respecto a la eficacia de la RP en entornos educativos actuales, se agruparon los estudios encontrados en esta revisión en: Ciencias Naturales, Ciencias Sociales y Aprendizaje de Vocabulario Extranjero.

En cuanto a los trabajos realizados en Ciencias Naturales, se ha visto que la RP es eficaz, resultados que se alinean con las revisiones realizadas por Agarwal et al. (2021) y Yang et al. (2021), en las que gran parte de los estudios incluidos fueron en asignaturas de ciencias, aunque es necesario mencionar que estas dos revisiones abarcan un rango de edad más amplio. Por lo tanto, serían necesarias otras revisiones específicas sobre el rango de edad analizado para poder comparar. Otro aspecto a destacar es que a pesar de que esta agrupación solo presenta dos estudios, el tamaño de la muestra supera los 100 sujetos, por lo tanto, son datos que hay que tener en consideración.

Respecto a los estudios realizados en asignaturas de Ciencias Sociales, los resultados son inconsistentes, ya que, de los cuatro estudios realizados, uno de ellos no mostró mejoría en el rendimiento académico del estudiantado (Jägerskog et al., 2019). Este resultado puede deberse a la falta de *feedback*. De hecho, Moreira et al. (2019) encontraron en su revisión que uno de los estudios en los que no se observó mejoría en la RP en comparación con el reestudio, tampoco se proporcionó retroalimentación. Además, en el metaanálisis de Yang et al. (2021) se encontró que el *feedback* modulaba significativamente el efecto de las pruebas en el aula. No obstante, cabe destacar que uno de los estudios analizados en esta revisión (Rowley y McCrudden, 2020), el alumnado no recibió retroalimentación y la RP sí que mejoró el rendimiento académico del alumnado, aunque posiblemente podría deberse a la forma de llevar a cabo la RP, ya que en Rowley y McCrudden (2020) el estudio consistió en 3 sesiones, en la primera tomaron notas sobre un texto, en la segunda sesión realizaron la fase de intervención con RP y por último, una prueba final de rendimiento, mientras que en Jägerskog et al. (2019) no tuvieron una sesión de tomar apuntes, simplemente prestaron atención a una presentación visuoverbal.

En lo que respecta al tercer grupo de artículos, relacionados con el Aprendizaje de Vocabulario Extranjero, se observó que al comparar la RP con otras estrategias como el

reestudio, aumentaba el rendimiento del estudiantado (Barenberg et al., 2021; Bertilsson et al., 2021; Jonsson et al., 2021; Wiklund-Hörnqvist et al., 2022), resultados que se ajustan con la revisión y metaanálisis realizada por Yang et al. (2021), aunque como se ha señalado anteriormente es difícil comparar los resultados de esta revisión con esa revisión, ya que abarca además de estudiantado de Secundaria y Bachillerato, otro grupo de alumnado que se encuentran fuera del rango de edad establecido para esta revisión. Cabe destacar que hubo estudios que realizaron experimentos donde compararon la RP con el reestudio en un escáner de resonancia magnética, mostrando el respaldo de los resultados comportamentales mencionados anteriormente por medio de los resultados de neuroimagen (Jonsson et al., 2021; y Wiklund-Hörnqvist et al., 2022). De hecho, se encontraron efectos significativos del *testing effect*, con activaciones en las regiones corticales y subcorticales lateralizadas a la izquierda, independientemente del rendimiento. Resultados parecidos encontraron Zhifang et al. (2020) puesto que en su estudio se observó que al analizar la RP con el reestudio, la RP conllevaba una mayor representación en la corteza prefrontal medial durante el post-test, independientemente del rendimiento, lo que sugiere que la RP puede facilitar la memoria por medio de los mecanismos de la corteza prefrontal medial en la integración de la memoria, la diferenciación y la consolidación. Aunque es necesario destacar que el estudio se llevó a cabo con estudiantes mayores de 18 años.

