



**UNIVERSIDAD DE ORIENTE
NÚCLEO DE SUCRE
ESCUELA DE HUMANIDADES Y EDUCACIÓN
COMISIÓN DE TRABAJO DE ASCENSO**

***Soporte Empírico de la Crítica de Wang, Thomas y sus
Colaboradores sobre la Efectividad del Método de Palabra Clave
en la Retención a Largo Plazo de Vocabulario en L2***

**Autora:
MSc. Yarimax Rodríguez**

Trabajo de Ascenso presentado para optar a la categoría de Agregado

Cumaná, octubre de 2022

DEDICATORIA

A mi mayor motor; mis hijos Maxell y Jesús Andrés, a mis adorados padres y mis mejores amigos; mis hermanos. Gracias por su apoyo incondicional.

AGRADECIMIENTO

A las personas que han colocado su apoyo y esfuerzo durante la realización y culminación de este trabajo.

ÍNDICE GENERAL

	Pág.
DEDICATORIA	i
AGRADECIMIENTO	ii
LISTA DE TABLAS.....	iv
RESUMEN	v
CAPITULO I	1
EL PROBLEMA.....	1
JUSTIFICACIÓN	5
OBJETIVO GENERAL.....	7
OBJETIVOS ESPECÍFICOS.....	7
CAPITULO II	9
REVISION DE LA LITERATURA	9
Retención a Largo Plazo	9
CAPITULO iii.....	31
METODOLOGIA.....	31
Tipo de Investigación	31
Población y Muestra.....	31
Procedimientos	32
CAPITULO IV	38
RESULTADOS	38
CAPITULO V	47
DISCUSION Y CONCLUSIONES	47
CONCLUSIONES	57
LIMITACIONES	58
REFERENCIAS.....	59
HOJAS DE METADATOS	67

LISTA DE TABLAS

Tabla 1. Exps. con diseños inter-participantes e interacción estadísticamente significativa.	39
Tabla 2. Parte inter-participantes de los Exps. e interacción estadísticamente significativa.	40
Tabla 3. Valores d del método de palabra clave ^a en la retención a largo plazo (Exps. con diseños inter-participantes de Wang, Thomas y sus colaboradores).	42
Tabla 4. Valores d del método de palabra clave ^a en la retención a largo plazo (parte inter-participantes) en los experimentos de Carney y Levin .	43
Tabla 5. Valores d de la versión estándar del método de palabra clave en los Exps. de Wang, Thomas y sus colegas y en la parte inter-participantes de los de Carney y Levin	44
Tabla 6. Valores d del método de palabra clave en la parte intra-participantes de los diseños en los Exps. de Wang, Thomas y sus colegas ^a y los de Carney y Levin	46



**UNIVERSIDAD DE ORIENTE
NÚCLEO DE SUCRE
ESCUELA DE HUMANIDADES Y EDUCACIÓN
COMISIÓN DE TRABAJO DE ASCENSO**

Soporte Empírico de la Crítica de Wang, Thomas y sus Colaboradores sobre la Efectividad del Método de Palabra Clave en la Retención a Largo Plazo de Vocabulario en L2

Autora:

MSc. Yarimax Rodríguez

RESUMEN

La presente investigación examinó el soporte empírico de la crítica de Wang, Thomas y sus colaboradores sobre la efectividad del método de palabra clave en la retención a largo plazo de vocabulario en L2 y su coincidencia con los hallazgos reportados en Carney y Levin (1998). La muestra la conformaron 14 estudios primarios (nueve de Wang, Thomas y sus colegas y cinco de Carney y Levin). Los hallazgos reportados en dichos estudios se examinaron desde dos perspectivas: el porcentaje de olvido y el desempeño en los tests tardíos. Desde la perspectiva del porcentaje de olvido, los resultados confirmaron el planteamiento de Wang, Thomas y sus colegas de que en diseños inter-participantes, el método de palabra clave exhibe porcentajes de olvido significativamente mayores que los de las estrategias control. Estos hallazgos fueron corroborados parcialmente por los reportados en Carney y Levin. Desde la perspectiva del desempeño en los tests tardíos, los valores d computados a los hallazgos reportados en Wang, Thomas y sus colaboradores fueron inconsistentes; mientras en algunos estudios el grupo de palabra clave aventajó al grupo control, en otros, sucedió lo contrario. En cambio, los valores d de los hallazgos reportados en Carney y Levin difirieron de los de Wang, Thomas y sus colaboradores; el método de palabra clave superó al método de repetición en todos los cinco experimentos. Con respecto a los diseños intra-participantes, los hallazgos corroboraron su señalamiento de que en ese tipo de diseño, la ventaja inicial del método de palabra clave se mantiene por unos días. Adicionalmente, se examinaron otros temas relacionados.

Palabras Clave: Método de palabra clave, vocabulario en L2, retención a largo plazo, porcentaje de olvido, diseños inter-participantes, diseños intra-participantes

CAPITULO I

EL PROBLEMA

El método de palabra clave (*The Keyword Method*) se ha convertido en uno de los métodos de enseñanza de vocabulario en segunda lengua (L2) más populares en las últimas cinco décadas, y, sin duda, en el más investigado. Tal y como fue desarrollado por Atkinson y sus colaboradores (Atkinson, 1975; Atkinson y Raugh, 1975; Raugh y Atkinson, 1975; Raugh, Schupbach, y Atkinson, 1977), el método de palabra clave consiste en la activación de dos enlaces: un enlace acústico y uno mental, en ese orden. En la activación del enlace acústico, el aprendiz selecciona una palabra o frase en su lengua nativa (L1), preferiblemente concreta (i.e., *la palabra clave o keyword*), que sea acústicamente similar a la palabra desconocida (i.e., *the L2 Word*). Seguidamente se activa el enlace mental en el que el aprendiz crea una imagen donde el referente de la palabra clave interactúa con el equivalente en L1 de la palabra por aprender. Por ejemplo, un hablante nativo del español pudiera aprender que la palabra inglesa *couch* significa sofá, usando la palabra *caucho* como la palabra clave y creando una imagen interactiva de un caucho sobre un sofá. La premisa subyacente del método de palabra clave es que la exposición a la palabra de la L2 (*couch*) evocará la palabra clave (*caucho*), la cual a su vez reevocará la imagen mental *del caucho sobre el sofá* que incluye el equivalente en L1 (*sofá*). En otras palabras, el camino al recuerdo (*the retrieval path*) de la palabra *couch* es como sigue:

couch → caucho → caucho sobre un sofá → sofá

La efectividad del método de palabra clave reportada en estos estudios iniciales inspiró la realización y posterior publicación de un considerable número de investigaciones que examinaron la eficacia del método para facilitar el aprendizaje de vocabulario en diversas L2s y en diferentes grupos etarios (Griffith, 1981; Hall, Wilson y Patterson, 1981; Levin y Pressley, 1980; Levin, Pressley, McCormick, Miller y Shriberg, 1979; Pressley, 1977; Pressley, Levin, Digdon,

Bryant y Ray, 1983; Pressley, Levin, Hall, Miller y Barry, 1980; Pressley, Levin y McCormick, 1980; Pressley, Levin y Miller, 1981b; Pressley, Levin, Wakamura, Hope, Bispo y Toye, 1980; Pressley, Samuel, Hershey, Bishop y Dickinson, 1981; Willerman y Melvin, 1979. Asimismo, se publicaron varios estudios que investigaron su eficiencia en la adquisición de vocabulario en inglés como L1 (e.g., Levin, McCormick, Miller, Barry y Pressley, 1982, Pressley, Levin, Kuiper, Bryant y Michener, 1982; Pressley, Levin y Miller. 1981a, 1982; Sweeney y Belleza, 1982) y de información en áreas de contenido (e.g., Levin, Shriberg y Berry, 1983; Levin, Shriberg, Miller, McCormick y Levin, 1980; Shriberg, Levin, McCormick y Pressley, 1982).

Sin embargo, la gran mayoría de estos estudios examinó la efectividad del método de palabra clave en la adquisición de vocabulario en L2, es decir, su eficacia para facilitar la asociación de la palabra en L2 y su equivalente en L1 tan pronto culminaba la intervención o tratamiento (recuerdo inmediato), lo cual, posiblemente, hizo que Levin y Pressley (1985) en su revisión sobre la efectividad de las estrategia mnemónicas en general y sobre la del método de palabra clave en particular, plantearan la necesidad de investigar la efectividad del método a largo plazo, un aspecto de extraordinaria importancia desde la perspectiva pedagógica. Específicamente señalaron:

El tema de si los efectos de la estrategia mnemónica son durables, es decir, se mantienen en el tiempo es de extrema importancia. Una cosa es que el método de palabra clave sea beneficioso en el desempeño en una prueba administrada inmediatamente después del aprendizaje, pero otra muy distinta es hallar beneficios similares en pruebas de retención administradas un mes después, una semana después, o incluso un día después. Tales efectos a largo plazo son absolutamente necesarios para convencer a los educadores del valor de un método o estrategia en particular (p. 159).

Estos investigadores reconocieron la existencia de muy pocos estudios sobre el tema y con ello su imposibilidad para emitir señalamientos concluyentes al respecto. El llamado de atención de estos investigadores sobre la necesidad de examinar la efectividad a largo plazo del método de palabra clave no tuvo una repercusión inmediata pues como bien lo señala Rodríguez, Y (2021, p.), en los

primeros 15 años desde que Atkinson presentó formalmente el método en 1975, sólo 10 estudios, publicados en 8 artículos, fueron codificados por la autora como estudios de retención de vocabulario en L2.

La situación experimentó un cambio radical en la década de los 90, impulsada por el programa de investigación desarrollado por Wang, Thomas, y sus colaboradores (Thomas y Wang, 1996, Exps. 1-2; Wang y Thomas, 1995, Exps. 2-3; Wang, Thomas, Inzana y Primicerio, 1993; Wang, Thomas y Ouellette, 1992). Aunque reconocieron la superioridad del método de palabra clave en el recuerdo inmediato de vocabulario en L2, cuestionaron su efectividad en la retención a largo plazo. Argumentaron que la efectividad a largo plazo, reportada en los estudios previos se debió a la implementación de diseños intra-participantes (*within-participants designs*) en los que los mismos aprendices toman el test inmediato y el test tardío. Desde su perspectiva, el test inmediato sirve como una práctica adicional y que dada la marcada superioridad del método de palabra clave en esta prueba, la ventaja inicial se mantenía después de unos días. De allí que en su programa de investigación implementaran el uso de diseños inter-participantes (*between-participants designs*) en los que el intervalo de retención se manipula como factor inter-participantes, es decir, un grupo de aprendices recibe el test inmediato y otro grupo el test tardío. Predijeron que, en este tipo de diseño, dada la ausencia de la prueba inmediata, el porcentaje de olvido (*forgetting rate*) del método de palabra clave sería muy superior al que exhibiría el método no mnemónico y que la ventaja inicial del método de palabra clave desaparecería completamente al cabo de pocos días. La diferencia en el porcentaje de olvido lo evidenciaría el que el efecto de la interacción entre las condiciones de aprendizaje y el intervalo de retención fuera estadísticamente significativo.

Entre 1992 y 1996, estos investigadores publicaron los resultados de 11 experimentos, distribuidos de la siguiente manera: 4 en Wang, Thomas y Ouellette (1992); 1 en Wang, Thomas, Inzana y Primicerio (1993); 3 en Wang y Thomas (1995) y 3 en Thomas y Wang (1996). En estos estudios, se comparó la efectividad del método de palabra clave con la del método de repetición (Thomas y

Wang, 1996; Wang, Thomas y Ouellette, 1992; Wang, Thomas, Inzana y Primicerio, 1993) y con la del método de contexto (Wang y Thomas, 1995). De ellos, se excluyeron 2 en la síntesis metodológica de Rodríguez, Y (2021): el Experimento 1 de Wang y Thomas (1995) porque las palabras experimentales fueron palabras poco frecuentes (*obscure words*) en L1 y el Experimento 3 de Thomas y Wang (1996) porque sólo se compararon variantes del método de palabra clave.

Los hallazgos reportados por este grupo de investigadores parecieron confirmar su cuestionamiento sobre la efectividad del método de palabra clave en la retención a largo plazo de vocabulario en L2. Con la excepción del Experimento 3 de Wang y Thomas (1995), en los restantes estudios, incluso en el Experimento 4 de Wang y otros (1992) en el que se utilizó un diseño intra-participantes, el método de palabra clave exhibió un porcentaje de olvido significativamente mayor al del método control. Aunque estos hallazgos inspiraron otros estudios (e.g., Ávila y Sadoski, 1996; Carney y Levin, 1998; Colón y Rodríguez, M (2003); Rodríguez, M y Sadoski, 2000), parecieron haber socavado el interés de los investigadores en el método de palabra clave en los últimos años (ver Rodríguez, Y, 2021). De los mencionados estudios, destacan los de Carney y Levin por haber realizado 5 experimentos con el propósito de replicar y extender los hallazgos reportados por Wang, Thomas y sus colaboradores. Contrariamente a sus antecesores, quienes utilizaron exclusivamente diseños inter-participantes en casi todos sus experimentos, Carney y Levin implementaron diseños mixtos que incluían una parte intra-participantes y otra inter-participantes. Sus hallazgos fueron sólo parcialmente consistentes con los reportados por Wang, Thomas y sus colaboradores.

En líneas generales, el impacto del programa de investigación de Wang, Thomas y sus colaboradores sobre la efectividad del método de palabra clave en la retención a largo plazo de vocabulario en L2 es innegable. Sin embargo, sorprende que luego de más de 2 décadas, no se hayan objetado aspectos relacionados entre sí y que fueron determinantes en su interpretación de los

mencionados hallazgos (e.g., la noción del porcentaje de olvido como elemento exclusivo de la ventaja a largo plazo (*long-term advantage*) y el uso del test de significación estadística como soporte empírico) ni que sus hallazgos hayan sido reinterpretados a la luz de los reportados en las réplicas de Carney y Levin (1998, Exps. 1-5). La presente investigación ofrece una evaluación exhaustiva del soporte empírico de los hallazgos reportados por Wang, Thomas y sus colaboradores. Por una parte, se interpretaron dichos hallazgos desde la perspectiva del desempeño relativo de los métodos para profundizar el análisis. Por otra parte, además de la tradicional prueba de la significación estadística, se utilizaron criterios estadísticos más novedosos (e.g., el tamaño de los efectos, *effect size*, *ES*) para reanalizar parte de los resultados presentados por estos investigadores. Finalmente, se examinó hasta qué punto los hallazgos reportados en las réplicas de Carney y Levin fueron consistentes con los de Wang, Thomas y sus colegas.

JUSTIFICACIÓN

Aunque Wang, Thomas y sus colaboradores reconocieron la enorme superioridad del método de palabra clave para promover el recuerdo inmediato de vocabulario en L2, cuestionaron su efectividad en la retención a largo plazo, particularmente en diseños inter-participantes, es decir, cuando la mitad de los estudiantes asignados aleatoriamente a este método toma el test inmediato mientras que la otra mitad lo hace algunos días después. Analizando la retención a largo plazo desde la perspectiva del porcentaje de olvido, predijeron que, en este tipo de diseño, la superioridad del método de palabra clave en el test inmediato desaparecería al cabo de unos pocos días. En otras palabras, plantearon que el método de palabra clave exhibiría un porcentaje de olvido superior al del método no mnemónico, lo cual sería corroborado por la presencia de un efecto estadísticamente significativo de la interacción entre los métodos de aprendizaje y los intervalos de retención.

Para examinar su hipótesis, estos investigadores desarrollaron una serie de estudios en los que manipularon varios aspectos, incluyendo el método control (e.g., repetición, contexto); el test tardío (2 días, 1 semana); el paradigma de aprendizaje (e.g., incidental, intencional); la L2 (e.g., francés, tagalog); el número de palabras experimentales (e.g., 20, 24); etc. En general, desde su perspectiva, los hallazgos que reportaron en la literatura parecieran corroborar su hipótesis. No obstante, el soporte empírico que sustenta la interpretación de dichos hallazgos no ha sido examinado suficientemente (cf. Carney y Levin, 1998; Gruneberg, 1998); por ejemplo, Wang, Thomas y sus colaboradores, al interpretar la retención de vocabulario en L2 exclusivamente desde la perspectiva del porcentaje de olvido, ofrecieron un análisis parcelado de los resultados. La presente investigación presenta una visión más completa del soporte empírico que debe sustentar el análisis real de dichos resultados al examinarlos también desde la perspectiva del desempeño relativo de las condiciones de aprendizaje, la otra interpretación reconocida por estos investigadores de lo que significa “la ventaja a largo plazo” (ver Wang y otros, 1992, pp. 520-521). Asimismo, esta otra perspectiva permitió usar criterios estadísticos más novedosos (e.g., tamaño del efecto, *ES*), en lugar del tradicional test de significación estadística de la hipótesis nula (NHST, por sus siglas en inglés) y sus correspondientes valores de probabilidad (*p values*), para analizar parte de la data presentada por este grupo de investigadores. Finalmente, examinó hasta qué punto los hallazgos reportados por los mencionados investigadores fueron consistentes con los reportados en las réplicas de Carney y Levin.

Formulación del Problema

¿Cuál es el soporte empírico del cuestionamiento de Wang, Thomas y sus colaboradores sobre la efectividad del método de palabra clave en la retención a largo plazo de vocabulario en L2 y en qué medida las réplicas de Carney y Levin (1998) se alinean con dicho soporte?

Interrogantes de la Investigación

1. ¿Corroborar la prueba de significación estadística de la interacción entre la condición de aprendizaje y el intervalo de retención la crítica de Wang, Thomas y sus colaboradores sobre la efectividad del método de palabra clave en la retención a largo plazo de vocabulario en L2?
2. ¿Sustentan los tamaños de los efectos (ESs) del método de palabra clave en los tests tardíos, estimados a partir de los hallazgos reportados por Wang, Thomas y sus colaboradores, su crítica sobre la efectividad del método de palabra clave en la retención a largo plazo de vocabulario en L2?
3. ¿Hasta qué punto los hallazgos reportados en las réplicas de Carney y Levin (1998) son consistentes con los reportados en los estudios de Wang, Thomas y sus colaboradores sobre la efectividad del método de palabra clave en la retención a largo plazo de vocabulario en L2?

OBJETIVO GENERAL

Examinar el soporte empírico de los hallazgos reportados en las investigaciones de Wang, Thomas y sus colaboradores sobre la efectividad del método de palabra clave en la retención a largo plazo de vocabulario en L2 y su nivel de coincidencia con los reportados en las réplicas de Carney y Levin (1998).

OBJETIVOS ESPECÍFICOS

1. Identificar el número de estudios de Wang, Thomas y sus colaboradores en el que el porcentaje de olvido del método de palabra clave relativo al del método control fue tan grande que hizo que el efecto de la interacción entre la condición de aprendizaje y el intervalo de retención fuera estadísticamente significativo.

2. Determinar hasta qué punto los tamaños de los efectos (ESs) del método de palabra clave en los tests tardíos, estimados a partir de los hallazgos reportados por Wang, Thomas y sus colaboradores, sustentan su crítica sobre la efectividad del método de palabra clave en la retención a largo plazo de vocabulario en L2.

3. Explorar hasta qué punto los hallazgos reportados en las réplicas de Carney y Levin (1998) son consistentes con los reportados por Wang, Thomas y sus colaboradores sobre la efectividad del método de palabra clave en la retención a largo plazo de vocabulario en L2.

CAPITULO II

REVISION DE LA LITERATURA

En el presente capítulo se describen los estudios realizados por Wang, Thomas y sus colaboradores entre 1992 y 1996, cuyos hallazgos proporcionaron el soporte empírico a su cuestionamiento sobre la efectividad del método de palabra clave en la retención a largo plazo de vocabulario en L2. Asimismo, se describen las réplicas de Carney y Levin (1998, Exps. 1-5) que abordaron directamente el mencionado cuestionamiento. Antes, sin embargo, se examina la noción de retención a largo plazo presentada por Wang, Thomas y Ouellette (1992, pp.520-521), destacándose la perspectiva que sustentó el mencionado cuestionamiento.

Retención a Largo Plazo

Dada la gran cantidad de evidencia que corroboraba la significativa superioridad del método de palabra clave en el aprendizaje y en los tests de recuerdo inmediato de vocabulario en L2, Wang y otros (1992), se preguntaron si esta ventaja se mantenía a largo plazo. A su juicio, la respuesta a dicha interrogante pasaba por considerar lo que exactamente significaba el término “ventaja a largo plazo”. En ese proceso, identificaron dos perspectivas distintas para su interpretación: *el porcentaje de olvido y la diferencia en el desempeño*.

Para examinar la perspectiva del porcentaje de olvido, citaron a Slamecka y McElree (1983), quienes definieron operacionalmente al olvido (*forgetting*) como la pérdida de información entre dos tests de retención, separados por un intervalo específico de tiempo. En ese sentido, la data puede graficarse de tal manera que la cantidad de información retenida se registra en el eje de la Y y el tiempo en el eje X. La pendiente (*slope*) de la función que relaciona estas dos variables refleja la tasa global de olvido que ha ocurrido entre los dos tests. La interpretación es clara: a mayor pendiente, mayor olvido. Usando esta interpretación, podría

determinarse si la pendiente de la función para el método de palabra clave es diferente a la de las demás estrategias de aprendizaje.

Por su parte, la perspectiva de la diferencia en el desempeño no toma en consideración el porcentaje de olvido. Si, por ejemplo, los porcentajes de olvido son constantes, todavía podría afirmarse que el método de palabra clave sigue siendo más efectivo a largo plazo porque la ventaja inicial se mantiene en el tiempo. Incluso, la función de olvido del método de palabra clave podría ser más inclinada, es decir, tener mayor pendiente que la del método comparativo y seguir siendo más efectivo siempre y cuando su desempeño global supere al del grupo control.