Además, se analizaron estudios que compararon diferentes condiciones de la RP. Por ejemplo, van den Broek et al. (2019) encontraron que a la hora de proporcionar *feedback* con pistas ortográficas, mnemotécnicas y *cognates* no mejoraron el rendimiento en comparación con el *feedback* con la respuesta correcta, aunque sí cuando las mismas pistas se usaron en el post-test, lo que se puede explicar por medio de la teoría del procesamiento de transferencia apropiada. Del mismo modo, el añadir imágenes a las preguntas de RP redujo el rendimiento académico (van den Broek et al., 2021), posiblemente debido a la disminución del esfuerzo mental necesario para extraer información de la memoria, pues en los tres experimentos se observó que el estudiantado respondió mucho más rápido y de manera más precisa en la condición de imágenes durante y tras la RP en comparación con las condiciones donde no había imagen, pudiéndose deberse a que fuese más sencillo.

Las MCQs escalonadas resultaron en un mayor rendimiento en comparación con las MCQs estándar. No obstante, es necesario señalar que solo aquellas personas que realizaron el esfuerzo de recordar la respuesta correcta durante el tiempo entre la pregunta y la presentación de las opciones de respuesta (RP), fueron los que puntuaron más alto en el post-test. Asimismo, en este mismo estudio se realizaron otros estudios con personas adultas y en el experimento 3A y 3B obtuvieron que la MCQs escalonada en un entorno online mejoraba el rendimiento en comparación con la MCQs estándar. Sin embargo, las

diferencias en torno a la relación entre el uso de la RP durante el tiempo de espera no fueron significativas en estos experimentos.

A lo largo del análisis de los estudios, se identificaron resultados que se consideran importantes de examinar. En lo que se refiere a transferencia, Barenberg et al. (2021), encontraron que la RP mejora el rendimiento académico en contextos de no transferencia, resultados que se alinean con la revisión llevada a cabo por Pan y Richard (2018). Sin embargo, es necesario mencionar que los estudios analizados en Pan y Richard (2018) se llevaron a cabo en laboratorios y con adultos jóvenes. Por otro lado, en el estudio de Jägerskog et al. (2019) en Ciencias Sociales no se observó una relación entre el uso de la RP y la transferencia de conocimientos. Cabe destacar que este resultado puede deberse a la falta de *feedback*, ya que, de acuerdo con Pan y Agarwal (2020), la transferencia en RP es más efectiva cuando se aplica el *feedback*. Asimismo, Agarwal (2019), van den Broek et al. (2019) (en el experimento 1 y 2), Bertilsson et al. (2021), Jonsson et al. (2021) y van den Broek et al. (2021) (en el experimento 2) encontraron que cuando el formato de la prueba RP es similar al post-test, se obtienen resultados más altos que cuando se modifica alguna de las características del formato inicial, lo que se puede explicar por medio de la teoría del procesamiento de transferencia apropiada (Morris et al., 1977), que establece que el rendimiento de la memoria depende de la similitud entre la prueba y el post-test. Por lo tanto, parece que la RP, en esta revisión, suele darse más en contextos de no transferencia, aunque tampoco se puede concluir que la RP solo se da en contextos de no transferencia, ya que solo se está comparando con un solo estudio de transferencia. Por lo tanto, es necesario que futuras investigaciones se centren en este tema en entornos educativos, ya que como comentaron Pan y Agarwal (2020) la transferencia es un aspecto esencial en la educación. Por lo tanto, en esta revisión, la RP se observa más en contextos de no transferencia, aunque no se puede concluir que solo ocurra en este tipo de contextos debido a la limitación de los estudios que analizan este aspecto.

Otro aspecto que se ha visto en los estudios analizados es que en la mayoría de los estudios no se aporta información sobre cómo se ha escogido el docente que impartía el contenido, únicamente se menciona que el mismo docente impartió el material (Jägerskog et al., 2019; y Molin et al., 2021) o que simplemente hubo un docente e investigador/a (Leggett et al., 2019; van den Broek et al., 2019; van den Broek et al., 2023). El estudio de Szeibert et al. (2023) fue el único que realizó entrevistas a varios docentes para escoger dos docentes que tuviesen las mismas creencias en cuanto a las matemáticas, la enseñanza y el aprendizaje de esta y, en van den Broek et al. (2021) se menciona que el docente está certificado. Por lo tanto, en la gran mayoría no se ha controlado el rol del docente como factor modulador de los resultados de aprendizaje del estudiantado, también conocido como el efecto de Hawthorne o “reactividad de los participantes” (Diaper, 1990; Paradis & Sutkin, 2017), lo que podría haber influido en los resultados. Se recomienda que en

futuras investigaciones se tenga en cuenta el rol del profesorado, como también comenta Agarwal et al. (2021) en su revisión.