Después de revisar la literatura sobre la efectividad del método de palabra clave en la retención a largo plazo y habiéndose inclinado por la perspectiva del porcentaje de olvido, Wang y otros (1992) criticaron el uso de diseños intra-participantes en todos esos estudios previos porque, en su opinión, este tipo de diseño no aporta información real con respecto al porcentaje de olvido ya que el test inmediato sirve como una práctica adicional o como una oportunidad para probar cuán apropiado son los caminos al recuerdo (*retrieval paths*); y en consecuencia, inflan el desempeño en el test tardío. Predijeron que en los diseños inter-participantes, el método de palabra clave exhibiría mayor porcentaje de olvido que el método control y lo indicaría el efecto estadísticamente significativo de la interacción entre la condición de aprendizaje y el intervalo de retención; el efecto no estadísticamente significativo de dicha interacción, reflejado por el paralelismo de las funciones de olvido de los métodos comparados, indicaría un porcentaje de olvido estadísticamente comparable. Como veremos seguidamente, Wang, Thomas y sus colaboradores desarrollaron un programa de investigación en el que impulsaron mayoritariamente la utilización de diseños inter-participantes.

Wang y otros (1992) realizaron cuatro experimentos en los que compararon la efectividad del método de palabra clave versus la del método de repetición. En los tres primeros experimentos implementaron un diseño factorial 2 x 2 en el que

manipularon tanto la condición de aprendizaje (palabra clave vs. repetición) como el intervalo de retención (inmediato vs. tardío) como factores inter-participantes. Sin embargo, en el Experimento 4 manipularon el intervalo de retención como un factor intra-participantes. En todos los experimentos, el test tardío se aplicó una semana después de haberse administrado el test inmediato.

En el primer experimento. 79 hablantes nativos del inglés, estudiantes de psicología, participaron voluntariamente en este estudio. Los participantes fueron asignados al azar a una de las dos condiciones de aprendizaje y a uno de los dos intervalos de retención. El número de participantes en cada condición experimental varió entre 17 y 22. Las palabras experimentales fueron 22 sustantivos franceses. Los participantes asignados al método de palabra clave recibieron instrucción detallada del funcionamiento del método, reforzada posteriormente a través de un ejemplo presentado por uno de los investigadores. Seguidamente se le entregó un folleto (*booklet*) a cada participante. La primera página contenía una lista con las 22 palabras experimentales y sus correspondientes palabras claves. Luego de escuchar la pronunciación de cada palabra experimental y de su correspondiente palabra clave, se les pidió pasar a la página siguiente, en la cual la lista aparecía secuenciada de manera diferente. De nuevo los participantes escucharon cada par de palabras (i.e., palabra experimental-palabra clave). Seguidamente les presentaron dos nuevas listas. Luego, se les dijo que se les daría el equivalente en inglés de cada palabra francesa y se les recordó la importancia de crear las imágenes interactivas. Entonces se les presentó cada palabra experimental, su palabra clave y su equivalente en inglés y se les concedió 10 segundos para estudiar cada ítem. Se les presentó una nueva lista con una secuencia diferente. Por su parte, los participantes asignados al método de repetición escucharon cada palabra experimental por dos segundos un total de cuatro veces. Luego les concedieron 10 segundos para estudiar cada palabra experimental y su equivalente en inglés. La lista completa se les presentó dos veces y se les pidió copiar el equivalente en inglés cinco veces en un espacio en blanco, reservado para tal fin al lado de cada par de palabras.

Los participantes asignados a la condición de intervalo de retención inmediata tuvieron 10 segundos para escribir el equivalente en inglés de cada palabra experimental luego de que ésta fuera pronunciada en voz alta por el investigador que dirigió el experimento; los que utilizaron el método de palabra clave debieron escribir además la palabra clave en paréntesis. Por su parte, los participantes asignados a la condición de recuerdo tardío completaron la misma prueba una semana más tarde.

Los resultados del análisis de varianza (ANOVA) indicaron que tanto los efectos principales del intervalo de retención y de la condición de aprendizaje como el efecto de su interacción alcanzaron significación estadística. Como se esperaba, se produjo un notable descenso en el recuerdo de los participantes después de una semana, $F(1, 22) = 70,94$, $MSe = 8,88$. Por otra parte, el efecto de la condición de aprendizaje favoreció al método de palabra clave, $F(1, 22) = 12,7$; sin embargo, este hallazgo debe interpretarse a la luz del efecto estadísticamente significativo de la interacción entre la condición de aprendizaje y el intervalo de retención, $F(1, 22) = 8,51$. Aunque el método de palabra clave fue muy superior al de repetición en el recuerdo inmediato de vocabulario, $F(1, 22) = 22,80$, no hubo diferencia estadísticamente significativa entre los métodos una semana más tarde, $F < 1$. De hecho, en el intervalo de una semana el grupo del método de palabra clave olvidó, en promedio, un poco más del doble que el grupo del método de repetición (13,30 vs. 6,45, respectivamente). Es interesante destacar que el recuerdo de las palabras claves fue bastante bueno, aún una semana más tarde. Los estudiantes que tomaron el test inmediato recordaron un promedio de 20,5 palabras claves mientras que los que tomaron el test tardío recordaron un promedio de 15 palabras claves.

En el segundo experimento participaron 40 hablantes nativos del inglés, cursantes de los primeros niveles de psicología. Las palabras experimentales en éste y en los restantes experimentos fueron 20 sustantivos concretos del *Tagalog*, el idioma nacional de Filipinas. Se utilizaron los mismos procedimientos del Experimento 1 con la excepción de que la asociación de las palabras

experimentales y las palabras claves, así como la exposición del grupo de repetición a las palabras experimentales se redujeron a dos presentaciones como consecuencia de los hallazgos observados en un estudio piloto.

Los resultados mostraron un patrón idéntico al observado en el experimento anterior. Hubo efectos altamente significativos para el intervalo de retención, $F(1, 12) = 81,77$, $MSe = 3,17$, condición de aprendizaje, $F(1, 12) = 58,35$, y la interacción entre condición x intervalo de retención, $F(1, 12) = 41,72$. De nuevo, el desempeño de los participantes asignados al método de palabra clave fue significativamente superior al de los asignados al método de repetición en la prueba inmediata, $F(1, 12) = 99,37$, pero estadísticamente equivalente en la prueba tardía, $F(1, 12) < 1$. Asimismo, las palabras claves se recordaron muy bien en ambas pruebas (19,4 y 18,0 en el test inmediato y en el test tardío, respectivamente).

En el Experimento 3, los investigadores implementaron el mismo diseño de investigación usados en los dos experimentos previos. Los participantes fueron 60 hablantes nativos del inglés, estudiantes de la carrera de psicología, y asignados al azar en la misma cantidad a una de las cuatro condiciones experimentales: palabra clave-retención inmediata; palabra clave-retención tardía; repetición-retención inmediata; repetición-retención tardía. Las palabras experimentales fueron 20 sustantivos concretos del Tagalog (17 de la lista utilizada en el Experimento 2 y 3 palabras adicionales). Los procedimientos fueron modificados a los fines de lograr que el desempeño inicial de ambos métodos fuera estadísticamente similar. En tal sentido, los estudiantes asignados a los grupos de repetición no fueron expuestos a la lista que contenía únicamente las palabras experimentales; en cambio, se les presentaron 10 listas de las palabras experimentales y sus equivalentes en inglés, con cada par presentado por 3 segundos. Por su parte, los procedimientos para los estudiantes asignados a los grupos de palabra clave fueron idénticos a los utilizados en los Experimentos 1 y 2, excepto que sólo se les presentó una sola lista contentiva de palabras experimentales, palabras clave y equivalentes en inglés y se les concedió 15

segundos para aprenderse cada tripleta. De allí que el tiempo total de estudio fue 500 segundos para los grupos del método de palabra clave y 600 para los del método de repetición.

Los resultados del ANOVA revelaron un efecto estadísticamente significativo para el intervalo de retención, $F(1, 16) = 98,21$, $MSe = 5,61$ y para la interacción entre la condición de aprendizaje y el intervalo de retención, $F(1, 16) = 11,80$. El efecto del Intervalo de retención indicó que, independientemente del método de aprendizaje, el recuerdo de vocabulario fue muy superior inmediatamente después del tratamiento que una semana más tarde. Por su parte, el efecto de la interacción alcanzó significación estadística porque, aunque la diferencia en recuerdo inmediato favoreció ligeramente al método de palabra clave, la del recuerdo tardío fue significativamente mayor para el método de repetición, $F(1, 16) = 10,01$. Como en los primeros dos experimentos, el promedio de palabras claves recordadas tanto en el test inmediato como en el tardío fue alto (18,8 y 15,8 palabras claves, respectivamente).

Finalmente, en el cuarto experimento, los 39 estudiantes participantes fueron asignados al azar a las condiciones de aprendizaje (20 al método de palabra clave y 19 al método de repetición), pero todos ellos completaron tanto la prueba inmediata como la tardía (una semana más tarde). Los procedimientos fueron idénticos a los utilizados en el Experimento 2, es decir, al grupo del método de palabra clave le presentaron dos listas con las palabras experimentales y las palabras claves correspondientes mientras que al grupo del método de repetición le presentaron dos listas que incluyeron sólo las palabras experimentales. Luego, a todos los participantes les presentaron dos listas de estudio que incluían los equivalentes en inglés (palabra experimental-equivalente en inglés para el grupo de repetición; palabra experimental-palabra clave-equivalente en inglés para el grupo de palabra clave). Para cada lista, los ítems se presentaron a un intervalo de 10 segundos. A todos los participantes se les administró un test de recuerdo inducido (*cued recall*), seguido de una prueba de pareo (*matching test*) al finalizar la primera sesión y al comienzo de la segunda sesión una semana más tarde.

Los resultados indicaron que los efectos principales (condición de aprendizaje e intervalo de retención) así como el efecto de su interacción fueron estadísticamente significativos, $F(1, 12) = 9,27$, $F(1, 12) = 48,89$ y $F(1, 12) = 5,44$, respectivamente. Los tests post hoc de la interacción indicaron que el método de palabra clave fue superior al de repetición tanto en la prueba de recuerdo inmediato, $F(1, 12) = 17,65$ como en el tardío, $F(1, 12) = 3,28$, aunque la diferencia fue mayor en el test inmediato. Finalmente, el promedio de palabras claves recordadas fue extremadamente alto (19,27 y 18,82 respectivamente en el test inmediato y en el test tardío).

Wang, Thomas, Inzana, y Primicerio (1993), en un intento por replicar los resultados obtenidos por Wang y otros (1992), compararon la efectividad del método de palabra clave y el método de repetición bajo el paradigma de aprendizaje intencional. Un total de 89 estudiantes de una clase de psicología general participaron en este estudio; sin embargo, dos de ellos fueron excluidos del análisis por su familiaridad con el lenguaje target. Los participantes fueron asignados al azar a una de dos condiciones de aprendizaje (método de palabra clave vs. método de repetición) y a uno de dos intervalos de retención (inmediato y tardío). Las palabras experimentales fueron 24 sustantivos concretos del Tagalog.