Además, en la mayoría de los estudios se compara la RP con el reestudio. A excepción del estudio de Rowley y McCrudden (2020), Molin et al. (2021) y Szeibert *et al.* (2023) compararon la RP con una clase tradicional. El hecho de que la RP se compare mayoritariamente con el reestudio se observa también en otras revisiones (Agarwal et al., 2021; Moreira et al., 2019; y Yang et al., 2021). Incluso viéndose que esta estrategia, de acuerdo con Callender y McDaniel (2009), no mejora el rendimiento académico en el aula. Como Agarwal et al. (2021) mencionan en su revisión que hay muchos estudios donde se muestra que la RP en comparación con el reestudio mejora el rendimiento académico, sería necesario compararse en futuras investigaciones con otras estrategias de aprendizaje eficaces en el entorno educativo.

Finalmente, tanto en la revisión de Agarwal et al. (2021) como por Moreira et al. (2019) no encontraron muchos estudios que analizaran las diferencias individuales. En esta revisión hay tres estudios que abordan este tema (Bertilsson et al., 2021; Jonsson et al., 2021; y Wiklund-Hörnqvist et al., 2022). En los tres se observó el *testing effect*, independientemente de la capacidad de memoria, la necesidad de cognición y la perseverancia (Bertilsson et al., 2021), del grupo de capacidad cognitiva (alta, intermedia y baja) y del intervalo de retención (Jonsson et al., 2021), y del grupo de NFC (alto y bajo) (Wiklund-Hörnqvist et al., 2022). En cuanto a la capacidad de memoria el estudio realizado de Yang et al. (2020) tuvo resultados similares. Asimismo, cabe destacar que en el estudio de Wiklund-Hörnqvist et al. (2022) se encontró que aquel estudiantado con mayor NFC tuvo un nivel de rendimiento superior en comparación con el grupo de NFC bajo. Por otro lado, en el estudio de Jonsson et al. (2021), se encontró que en los tres intervalos de retención se observaron diferencias significativas entre aquel estudiantado que pertenecía al grupo de baja capacidad cognitiva e intermedia, y de entre intermedia y alta capacidad cognitiva. No obstante, también se observó que el alumnado perteneciente al grupo de habilidad cognitiva baja y RP, rindió igual o más en comparación con el grupo de habilidad cognitiva intermedia y reestudio. Lo mismo pasó con aquel estudiantado que pertenecía al grupo de habilidad cognitiva intermedia y RP en comparación con el grupo de habilidad cognitiva alta y reestudio. Estos resultados se alinean con el estudio llevado a cabo por Yang et al. (2020).

Las limitaciones de esta revisión radican en que solo se han incluido estudios publicados entre 2019 y 2024, lo que implica que se han podido excluir artículos sobre la RP publicados fuera de dicho periodo. Otra limitación podría ser que se hayan omitido ciertos estudios que tratasen la RP, ya que la RP se conoce por diversos términos como “*testing effect*”, “*test-enhanced learning*”, lo que en ocasiones puede crear confusión con otras actividades como el test estandarizado, como ya comentó Agarwal et al. (2021). Por otro lado, en ocasiones, existen ciertos artículos que confunden la RP con la práctica

espaciada, lo que dificulta realizar el cribado de artículos durante la búsqueda. De hecho, muchos de los estudios pertenecientes a la agrupación de aprendizaje de vocabulario extranjero se podrían considerar práctica espaciada puesto que tenían un total de palabras, y en cada post-test, se les añadía una parte de ese conjunto, lo que concuerda con la definición que proporciona Carpenter y Agarwal (2020) de que la práctica espaciada consiste en gestionar el tiempo de estudio de manera que este se distribuya en múltiples sesiones a lo largo del tiempo. Por lo tanto, es necesario que haya un consenso en cuanto a los conceptos que utilizar en esta área. Asimismo, ciertos estudios no comparan la RP con otra estrategia de aprendizaje, sino que varían ciertas características de la RP y comparan dos condiciones de RP, lo que dificulta conocer si la RP es eficaz. Esto ha ocurrido sobre todo en los estudios de Aprendizaje de Vocabulario Extranjero. A pesar de todas estas limitaciones, se puede considerar que todos estos estudios aportan información relevante para comprender mejor la RP.