Los participantes del grupo asignado al método de palabra clave fueron expuestos a dos listas que incluían las palabras experimentales y sus respectivas palabras claves. Los pares de palabras fueron presentados visualmente y pronunciados en voz alta por uno de los investigadores cada 4 segundos. Seguidamente se les presentaron otras dos listas que incluían además los correspondientes equivalentes en inglés. Cada tripleta (Palabra experimental-palabra clave-equivalente en inglés) se les presentó durante 10 segundos. Todas las listas de pares y tripletas tuvieron secuencias distintas de presentación. Los participantes en la condición de repetición fueron expuestos a 10 presentaciones con secuencias generadas al azar en las que cada par de palabras (palabra experimental-equivalente en inglés) aparecía por 3 segundos y era pronunciada

por uno de los investigadores. Luego de la fase de adquisición, los participantes completaron una tarea distractora de 3 minutos. Una vez completada dicha tarea se administró el test inmediato. La retención se midió a través de un test de recuerdo inducido, seguido de un test de pareo asociativo. Los participantes dispusieron de 5 minutos para completar cada test. En el test de recuerdo inducido, el grupo de repetición debió escribir el equivalente en inglés al lado de cada palabra experimental mientras que el grupo del método de palabra clave debió escribir además la palabra clave. En el test de pareo asociativo, todos los participantes debieron aparejar las palabras experimentales con sus correspondientes equivalentes en inglés. Debido a que se utilizó el paradigma de aprendizaje intencional, se les informó a todos los participantes antes de iniciar el experimento que se administraría un test de recuerdo luego de completar el mencionado experimento. A los participantes asignados al test tardío se les recordó regresar en 2 días para aplicárseles dicho test.

Los resultados del ANOVA para el test de recuerdo inducido indicaron efectos estadísticamente significativos para la condición de aprendizaje, $F(1, 85) = 4,94$, $MSe = 26,24$, $p < 0,03$, el intervalo de retención, $F(1, 85) = 36,92$, $p < 0,001$ y para su interacción, $F(1, 85) = 3,83$, $p < 0,06$. Los análisis de la interacción revelaron que, aunque el desempeño de ambos métodos fue similar en el test inmediato, varió notablemente dos días después; el método de repetición superó ampliamente al método de palabra clave. De hecho, el grupo de palabra clave olvidó casi el doble de palabras experimentales que el grupo de repetición. Por su parte, los resultados para la prueba de pareo asociativo revelaron efectos estadísticamente significativos para el intervalo de retención, $F(1, 85) = 15,79$, $MSe = 25,42$, $p < 0,001$ y para la interacción entre condición de aprendizaje e intervalo de retención, $F(1, 85) = 6,49$, $p < 0,02$. Los análisis de este efecto interactivo indicaron que, aun cuando el método de palabra clave aventajó al de repetición en el test inmediato, esta ventaja desapareció en el test tardío. De hecho, en este test el grupo de repetición superó al grupo del método de palabra clave. Finalmente, el recuerdo de palabras claves fue bastante alto, aunque descendió ligeramente después de dos días (22,42 vs. 20,27).

Wang y Thomas (1995) realizaron tres experimentos en los que compararon la efectividad del método de palabra clave y el método de contexto. El Experimento 1 fue excluido de la presente investigación porque las palabras experimentales, tomadas del estudio de McDaniel, Pressley y Dunay (1987), fueron palabras poco frecuentes (*obscure words*) en inglés, la L1 de los participantes.

En el Experimento 2, 48 estudiantes universitarios de un curso de psicología general fueron asignados al azar a una de cuatro condiciones experimentales (palabra clave-inmediato; palabra clave-tardío; contexto-inmediato, contexto-tardío). Los investigadores implementaron un paradigma de aprendizaje incidental. Las palabras experimentales fueron los mismos 24 sustantivos concretos del tagalog usados en Wang y otros (1993). Luego de ser instruidos en la lógica del método de palabra clave, los participantes asignados a dicho método fueron expuestos a cada tripleta (palabra experimental-palabra clave-equivalente en inglés) por 12 segundos. La lista completa fue presentada dos veces. Por su parte, los asignados al método de contexto fueron expuestos en tarjetas (*flashcards*) por 9 segundos a la palabra experimental y dos oraciones en inglés que incluían la palabra experimental. Seguidamente se volteaba la tarjeta por 3 segundos para que los estudiantes pudieran leer el equivalente en inglés. Al igual que los asignados al método de palabra clave, tuvieron dos oportunidades para aprender el significado de las palabras experimentales. Finalmente, los participantes en la condición de recuerdo inmediato completaron una tarea distractora de 3 minutos, seguida de un test de recuerdo inducido de 5 minutos. A los participantes asignados al método de palabra clave se les concedió otros 5 minutos para completar el test de recuerdo de las palabras claves. Los participantes en la condición de recuerdo tardío completaron el test de recuerdo inducido 2 días después.

Los resultados del ANOVA indicaron que el efecto principal del intervalo de retención y el de su interacción con la condición de aprendizaje fueron

estadísticamente significativos, $F(1, 19) = 36,57$, $p < 0,001$, $MSe = 9,39$ y $F(1, 19) = 28,51$, $p < 0,001$. El desempeño del grupo de palabra clave en el test inmediato fue muy superior al del grupo de contexto ($M = 18,50$, $DS = 3,15$ vs. $M = 8,70$, $DS = 3,77$), $t(11) = 5,02$, $p < 0,001$. Sin embargo, en el test tardío, la situación cambió radicalmente pues el desempeño del grupo de contexto fue superior al de palabra clave ($M = 7,80$, $DS = 2,17$ vs. $M = 3,80$, $DS = 2,49$), $t(8) = 2,71$, $p < 0,03$. Finalmente, tanto el recuerdo inmediato de las palabras claves ($M = 23,28$) como el tardío ($M = 19,66$) fue muy elevado.

En el Experimento 3, los investigadores usaron un diseño mixto en el que la condición de aprendizaje (método de palabra clave, método de contexto) y el intervalo de retención (inmediato, tardío) fueron manipulados como factores inter-participantes mientras que la frecuencia con la que se presentaron las 24 palabras experimentales (1 vez, 3 veces, 5 veces) fue manipulada como un factor intra-participantes, es decir, se presentaron 8 palabras una vez; 8 palabras tres veces y las restantes 8 cinco veces para conformar una lista de 72 ítems. Un total 60 estudiantes con características similares a los del experimento anterior (hablantes nativos del inglés, registrados en cursos de psicología general y sin competencia en tagalog) fueron asignados al azar en números iguales ($n = 15$) a una de cuatro condiciones inter-participantes (palabra clave-inmediato; palabra clave-tardío; contexto-inmediato, contexto-tardío). Se construyeron dos listas con secuencias diferentes para minimizar el efecto de la secuencia de presentación. En dichas listas se controló el número de ítems entre las repeticiones de las palabras experimentales presentadas 3 veces y 5 veces. Las palabras experimentales fueron presentadas cada 10 segundos para un total de 720 segundos (72×10) o 12 minutos ($720/60$).

Los resultados del ANOVA mostraron que los efectos principales de la condición de aprendizaje, el intervalo de retención y la frecuencia fueron estadísticamente significativos, indicando que el método de palabra clave fue superior al de contexto, $F(1, 24) = 17,10$, $p < 0,001$, $MSe = 2,33$; que el desempeño de los participantes fue significativamente mejor en el test inmediato

que 2 días más tarde, $F(1, 24) = 44,66$, $p < 0,001$; y que las palabras presentadas 5 veces se recordaron mejor que las presentadas 3 veces y éstas se recordaron más fácilmente que las presentadas una vez, $F(2, 48) = 139,54$, $p < 0,001$, $MSe = 0,87$. Asimismo, la interacción triple alcanzó significación estadística, $F(2, 48) = 2,89$, $p < 0,05$.

Para examinar la mencionada interacción triple, los investigadores realizaron tres ANOVAs, una para cada nivel de frecuencia. Los resultados de estas tres ANOVAs indicaron que la única condición x intervalo de retención que alcanzó significación estadística fue para los ítems practicados una vez, $F(1, 24) = 5,99$, $p < 0,03$, $MSe = 1,40$. En otras palabras, el patrón de porcentaje de olvido distinto observado en la interacción 1 x condición fue responsable de que se produjera la significativa interacción triple. En cambio, cuando los estudiantes tuvieron más de una oportunidad de practicar las palabras experimentales, los porcentajes de olvido para los métodos de palabra clave y contexto fueron estadísticamente comparables (i.e., paralelas). Es importante destacar que cuando los ítems se estudiaron solamente una vez, el porcentaje de olvido fue mayor en el método de palabra clave que en el método de contexto.

Thomas y Wang (1996) realizaron tres experimentos, pero el tercero fue excluido de la presente investigación debido a que en él no se comparó el método de palabra clave con ningún otro método de aprendizaje de vocabulario de FL, sino dos variantes del método de palabra clave. En el Experimento 1, los investigadores emplearon un diseño factorial 2 x 3 en el que tanto el intervalo de retención (5 minutos, 2 días) como la estrategia de aprendizaje (palabra clave generada, palabra clave dada, repetición) fueron manipulados como factores inter-participantes. Las palabras experimentales fueron 24 sustantivos concretos del tagalog. Los participantes fueron 85 hablantes nativos del inglés, matriculados en un curso introductorio de Psicología y sin competencia en tagalog. A los estudiantes asignados a la generación de las palabras claves se les instruyó sobre la lógica del método de palabra clave y se les proporcionó varios ejemplos, usando

el español como L2. Seguidamente se les presentó cada palabra experimental y se les pidió que escribieran la mejor palabra clave que les viniera a la mente. Luego se les distribuyó una lista con las palabras experimentales, los equivalentes en inglés y un espacio en blanco para escribir la palabra clave. Se les concedió 6 minutos para estudiar la lista y escribir las palabras claves. Se les recordó formar imágenes interactivas vívidas que combinaran las palabras claves y los equivalentes en inglés. Se les pidió colocar una marca (*check*) al lado de cada ítem para el que habían formado una imagen interactiva. Después de la fase de estudio, se les administró un test de conocimiento general (e.g., “¿Cuál río es más largo: el Nilo o el Amazona?”). Los participantes asignados al intervalo de retención de 5 minutos tomaron el test de recuerdo inducido, el cual incluyó la lista de las 24 palabras experimentales y espacios en blanco para escribir las palabras claves y los equivalentes en inglés. Se les concedió 5 minutos para completar dicho test. Por su parte, a los asignados al grupo tardío se les pidió regresar 2 días después para aprender un material nuevo con una estrategia de aprendizaje diferente. Al retornar tomaron el test de recuerdo inducido. Al grupo asignado al método de palabra clave estándar se le presentaron dos listas con las palabras tagalog y las palabras claves, cada par a intervalos de 5 segundos. El resto del procedimiento fue idéntico al utilizado con el grupo de participantes que generó las palabras claves. Finalmente, al grupo de repetición se le presentó una lista con las 24 palabras experimentales y sus respectivos equivalentes en inglés a un intervalo de 15 segundos por palabra experimental. Cada par (palabra tagalog-equivalente en inglés) apareció escrito en la parte de arriba de una página del folleto, seguido de 10 líneas. A los participantes de esta condición se les pidió escribir el equivalente en inglés tantas veces como fuese posible en los 15 segundos mientras pensaban en la palabra tagalog. Al finalizar esta sesión, se les administró el test de conocimiento general y el test de recuerdo inducido inmediato o tardío.

Los resultados del ANOVA 2x3 produjeron efectos principales estadísticamente significativos para el intervalo de retención, $F(1, 79) = 110,15$, $MSe = 21,23$, $p < 0,001$, y condición de aprendizaje, $F(2, 79) = 7,21$, $p < 0,002$. Su interacción también alcanzó significación estadística, $F(2, 79) = 21,56$, $p < 0,001$.

Tests posteriores indicaron que en la prueba de recuerdo inmediato, el desempeño del grupo asignado al método de palabra clave estándar fue superior al de los otros dos grupos, los cuales no difirieron estadísticamente. Sin embargo, después de 2 días, el grupo de repetición fue capaz de recordar estadísticamente más palabras que cualquiera de las dos variantes del método de palabra clave. Finalmente, el recuerdo de las palabras claves fue muy alto: el grupo que generó las palabras claves recordó 21,6 y 21,2 palabras claves después de 5 minutos y después de 2 días, respectivamente mientras que el grupo que las recibió, recordó 23,9 palabras claves después de 5 minutos y 20,3 al cabo de 2 días.

En el Experimento 2, Thomas y Wang implementaron un diseño factorial 2 x 3 con dos variables manipuladas como factores inter-participantes: intervalo de retención (5 minutos y 2 días) y condición de aprendizaje (palabra clave estándar, palabra clave-dibujo y repetición). Los participantes fueron 103 estudiantes de un curso introductorio de psicología, sin ningún conocimiento del tagalog. Hubo entre 8 y 21 participantes en cada una de las seis condiciones experimentales. Las palabras experimentales fueron 24 palabras del tagalog, 9 palabras nuevas y 15 palabras usadas en el Experimento 1.

Los participantes del grupo de palabra clave estándar recibieron instrucción sobre el funcionamiento del método, especialmente sobre la importancia de generar imágenes interactivas vívidas. Practicaron el enlace acústico, utilizando varios ejemplos con palabras del español. Seguidamente se les presentó la lista de las 24 palabras tagalog y sus correspondientes palabras claves dos veces con una secuencia diferente a un intervalo de 10 segundos. Antes de la segunda presentación de la lista, se les recordó formar las imágenes interactivas. Al finalizar la sesión de estudio, los participantes asignados al intervalo de retención de 5 minutos tomaron la prueba de conocimiento general y seguidamente el test de recuerdo inducido. El test consistía en la lista de las 24 palabras tagalog, sus correspondientes palabras claves y espacios en blanco para escribir los equivalentes en inglés. Se les concedió 5 minutos para completar el mencionado test. Por su parte, los participantes asignados al intervalo de retención de 2 días

recibieron el mismo tratamiento, excepto que se les administró el test de recuerdo inducido 2 días más tarde. A los del grupo de palabra clave-dibujo también se les instruyó sobre el funcionamiento del método de palabra clave. Sin embargo, las fichas que mostraban la palabra tagalog y la palabra clave incluían también los dibujos que denotaban la palabra clave y el equivalente en inglés. Dichos dibujos aparecían directamente sobre las palabras impresas. Se les pidió que generaran una imagen interactiva que involucrara a los ítems dibujados. El resto del procedimiento fue idéntico al seguido por el grupo del método de palabra clave estándar, excepto que luego del clásico test de recuerdo inducido, se les administró un test de recuerdo inducido por los dibujos. Se les mostró la ficha con la palabra tagalog, su palabra clave y el dibujo de la palabra clave por 10 segundos y se les pidió que escribieran el equivalente en inglés. Finalmente, a los asignados al método de repetición se les presentó cada par de palabra tagalog-equivalente en inglés a razón de 2 segundos por par. En total, cada par se les presentó 5 veces, secuenciadas al azar. La mitad de este grupo recibió el clásico test de recuerdo inducido inmediatamente después de completar la prueba de conocimiento general mientras que la otra mitad se retiró luego de la prueba de conocimiento general y regresó 2 días después para tomar el test de recuerdo inducido.

Los resultados del ANOVA 2 x 3 indicaron que hubo un efecto estadísticamente significativo para el intervalo de retención, $F(1, 97) = 126,82$; $MSe = 21,27$; $p < 0.001$. Asimismo, el efecto de la interacción entre el intervalo de retención y la condición de aprendizaje alcanzó significación estadística, $F(2, 97) = 3,26$; $p = 0,04$. Esta interacción se debió a que mientras en el test inmediato el desempeño de los grupos comparados fue estadísticamente comparable, 2 días más tarde, el grupo asignado al método de palabra clave estándar recordó significativamente menos palabras que el asignado al método de repetición; no hubo diferencias estadísticamente significativas entre el grupo palabra clave-dibujo y los otros dos grupos.

Con el propósito de replicar y extender los hallazgos reportados por Wang, Thomas y sus colaboradores, Carney y Levin (1998) realizaron 5 experimentos en los que compararon la efectividad del método de repetición versus la de diferentes variantes del método de palabra clave. En el Experimento 1, 77 estudiantes de cursos avanzados de psicología participaron voluntariamente en el mencionado estudio. Se utilizó un diseño factorial 2 x 2. En otras palabras, los participantes fueron asignados al azar a uno de dos métodos de aprendizaje: repetición y la variación estándar del método de palabra clave y a uno de dos intervalos de retención: inmediato y 1 semana más tarde. Las palabras experimentales fueron 24 sustantivos concretos del tagalog.

A los aprendices del método de repetición se les mostró la lista de palabras a aprender y sus correspondientes significados y se les pidió repetir una y otra vez cada palabra experimental y su significado mientras que a los del método de palabra clave se les dio una explicación del funcionamiento del método de palabra clave. Seguidamente se les presentó la lista de palabras experimentales y las palabras claves y se les pidió generar sus propias imágenes interactivas. Al final de esta sección, todos los participantes estudiaron 5 palabras de práctica. La sección siguiente representó la fase de familiarización, en la que a todos los participantes se les presentó la lista con las 24 palabras experimentales (a los participantes del método de palabra clave se les proporcionaron también las palabras claves correspondientes). La lista se presentó dos veces y cada palabra se mostró por 10 segundos. En la fase de estudio propiamente dicha, se les recordó a los participantes utilizar su método de aprendizaje con dos palabras de práctica, las cuales se mostraron a un intervalo de 15 segundos. A continuación, se les mostró la lista de las 24 palabras tagalog dos veces a razón de 15 segundos por palabra. Mientras al grupo de repetición se le instruyó estudiar las 24 palabras experimentales y sus correspondientes significados utilizando la repetición, al grupo del método de palabra clave se les pidió estudiar las 24 palabras y sus significados usando las palabras claves dadas y las imágenes que habían generado. Después de una tarea de 5 minutos que sirvió como distractor, la mitad del grupo de repetición y la mitad del grupo del método de palabra clave

tomaron el test inmediato. Para cada ítem, se le solicitó al grupo del método de palabra clave suministrar tanto el significado como la palabra clave. El resto de los estudiantes no tomó el test inmediato, sino que respondió algunas preguntas sobre su experiencia en el experimento; por ejemplo, “¿Cuán fácil te resultó usar tu método de aprendizaje?” “¿Usarías este método de nuevo?”. Todos los participantes retornaron siete días después para tomar una prueba tardía de recuerdo (primer test para la mitad de los participantes y segundo para la otra mitad) y completar un cuestionario.

Los resultados indicaron que el grupo que utilizó el método de palabra clave superó ampliamente al grupo que utilizó el método de repetición en el test inmediato, $t(36) = 3,80$, $p < 0,001$, $ES = 1,23$. Sin embargo, para estos mismos estudiantes (el tradicional diseño intra-participantes), la ventaja inicial desapareció una semana después. Este hallazgo es contrario al reportado por Wang, Thomas y Ouellette, 1992, Exp. 4). Para los estudiantes que no tomaron el test inmediato (el diseño inter-participantes del estudio), el desempeño de ambos métodos fue estadísticamente comparable en el test tardío, $t(37) = 1,12$, $p > 0,20$, lo cual es consistente con lo reportado por Wang y otros (1992, Exps. 1-2). El bajo desempeño de los estudiantes del método de palabra clave en el test tardío no fue atribuible al olvido de las palabras claves pues fueron capaces de recordar un promedio de 90 y 80% cuando tomaron el test inmediato y cuando no lo hicieron, respectivamente.

El Experimento 2 fue diseñado para investigar el argumento expuesto por Wang y otros (1992) de que el tomar el test inmediato hace que la superioridad mnemónica se mantenga en el test tardío, tal como lo reportaron en su Experimento 4. Difirió del Experimento 1 en dos aspectos importantes. En primer lugar, se adoptó un diseño diferente. Inicialmente, los participantes completaron un test que incluía únicamente la mitad de los ítems estudiados (12 de 24) y al retornar una semana después completaron otro con todos los 24 ítems incluidos. Los dos subconjuntos de 12 ítems fueron seleccionados al azar y contrabalanceados dentro de las condiciones experimentales. Asimismo, en los

materiales de estudio y en el test tardío, los ítems de las dos listas se secuenciaron al azar, por ejemplo, un ítem que había sido examinado seguía a uno que no lo había sido, o viceversa).

En segundo lugar, pusieron a prueba la hipótesis de la práctica del camino al recuerdo de vocabulario (*retrieval practice*) para contraponerla a la explicación de Wang y otros (1992) de que el test inmediato sirve como práctica y que como los estudiantes que utilizan el método de palabra clave recuerdan más definiciones en el test inmediato, obtienen una práctica adicional para muchos de los ítems estudiados. La hipótesis de Carney y Levin (1998), por su parte, planteaba que el test inmediato podía convertirse en una oportunidad para que los estudiantes del método de palabra clave repasasen sus caminos al recuerdo de vocabulario, es decir, la activación de los enlaces, aumentando de esta manera la probabilidad de recordar el significado de las palabras experimentales en el próximo encuentro.