En resumen, los resultados de esta revisión indican que la RP es generalmente eficaz en alumnado de Educación Secundaria y Bachillerato, aunque con alguna variación según el contenido impartido y las condiciones de los estudios. Se observó que, en materias relacionadas con Ciencias Naturales y Aprendizaje de Vocabulario Extranjero, los datos son más consistentes, mejorando la RP el rendimiento académico. Sin embargo, en el caso de las Ciencias Sociales, los resultados fueron irregulares, posiblemente debido a la falta de retroalimentación. Además, los resultados sugieren que la aplicación de RP se ha realizado sobre todo en contextos de no transferencia. Por ello, se requiere de más investigaciones para entender mejor su impacto en la transferencia de conocimientos. Igualmente, es fundamental la consideración del rol del docente y la comparación con otras estrategias de aprendizaje, distintas al reestudio, en futuras investigaciones. Finalmente, los estudios revisados indican que la RP beneficia a estudiantes de diversas capacidades cognitivas en el Aprendizaje de Vocabulario Extranjero. Futuros estudios podrían enfocarse en comparar la RP con otras estrategias de aprendizaje y en investigar la influencia de factores como el *feedback* y el rol del docente con la finalidad de ofrecer una comprensión más integral de su eficacia.

Por lo tanto, la RP parece una estrategia de aprendizaje prometedora para estudiantes de Secundaria y Bachillerato. Sin embargo, se necesitan más investigaciones para comprender plenamente su eficacia y optimizar su uso en el aula.

6. Propuesta de intervención

A continuación, se va a proponer un ejemplo de RP en la asignatura de “Relaciones en el equipo de trabajo”, del Ciclo Formativo de grado medio de Técnico/a en Cuidados en Enfermería. Está dirigida específicamente al grupo de primer curso de Auxiliar de Enfermería, grupo formado mayoritariamente por adolescentes, con quien estuve en contacto durante el Prácticum.

Esta propuesta de intervención se realizará en el módulo de Formación y Orientación Laboral (FOL), que pertenece al grupo de Ciencias Sociales, en concreto en la unidad didáctica “Conflicto y Negociación”. En cuanto a los objetivos de esta unidad son: identificar los principales tipos de conflicto en el ámbito laboral y sus fuentes; conocer los diferentes métodos para resolver conflictos; y aplicar las diferentes tácticas para negociar de forma eficaz.

De acuerdo con los resultados obtenidos de esta revisión, toda RP debe contener una fase de estudio, una fase de realización de tareas de RP y un post-test tras un periodo de tiempo que puede variar de 5 minutos tras la finalización de las tareas de RP, hasta 4 semanas después. Según los resultados de esta revisión, en disciplinas de Ciencias Sociales los efectos de la RP estos son inconsistentes, posiblemente por la falta de *feedback*. Por lo tanto, en esta propuesta se proporcionará retroalimentación.

Para ello, se utilizará la RP de la siguiente manera:

- La primera sesión consistirá en presentar la unidad didáctica por medio de un Power Point, posteriormente, se realizará una prueba de preguntas cortas donde se les preguntará sobre la definición de los conceptos claves de la unidad. Un ejemplo de ello podría ser: Define el conflicto o ¿qué métodos existen para resolver un conflicto?
- Seguidamente, se les proporcionará *feedback* por medio de una hoja con las respuestas de cada pregunta. Por ejemplo: el conflicto es una confrontación o desacuerdo entre ideas, valores, objetivos o posturas dentro de un ámbito de la organización.
- Tras una semana, ya que este grupo tiene solo dos horas el miércoles, se les entregará la misma prueba antes de iniciar la sesión. Seguirá el mismo formato, ya que en los estudios revisados se ha visto que en aquellas pruebas en las que coincide el formato de la prueba RP con la prueba post-test, la RP suele ser efectiva.
- Tras dos semanas, se les volverá a examinar de este contenido junto a la unidad de “Equipo de Trabajo”. Esta unidad tiene un nivel de dificultad similar a la unidad de “Conflicto y Negociación”, por lo tanto, por medio de la prueba objetiva final se podrá analizar si la RP ha mejorado el contenido de “Conflicto y

Negociación” en comparación con el contenido de la unidad de “Equipo de Trabajo” que se habrá impartido sin RP.