Un total de 118 estudiantes universitarios participaron en este estudio. Fueron asignados al azar a una de 4 condiciones experimentales: repetición estándar, método de palabra clave estándar, *retrieval practice* (repetición) y *retrieval practice* (método de palabra clave). Mientras las condiciones “estándar” fueron muy similares a las del Experimento 1, las del “*retrieval practice*” proporcionaron a los estudiantes test de práctica y feedback para cada ítem. En las condiciones de “*retrieval practice*”, los participantes tuvieron 20 segundos para escribir el significado de cada palabra tagalog (el grupo del método de palabra clave debió escribir además la correspondiente palabra clave). Seguidamente debieron ver el significado de la palabra y usar el tiempo restante para revisar la palabra tagalog (y la palabra clave) y su significado. Después de una tarea distractora de 5 minutos, todos los participantes tomaron el test de recuerdo inmediato sobre uno de los dos conjuntos de 12 palabras tagalog. Luego de 7 días, los estudiantes regresaron para tomar un test, sin previo aviso, sobre todas las 24 palabras tagalog y completar un cuestionario.

Los resultados indicaron que la *retrieval practice* no mejoró significativamente la efectividad del método de palabra clave en ninguna de las

tres mediciones analizadas: recuerdo inmediato y las dos de recuerdo tardío (ítems examinados e ítems no examinados). Asimismo, indicaron que el método de palabra clave superó ampliamente al método de repetición en el test inmediato, $t(114) = 3,83$ y $3,56$ para las condiciones estándar y *retrieval practice*, respectivamente; ambas $ps < 0,001$, $ES = 1,00$ y $0,93$. El tamaño de los efectos fue mucho más pequeño en el test tardío. Para las palabras examinadas previamente (parte intra-participantes del diseño), el método de palabra clave fue ligeramente mejor que el de repetición, $t(110) = 1,84$ y $1,74$ respectivamente para las condiciones estándar y *retrieval practice*, ambas $ps < 0,05$, $ES = 0,48$ y $0,3$. Este modesto efecto a favor del método de palabra clave en el test tardío de un contexto intra-participantes es consistente con el reportado en Wang y otros (1992, Exp. 4). Para los ítems no examinados previamente (parte inter-participantes del diseño) la efectividad de los métodos de aprendizaje fue estadísticamente comparable, $t(110) = 1,17$ y $1,01$, ambas $ps > 0,20$ para las condiciones estándar y *retrieval practice*, respectivamente. Es importante señalar que ambas diferencias favorecieron al método de palabra clave.

En el Experimento 3, 107 participantes fueron asignados aleatoriamente a una de tres condiciones experimentales: repetición, palabra clave + interacción (palabras claves e imágenes interactivas proporcionadas por los investigadores), o palabra clave + sonido + interacción (además de lo anterior, se agregó un sonido asociado a cada palabra clave). Los participantes fueron estudiantes universitarios en cursos de psicología educativa. El material de estudio para los grupos de repetición y de palabra clave + interacción fue muy parecido al utilizado en el Experimento 1 con la adición de una oración que describía una interacción entre la palabra clave y el significado en la segunda condición. Para el grupo de palabra clave + sonido + interacción se preparó un material especial en el que cada palabra clave fue asociada con un sonido característico (e.g., una pelota “rebotando en un piso de concreto”. Cada imagen sonora fue incorporada a la imagen interactiva proporcionada en el material de estudio,

Los estudiantes tuvieron 7 minutos para leer una sección titulada “tu estrategia” que describía el método que le había sido asignado. Además, esta sección incluía 5 ítems de práctica y su respectivo test. Al finalizar dicha sección, completaron la fase de familiarización en la que al grupo de repetición se le presentó dos veces una lista de 24 palabras tagalog a un intervalo de 10 segundos por palabra. De igual manera, al grupo de palabra clave + interacción se le presentó dos veces la lista de las 24 palabras y sus correspondientes palabras claves. Luego de completar una tarea distractora de 2 minutos, se administró el test inmediato para la mitad de las palabras tagalog (12 de 24). Los estudiantes dispusieron de 7 minutos para completar dicho test. Dependiendo de su condición, los grupos de palabra clave debieron escribir también las palabras claves o las palabras claves y los sonidos claves. Los estudiantes regresaron 5 días después para completar un test tardío no anunciado sobre todas las 24 palabras experimentales.

Los resultados de las tres pruebas analizadas (recuerdo inmediato, recuerdo tardío para las palabras examinadas y recuerdo tardío para las palabras no examinadas) indicaron que los participantes en ambas condiciones mnemónicas fueron superiores al método de repetición, todos los $ESs > 0,80$. Asimismo, no hubo diferencia estadísticamente significativa entre las dos variantes del método de palabra clave, indicando que la hipótesis de que la adición de un enlace sonoro fortalecería aún más los enlaces activados en la condición de palabra clave + interacción, no fue confirmada en este estudio. Finalmente, los resultados confirmaron parcialmente los hallazgos reportados por Wang y sus colegas; fueron consistentes para los ítems examinados previamente, pero en contraste marcado para los no examinados previamente.

En el Experimento 4, un total de 94 estudiantes universitarios, inscritos en clases de psicología educativa, fueron asignados aleatoriamente a tres condiciones de aprendizaje: repetición, método de palabra estándar y método de palabra clave + interacción. Las palabras experimentales fueron 24 sustantivos tagalog, las cuales fueron divididas al azar en 3 sets de 8 palabras. Como en el

Experimento 3, a los estudiantes se les presentó dos veces la lista de 24 palabras tagalog y sus significados a un intervalo de 15 segundos por palabra. Después de concluida la sección de estudio y la tarea distractora de 2 minutos, los estudiantes completaron el test inmediato. Retornaron 2 días después y sorpresivamente se les administró un test tardío. Finalmente, regresaron 5 días más tarde para tomar otro test tardío sin previa notificación y completar un cuestionario sobre cómo habían estudiado.

Los mencionados tests se construyeron de la manera siguiente: Ocho palabras fueron examinadas en los tres tests. Un segundo set de 8 palabras fue examinado dos y cinco días después. Finalmente, un tercer set de 8 palabras fue examinado únicamente 5 días después. En consecuencia, el test inmediato incluyó 8 palabras, el primer test tardío (2 días después) 16 palabras y el segundo test tardío (5 días después) 24 palabras. Los participantes asignados a las dos variantes del método de palabra clave debieron recordar también las respectivas palabras claves.

Los resultados indicaron que no hubo diferencia estadísticamente significativa entre las dos variantes del método de palabra clave en ninguno de los tests. Asimismo, indicaron que en las palabras examinadas dos y tres veces, las variantes del método de palabra clave fueron ampliamente superiores al método de repetición, todas las $ps < 0,001$, todos los $ES > 1.23$. Por cierto, el porcentaje de olvido del método de repetición y el de las variantes del método de palabra clave fue estadísticamente comparable. Estos hallazgos son consistentes con lo planteado por Wang y otros (1992, Exp. 4) para diseños intra-participantes. Finalmente, para las palabras examinadas una vez, el efecto de la interacción entre la condición de aprendizaje y el intervalo de retención fue estadísticamente significativo, $F(2, 84) = 3,33$; $p < 0,05$. Asimismo, el porcentaje de olvido del método de palabra clave fue superior al del método de repetición, $t(84) = 2,43$; $p = 0,017$. Sin embargo, el desempeño de las dos variantes del método de palabra clave en el recuerdo de vocabulario fue estadísticamente superior al del método

de repetición en todas las mediciones, aún después de 5 días, todas las $ps < 0,002$, todos los $ESs > 0,79$.

Por último, en el Experimento 5, los participantes fueron 76 estudiantes universitarios en cursos de psicología educativa. Fueron asignados al azar a una de tres condiciones de aprendizaje: repetición intencional, palabra clave intencional y palabra clave incidental. Las palabras experimentales fueron las mismas 24 palabras tagalog usadas en los experimentos anteriores. Inicialmente, los estudiantes tuvieron 6 minutos para leer la sección titulada “tu estrategia” que describía el método al que habían sido asignados. En ella, se le pedía al grupo de repetición aprender el nuevo vocabulario repitiendo cada palabra y su significado una y otra vez, y se le informaba que se le administraría un test de recuerdo. Por su parte, al grupo de palabra clave intencional se le explicó el funcionamiento del método y se le informó que habría un test mientras que al grupo de palabra clave incidental, además de dicha explicación, se les dijo que ese día evaluarían algunos materiales de aprendizaje. Asimismo, se les dijo que dichos materiales habían sido diseñados para ayudar a los estudiantes a aprender un conjunto de 24 palabras y que su evaluación sería muy beneficiosa para mejorar los mencionados materiales. Todos los participantes practicaron 5 ítems en esta sección.

La siguiente sección fue la de familiarización. Al grupo de repetición se le presentó la lista de las 24 palabras tagalog; al grupo de palabra clave intencional se le presentó la lista de las 24 palabras y sus correspondientes palabras claves mientras que al grupo de palabra clave incidental se le pidió evaluar cada una de las 24 palabras claves en términos de su correspondencia (adecuación) con la palabra a aprender. En la sección final, se les recordó a los participantes apegarse a su estrategia e implementarla con 2 ítems de práctica, presentados en un intervalo de 20 segundos por palabra. Seguidamente se les pidió estudiar las 24 palabras experimentales a razón de 20 segundos por palabra o en el caso de la palabra clave incidental, evaluar las palabras claves. Luego de completar una tarea distractora de 5 minutos, se les concedió 6 minutos para completar el test inmediato con la mitad (12 de 24) de las palabras experimentales. Regresaron 2

días más tarde para completar un test no anunciado con todas las 24 palabras. De ellas, 12 no habían sido examinadas en el test inmediato. En ambos tests, las dos variantes del método de palabra clave debieron también recordar las palabras claves.

Los resultados indicaron que en el test inmediato, ambas variantes del método de palabra clave fueron estadísticamente superiores al método de repetición, $t(73) = 4,98$ y $2,53$ para el grupo de palabra clave intencional y el grupo de palabra clave incidental, respectivamente, $ES = 1,41$ y $0,72$. En el test tardío, solamente el método de palabra clave intencional fue estadísticamente superior al método de repetición. La ventaja fue sustancial para los 12 ítems que habían sido examinados previamente, 72 vs. 48%, $t(70) = 3,45$, $ES = 0,99$, y más modesta, aunque estadísticamente significativa para los 12 ítems no examinados previamente, 25 vs. 15%, $t(70) = 2,24$, $ES = 0,64$.

CAPITULO III

METODOLOGIA

Tipo de Investigación

La presente investigación es secundaria ya que se utilizó el material publicado por otros investigadores para abordar el tema examinado en ella. La utilización de dicho material fue amplia y detallada pues comprendió desde aspectos metodológicos hasta los hallazgos reportados en cada uno de los estudios incluidos. Además, permitió tomar parte de los datos presentados en los estudios primarios y reanalizarlos desde una perspectiva distinta.