7. Referencias

- Agarwal, P. K. (2019). Retrieval practice & Bloom's taxonomy: Do students need fact knowledge before higher order learning? *Journal of Educational Psychology*, *111*(2), 189–209. <https://doi-org.publicaciones.umh.es/10.1037/edu0000282>
- Agarwal, P.K., Nunes, L.D. & Blunt, J.R. (2021). Retrieval Practice Consistently Benefits Student Learning: a Systematic Review of Applied Research in Schools and Classrooms. *Educ Psychol Rev*, *33*, 1409–1453. <https://doi.org/10.1007/s10648-021-09595-9>
- Agarwal, P. K., Roediger, H. L., III, McDaniel, M. A., & McDermott, K. B. (2018). *Cómo utilizar la práctica de recuperación para mejorar el aprendizaje*. Washington University St. Louis. Recuperado de: https://drive.google.com/drive/folders/1g4fwhtGpLGb1UxUY0V_8_0iC8ZPYjNkk
- Barenberg, J., Berse, T., Reimann, L. & Dutke, S. (2021). Testing and transfer: Retrieval practice effects across test formats in English vocabulary learning in school. *Appl Cognit Psychol.*, *35*: 700–710. <https://doi.org/10.1002/acp.3796>
- Bertilsson, F., Stenlund, T., Wiklund-Hörnqvist, C., & Jonsson, B. (2021). Retrieval Practice: Beneficial for All Students or Moderated by Individual Differences? *Psychology Learning & Teaching*, *20*(1), 21-39. <https://doi.org/10.1177/1475725720973494>
- Callender, A. A., & McDaniel, M. A. (2009). The limited benefits of rereading educational texts. *Contemp. Educ. Psychol.*, *34*, 30–41. doi: 10.1016/j.cedpsych.2008.07.001ç
- Carpenter, S. K., & Agarwal, P. K. (2020). How to use spaced retrieval practice to boost learning. <http://www.retrievalpractice.org>
- Catherine-Noémie, A. G., Lorena, D., Martin, G., Nikolai, A., & Nico, B., (2022). Benefit from retrieval practice is linked to temporal and frontal activity in healthy young and older humans, *Cerebral Cortex Communications*, *Volume 3*, Issue 1, tgac009, <https://doi.org/10.1093/texcom/tgac009>
- Diaper, G. (1990). The Hawthorne effect: a fresh examination. *Educational Studies*, *16*, 261–267.
- Dirkx, K.J.H., Camp, G., Kester, L., & Kirschner, P.A. (2019). Do secondary school students make use of effective study strategies when they study on their own? *Appl Cognit Psychol.*, *33*: 952–957. <https://doi.org/10.1002/acp.3584>

- Jägerskog, A. S., Jönsson, F. U., Selander, S., & Jonsson, B. (2019). Multimedia learning trumps retrieval practice in psychology teaching. *Scandinavian Journal of Psychology*. Advance online publication. <https://doi-org.publicaciones.umh.es/10.1111/sjop.12527>
- Jonsson, B., Wiklund-Hörnqvist, C., Stenlund, T., Andersson, M., & Nyberg, L. (2021). A learning method for all: The testing effect is independent of cognitive ability. *Journal of Educational Psychology*, 113(5), 972–985. <https://doi-org.publicaciones.umh.es/10.1037/edu0000627>
- Langbroek, S., Duchatelet, D., Joosten-ten Brinke, D., & Camp, G. (2024). Enhancing teachers' instruction on how to study: an exploration of the effectiveness of learning strategies for particular secondary school subjects. *Frontiers in Education*, 9, 2024. <https://doi.org/10.3389/feduc.2024.1340120>
- Leggett, J.M.I., Burt, J.S., & Carroll, A. (2019). Retrieval practice can improve classroom review despite low practice test performance. *Appl Cognit Psychol.*, 33: 759–770. <https://doi-org.publicaciones.umh.es/10.1002/acp.3517>
- Molin, F., Cabus, S., Haelermans, C., & Groot, W. (2021). Toward Reducing Anxiety and Increasing Performance in Physics Education: Evidence from a Randomized Experiment. *Res Sci Educ*, 51 (Suppl 1), 233–249. <https://doi.org/10.1007/s11165-019-9845-9>
- Moreira, B.F.T., Pinto, T.S.S., Starling, D.S.V. and Jaeger, A. (2019). Retrieval Practice in Classroom Settings: A Review of Applied Research. *Front. Educ.*, 4:5. doi: 10.3389/feduc.2019.00005
- Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer-appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, 16(5), 519–533
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., ... Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372:n71. <https://doi.org/10.1136/bmj.n71>.
- Pan, S. C., & Agarwal, P. K. (2020). Retrieval practice and transfer of learning: fostering students' application of knowledge. <http://www.retrievalpractice.org>
- Pan, S. C., & Rickard, T. C. (2018). Transfer of test-enhanced learning: meta-analytic review and synthesis. *Psychonomic Bulletin*, 144(7), 710–756.