Población y Muestra

En la presente investigación, la población estuvo representada por todos los estudios que conformaron el programa de investigación desarrollado por Wang, Thomas y sus colaboradores entre 1992 y 1996 para sustentar su crítica sobre la efectividad del método de palabra clave en la retención a largo plazo de vocabulario. Asimismo, fue integrada por un número importante de estudios (e.g., Ávila y Sadoski, 1996; Carney y Levin, 1998; Colón y Rodríguez, M., 2000; Rodríguez, M., y Sadoski, 2000), inspirados en los hallazgos reportados por esos investigadores. Sin embargo, para ser incluidos en la muestra definitiva, los estudios primarios debieron cumplir con ciertos criterios. Los criterios variaron para los estudios liderados por Wang y Thomas y para los que inspiraron sus hallazgos; para los primeros se aplicaron los siguientes dos criterios:

1. Las palabras experimentales pertenecían a una L2 para los participantes; y
2. Variante(s) del método de palabra clave fueron comparadas con el método de repetición o con el método de contexto.

El programa de investigación liderado por Wang y Thomas incluyó 11 estudios primarios, a saber: cuatro en Wang, Thomas y Ouellette (1992), uno en

Wang, Thomas, Inzana y Primicerio (1993); tres en Wang y Thomas (1995); y tres en Thomas y Wang (1996). La aplicación de los mencionados criterios hizo que se excluyeran 2 estudios: el Experimento 1 de Wang y Thomas (1995) y el Experimento 3 de Thomas y Wang (1996). En el primer caso, las palabras experimentales fueron palabras inglesas poco frecuentes (*obscure words*), la L1 de los participantes mientras que en el segundo caso, sólo se compararon variantes del método de palabra clave.

Para los segundos, además de los mencionados criterios, debieron examinar las conclusiones más importantes generadas a partir de los hallazgos reportados por Wang, Thomas y sus colegas. El hecho de que Carney y Levin (1998) en su artículo, reportaran los resultados de cinco estudios, realizados con el propósito de replicar y extender los hallazgos de Wang, Thomas y sus colaboradores, garantizó su inclusión en la muestra final de la presente investigación; el resto de los candidatos, al no cumplir con los criterios de inclusión establecidos, fueron excluidos. En tal sentido, la muestra final estuvo integrada por cinco artículos que incluyeron 14 estudios primarios, distribuidos de la siguiente manera: Wang y otros (1992, Exps. 1-4); Wang y otros (1993); Wang y Thomas (1995, Exps. 2-3); Thomas y Wang (1996; Exps. 1-2) y Carney y Levin (1998, Exps. 1-5).

Procedimientos

Una vez obtenidos los artículos que contenían los estudios primarios incluidos en la muestra definitiva de la presente investigación, se procedió a ordenarlos cronológicamente para revisarlos posteriormente en ese mismo orden. Sin embargo, dado que el tema central a examinar en la presente investigación fue el soporte empírico de la crítica de Wang, Thomas y sus colaboradores sobre la efectividad del método de palabra clave en la retención a largo plazo de vocabulario en L2, se decidió completar esta tarea antes de proceder a revisar los estudios primarios de Carney y Levin (1998).

Sucintamente, la crítica de Wang, Thomas y sus colegas se centra en el porcentaje de olvido del método de palabra clave y en el tipo de diseño implementado. En su opinión, el porcentaje de olvido de este método es mayor que el de diversos grupos control (e.g., repetición, contexto), pero está condicionado por el diseño utilizado. Mientras los diseños intra-participantes lo subestiman al inflar el desempeño del método en los tests tardíos, los inter-participantes generan porcentajes de olvido reales. Predijeron que con diseños intra-participantes, la marcada ventaja del método de palabra clave en el test inmediato se mantendría después de unos días; sin embargo, con diseños inter-participantes, la ventaja inicial del método desaparecería o incluso se revertiría al cabo de pocos días. Este hallazgo con diseños inter-participantes se reflejaría en el efecto estadísticamente significativo de la interacción entre la condición de aprendizaje y el intervalo de retención. En la presente investigación, el soporte empírico de la crítica de Wang, Thomas y sus colaboradores se examinó desde dos perspectivas diferentes: la perspectiva del porcentaje de olvido y la del desempeño del método de palabra clave.

Análisis desde la Perspectiva del Porcentaje de Olvido. Desde esta perspectiva, la adoptada por estos investigadores, el procedimiento fue relativamente sencillo. Consistió en revisar los resultados del ANOVA de cada uno de los estudios donde se usó un diseño inter-participantes para verificar si el efecto de la mencionada interacción había alcanzado significación estadística; en el único estudio donde se utilizó un diseño mixto (Wang y Thomas, 1995, Exp. 3), también se revisó la interacción entre estas dos variables. En ese sentido, se identificó el número de estudios en los que dicho efecto fue estadísticamente significativo. La interpretación fue muy clara y directa: a mayor número de estudios, mayor soporte empírico.

Ahora bien, para examinar hasta qué punto los hallazgos reportados en los estudios de Carney y Levin (1998) avalaban los reportados por Wang, Thomas y sus colegas, se empleó un procedimiento un poco más complejo. Dado que en los

cinco experimentos de Carney y Levin se implementó un diseño mixto, es decir, con una parte inter-participantes y otra intra-participantes, se examinó la parte inter-participantes del diseño. Con excepción del Experimento 1 en el que la mitad de los participantes completó tanto el test inmediato como el tardío (parte intra-participantes) mientras que la otra mitad únicamente tomó el test tardío (parte inter-participantes), en los restantes experimentos, las palabras experimentales fueron divididas aleatoriamente en dos o tres conjuntos y la parte inter-participantes del diseño estuvo representada por el conjunto de palabras examinadas una sola vez. Determinar si el efecto de la interacción entre la condición de aprendizaje y el intervalo de retención alcanzó significación estadística fue más difícil en estos estudios porque esta información fue proporcionada únicamente en el Experimento 4. En todos los demás experimentos, hubo que chequear el desempeño de los métodos comparados (palabra clave vs. repetición) tanto en el test inmediato como en el test tardío o para los ítems no examinados previamente. Incluso, en ocasiones, hubo que graficar las funciones de olvido de los métodos para tomar la decisión apropiada.

Análisis desde la Perspectiva del Desempeño del Método de Palabra Clave. A los fines de presentar una visión más completa del soporte empírico de la crítica de Wang, Thomas y sus colaboradores sobre la efectividad del método de palabra clave en la retención a largo plazo de vocabulario en L2, se decidió analizarlo desde la perspectiva del desempeño del método de palabra clave, específicamente en los tests tardíos. Después de todo, desde el punto estrictamente práctico, lo que interesa es conocer cuál de los métodos garantiza mayor retención a largo plazo de vocabulario.

Desde esta perspectiva, el análisis permitió examinar la novedosa noción de la significación práctica (tamaño de los efectos, *effect sizes*, *ESs*) en lugar de la tradicional significación estadística. Aunque existen varios índices de *ESs* (e.g., *d*, *r*, η^2 , R^2), en la presente investigación secundaria se empleó el índice *d* de Cohen (1988) porque en los estudios primarios se comparó la efectividad del método de

palabra clave versus la del método de repetición o la del método de contexto. El valor de d refleja la diferencia de los promedios (medias, means, M_s) entre dos grupos en unidades de desviación estándar (DS). Sin embargo, los criterios generalizados de Cohen para categorizar los valores d como pequeño ($d = 0,2$), mediano ($d = 0,5$) y grande ($d = 0,8$) fueron reemplazados por los reportados en Plonsky y Oswald (2014), específicamente para las investigaciones en L2. Luego de examinar 346 estudios primarios y 91 meta-análisis en L2, estos investigadores hallaron evidencia empírica que indicó que la escala de Cohen subestima el rango de los efectos que generalmente se obtienen en las investigaciones de L2. Para las diferencias de las medias entre grupos, obtuvieron los siguientes criterios: pequeño ($d = 0,40$), mediano ($d = 0,70$) y grande ($d = 1,00$).

Para calcular los valores d , se aplicaron las tres fórmulas sugeridas en Norris y Ortega (2000, pp. 442-444).

$$d = \frac{M_e - M_c}{S_w} \quad (1)$$

Donde M_e es la media del grupo experimental o el grupo que recibe el tratamiento, M_c es la media del grupo control o el grupo a comparar y S_w es la desviación estándar combinada de los grupos experimental y control. La S_w se calcula aplicando la fórmula siguiente:

$$S_w = \frac{(n_1 - 1) DS_1 + (n_2 - 1) DS_2}{(n_1 - 1) + (n_2 - 1)} \quad (1.1)$$

Donde n es el tamaño de la muestra de cualquiera de los grupos mientras que DS es la desviación estándar de los respectivos grupos. Como se puede apreciar, aplicar estas fórmulas para calcular los valores d requiere únicamente los estadísticos descriptivos más básicos (e.g., n , M , DS).

Sin embargo, en algunos estudios primarios no se reportaron todos los mencionados estadísticos descriptivos básicos sino los valores exactos de F o t en cuyo caso, hubo que utilizar una de las dos fórmulas siguientes:

$$d = \frac{(\sqrt{F})(n_1+n_2)}{(\sqrt{df_{error}})(\sqrt{n_1n_2})} \quad (2)$$

Donde F es el valor reportado en un análisis de varianza (ANOVA), n es el tamaño de la muestra de los respectivos grupos, df_{error} son los grados de libertad de término error y t es el valor reportado en una prueba t .

$$d = \frac{t(n_1+n_2)}{(\sqrt{df_{error}})(\sqrt{n_1n_2})} \quad (3)$$

Al no reportar los valores d en sus comparaciones, Wang, Thomas y sus colaboradores obligaron a utilizar alguna de las mencionadas fórmulas para calcularlos. Sin embargo, la tarea se complicó, especialmente en los estudios iniciales (i.e., Wang y otros, 1992, Exps. 1-2; Wang y otros, 1993) debido a que tampoco reportaron los estadísticos descriptivos básicos necesarios para aplicar la Fórmula 1. En esos casos, se recurrió a las gráficas presentadas y a otro tipo de información para emplear dicha fórmula y calcular los valores *aproximados* de d . Afortunadamente, en algunos estudios (Wang y otros, 1992, Exps. 3; Wang y Thomas, 1995, Exp. 2) incluyeron los valores exactos de F o t y sus correspondientes grados de libertad (degrees of freedom, df) e información suficiente para aplicar las Fórmulas 2 y 3, respectivamente; en los estudios posteriores (Thomas y Wang, 1996, Exps. 1-2) incluyeron los mencionados estadísticos descriptivos básicos para estimar los valores d .

Con respecto a los hallazgos de las réplicas de Carney y Levin (1998), la situación fue mucho más sencilla porque estos investigadores reportaron todos los

estadísticos descriptivos básicos para calcular los valores d . De hecho, reportaron estos valores para muchos de los contrastes examinados, aplicando la Fórmula 1. Específicamente, señalaron: “El tamaño del efecto (ES) es la diferencia de las medias dividida por la división estándar combinada” (p. 280).