- Paradis, E., & Sutkin, G. (2017). Beyond a good story: from Hawthorne effect to reactivity in health professions education research. *Medical Education*, *51*(1), 31–39.
- Rowley, T. & McCrudden, M.T. (2020). Retrieval practice and retention of course content in a middle school science classroom. *Appl Cognit Psychol*. *34*: 1510–1515. <https://doi-org.publicaciones.umh.es/10.1002/acp.3710>
- Szeibert, J., Muzsnay, A., Szabó, C. et al. (2023). A Case Study of Using Test-Enhanced Learning as a Formative Assessment in High School Mathematics. *Int J of Sci and Math Educ* *21*, 623–643. <https://doi.org/10.1007/s10763-022-10264-8>
- Tullis, J.G. & Maddox, G.B. (2020). Self-reported use of retrieval practice varies across age and domain. *Metacognition Learning* *15*, 129–154. <https://doi.org/10.1007/s11409-020-09223-x>
- van den Broek, G. S. E., Gerritsen, S. L., Oomen, I. T. J., Velthoven, E., van Boxtel, F. H. J., Kester, L., & van Gog, T. (2023). Optimizing multiple-choice questions for retrieval practice: Delayed display of answer alternatives enhances vocabulary learning. *Journal of Educational Psychology*, *115*(8), 1087–1109. <https://doi-org.publicaciones.umh.es/10.1037/edu0000810>
- van den Broek, G. S. E., Segers, E., van Rijn, H., Takashima, A., & Verhoeven, L. (2019). Effects of elaborate feedback during practice tests: Costs and benefits of retrieval prompts. *Journal of Experimental Psychology: Applied*, *25*(4), 588–601. <https://doi-org.publicaciones.umh.es/10.1037/xap0000212>
- van den Broek, G. S. E., van Gog, T., Jansen, E., Pleijsant, M., & Kester, L. (2021). Multimedia effects during retrieval practice: Images that reveal the answer reduce vocabulary learning. *Journal of Educational Psychology*, *113*(8), 1587–1608. <https://doi-org.publicaciones.umh.es/10.1037/edu0000499>
- Wiklund-Hörnqvist, C., Stillesjö, S., Andersson, M., Jonsson, B., & Nyberg, L. (2022). Retrieval Practice Is Effective Regardless of Self-Reported Need for Cognition - Behavioral and Brain Imaging Evidence. *Frontiers in Psychology*, *12*(February), 1–10. <https://doi.org/10.3389/fpsyg.2021.797395>
- Yang, C., Luo, L., Vadillo, M.A., Yu, R. & Shanks, D.R. (2021 Apr). Testing (quizzing) boosts classroom learning: A systematic and meta-analytic review. *Psychol Bull.*, *147*(4):399-435. doi: 10.1037/bul0000309. Epub 2021 Mar 8. PMID: 33683913.
- Yang, C., Sun, B., Potts, R., Yu, R., Luo, L., & Shanks, D. R. (2020). Do working memory capacity and test anxiety modulate the beneficial effects of testing on new learning? *Journal of Experimental Psychology: Applied*, *26*(4), 724–738. <https://doi-org.publicaciones.umh.es/10.1037/xap0000278>

- Ye, Z., Shi, L., Li, A., Chen, C. & Xue, G. (2020 May). Retrieval practice facilitates memory updating by enhancing and differentiating medial prefrontal cortex representations. *Elife.*, 18;9:e57023. doi: 10.7554/eLife.57023. PMID: 32420867; PMCID: PMC7272192.
- Zhifang, Y., Liang, S., Anqi, L., Chuansheng C. & Gui, X. (2020). Retrieval practice facilitates memory updating by enhancing and differentiating medial prefrontal cortex representations *eLife* 9:e57023.