Finalmente, se calcularon los valores d para examinar la predicción de Wang, Thomas y sus colegas sobre el desempeño del método de palabra clave en diseños intra-participantes. Este tipo de diseño lo usaron en el Experimento 4 de Wang y otros (1992) y en parte del diseño mixto de Wang y Thomas (1995, Exp. 3). Asimismo, se computaron para la parte intra-participantes de los estudios de Carney y Levin (1998) a los fines de evaluar la consistencia de estos hallazgos con los estimados para los mencionados dos estudios de Wang, Thomas y sus colaboradores. Con la excepción de Wang y otros (1992, Exp. 4) en el que se reportó el valor F y su correspondiente df para el test tardío y se aplicó la Fórmula 2 para calcular el valor de d , en todos los demás experimentos se reportaron los estadísticos descriptivos básicos (e.g., n , M , SD) que permitieron emplear la Fórmula 1 para obtener los índices d .

CAPITULO IV

RESULTADOS

Soporte Empírico de la Crítica de Wang, Thomas y sus Colaboradores

Wang, Thomas y sus colaboradores discutieron la noción de “ventaja a largo plazo” desde dos perspectivas diferentes: la diferencia en el desempeño de los métodos comparados y el porcentaje de olvido exhibido por dichos métodos. Se inclinaron por esta última para examinar la efectividad del método de palabra clave en la retención a largo plazo de vocabulario en L2. Asumiendo esta perspectiva, criticaron el uso de diseños intra-participantes en todos los estudios previos de retención a largo plazo que involucraban al método de palabra clave porque en este tipo de diseño, la marcada superioridad del método mnemónico en la retención inmediata infla su verdadero desempeño en el test administrado días más tarde. En su opinión, un diseño inter-participantes combinado con un paradigma de aprendizaje incidental producirían mediciones reales del porcentaje de olvido. En este contexto, el método de palabra clave exhibiría un mayor porcentaje de olvido que el método control, indicado por el efecto estadísticamente significativo de la interacción entre el método de aprendizaje y el intervalo de retención.

En virtud de que los diseños inter-participantes representan el basamento sustantivo de la crítica de este grupo de investigadores, se les dio una cobertura mucho más completa y detallada que la dada a los diseños intra-participantes, elementos complementarios de dicha crítica.

Diseños Inter-participantes

Perspectiva del Porcentaje de Olvido. La Tabla 1 presenta el número de experimentos en el que Wang, Thomas y sus colaboradores manipularon la condición de aprendizaje y el intervalo de retención como factores inter-participantes y en el que el efecto interactivo entre estas dos variables alcanzó

significación estadística. Como se aprecia en la mencionada tabla, estos investigadores realizaron 8 experimentos en los que implementaron un diseño inter-participantes para poner a prueba su cuestionamiento sobre la efectividad del método de palabra clave en la retención a largo plazo de vocabulario en L2. Con la excepción del estudio de Wang y otros (1993), todos los experimentos usaron un paradigma de aprendizaje incidental. Asimismo, se aprecia que el efecto de la interacción entre la condición de aprendizaje y el intervalo de retención fue estadísticamente significativo en 7 de los 8 experimentos registrados, corroborando las predicciones de Wang, Thomas y sus colegas.

Tabla 1. Exps. con diseños inter-participantes e interacción estadísticamente significativa		
Autores y año de publicación	Experimento	Sig. Estadística
	1	✓
Wang, Thomas y Ouellette (1992)	2	✓
	3	✓
Wang, Thomas, Inzana y Primicerio (1993) ^a		✓
Wang y Thomas (1995) ^b	2	✓
	3	-
Thomas y Wang (1996)	1	✓
	2	✓

Nota. ^a En este artículo se reportó un solo estudio y fue bajo un paradigma de aprendizaje intencional; ^b en el Exp. 3, el diseño fue mixto. Aunque la condición de aprendizaje y el intervalo de retención se manipularon como factores inter-participantes, la frecuencia de presentación de las palabras experimentales se manipuló como factor intra-participantes (8 palabras presentadas una vez; 8 palabras presentadas 3 veces y 8 presentadas 5 veces). La interacción triple fue la única interacción que alcanzó significación estadística.

Por su parte, Carney y Levin (1998) realizaron 5 experimentos para replicar y extender los hallazgos reportados por Wang, Thomas y sus colaboradores. Utilizaron un diseño mixto en cada uno de sus estudios, es decir, incluyeron una parte intra-participantes y otra inter-participantes. La información sobre la parte inter-participantes de los diseños se presenta en la Tabla 2, a los fines de

examinar hasta qué punto dicha información es consistente con la registrada en la tabla anterior.

Tabla 2. Parte inter-participantes de los Exps. e interacción estadísticamente significativa		
Autores y año de publicación	Experimento	Sig. Estadística
Carney y Levin (1998)	1	✓
	2	-
	3	-
	4	✓
	5	✓

Como lo muestra la Tabla 2, los hallazgos reportados por Carney y Levin (1998) para la parte inter-participantes de sus experimentos corroboraron sólo parcialmente los reportados por Wang, Thomas y sus colegas ya que únicamente en 3 de los 5 experimentos, el efecto de la interacción entre la condición de aprendizaje y el intervalo de retención fue estadísticamente significativo, indicando que el porcentaje de olvido del método de palabra clave fue significativamente mayor al exhibido por el método de repetición para los ítems no examinados previamente. En los restantes dos experimentos, específicamente en los Experimentos 2 y 3, la diferencia en el porcentaje de olvido entre los métodos no alcanzó significación estadística. Sin embargo, en los 5 experimentos, el promedio del porcentaje de olvido del método de palabra clave superó el del método de repetición.

Perspectiva de la Diferencia en el Desempeño. Como se señaló en la sección inicial del Capítulo II, esta perspectiva se enfoca exclusivamente en las diferencias absolutas del desempeño entre los métodos mientras deja de lado lo relativo al porcentaje de olvido. Desde esta perspectiva, se toma en consideración el desempeño tanto en el test inmediato de recuerdo inducido como en el test administrado días más tarde. Dada la marcada superioridad del método de palabra clave en la retención inmediata, reconocida y confirmada por Wang, Thomas y sus

colaboradores y corroborada por Carney y Levin (1998), pareciera que la aplicación de esta perspectiva tal y como se plantea estaría sesgada a favor del método de palabra clave. Ahora bien, desde el punto de vista estrictamente práctico, interesaría más el desempeño en la retención a largo plazo que el desempeño en la retención inmediata. En consecuencia, se decidió examinar el desempeño del método de palabra clave relativo al de diversos métodos control (e.g., repetición, contexto) en la retención a largo plazo de vocabulario en L2. Esta decisión ofreció la oportunidad de reemplazar criterios estadísticos tradicionales (e.g., test de significación estadística, valores p) por otros más novedosos y menos vulnerables (e.g., d de Cohen). El cálculo de los valores d no tuvo exento de dificultades; por ejemplo, dado que, en los hallazgos reportados por Wang, Thomas y sus colaboradores para algunos de sus experimentos (e.g., Wang y otros, 1992, Exps. 1-2) no se incluyó suficiente información (e.g., n , M , DS) para computar los valores exactos del tamaño de los efectos, los que se presentan en la Tabla 3 son aproximaciones de dichos valores.

La Tabla 3 indica que los tamaños de los efectos estimados a partir de los hallazgos reportados por Wang, Thomas y sus colaboradores mostraron una marcada inconsistencia. Por una parte, el método de palabra clave resultó ser más efectivo en la retención a largo plazo de vocabulario en L2 que el método control en 3 de los 8 experimentos registrados (valores d pequeños en Wang y otros, 1992, Exps. 1-2, test tardío: 1 semana después; y mediano en Wang y Thomas (1995, Exp.3, test tardío: 2 días). Por otra parte, el método de palabra clave fue mucho menos efectivo que el método control en los restantes 5 estudios (los valores d oscilaron entre -0,67 y -2,14).

Tabla 3. Valores d del método de palabra clave ^a en la retención a largo plazo (Exps. con diseños inter-participantes de Wang, Thomas y sus colaboradores)		
Autores y año de publicación	Experimento	d
	1	0,28
Wang, Thomas y Ouellette (1992)	2	0,56
	3	-1,58
Wang, Thomas, Inzana y Primicerio (1993)		-0.90
Wang y Thomas (1995)	2	-1,92
	3	0,81
Thomas y Wang (1996) ^b	1	-2,14
	2	-0,67

Nota. El signo (-) delante del valor d indica que el desempeño del método de palabra clave fue inferior al del método de control; ^a con excepción de los experimentos de Thomas y Wang (1996), el método de palabra clave consistió en la versión estándar del método (palabras claves suministradas por el investigador e imágenes interactivas generadas por los estudiantes); ^b el valor d estimado fue el combinado de las dos variantes utilizadas: la estándar y la variante en la que ambos enlaces fueron generados por los aprendices en el Experimento 1; y en el Experimento 2, la clásica variante estándar y otra variante estándar que además incluyó los dibujos de las palabras claves y los de sus correspondientes equivalentes en inglés.

Por su parte, los tamaños de los efectos del método de palabra clave en los estudios de Carney y Levin (1998), presentados en la Tabla 4, muestran una gran consistencia (los valores d todos positivos), indicando un mejor desempeño del método de palabra clave que el del método de repetición en los cinco experimentos reportados, particularmente en aquellos en los que los aprendices recibieron además las imágenes interactivas (e.g., Exps. 3-4).

Tabla 4. Valores d del método de palabra clave^a en la retención a largo plazo (parte inter-participantes) en los experimentos de Carney y Levin

Autores y año de publicación	Experimento	d
Carney y Levin (1998)	1	0,36
	2 ^b	0,37
	3 ^c	0,98
	4 ^d	1,11
	5 ^e	0,38

Nota. ^a Con excepción del Experimento 1 en el que el método de palabra clave fue representado por la versión estándar del método, en los restantes experimentos se usaron dos variantes; ^b El valor d computado resultó de la combinación de los valores d de las variantes estándar y práctica de los caminos al recuerdo; ^c El valor d computado de la combinación de los valores d de las variantes palabra clave + interacción (palabras claves e imágenes interactivas dadas) y palabra clave/sonido + interacción; ^d variantes: estándar y palabra clave + interacción; ítems nuevos examinados 2 días después e ítems nuevos examinados 5 días después; ^e variantes: estándar intencional y estándar incidental.

Ahora bien, dado que la variante estándar del método de palabra clave estuvo presente en todos los experimentos de Wang, Thomas y sus colaboradores en los que compararon la efectividad de los métodos de palabra clave y de repetición (Thomas y Wang, 1996; Wang y otros, 1992, 1993) y en 4 de los 5 experimentos de Carney y Levin (1998), pareció interesante calcular los valores d de dicha variante en los mencionados estudios para determinar con mayor precisión su desempeño relativo en la retención a largo plazo de vocabulario en L2. Los valores d computados se presentan en la Tabla 5.

El patrón de resultados que se muestra en dicha tabla es prácticamente idéntico al mostrado en las Tablas 3 y 4. Los valores d estimados de los hallazgos reportados por Wang, Thomas y sus colegas se caracterizaron por una marcada inconsistencia; por ejemplo, en Wang y otros (1992, Exps. 1-2) se obtuvieron valores positivos, aunque modestos, indicando que el desempeño de la versión estándar del método de palabra clave fue ligeramente superior al del método de repetición. Sin embargo, en los restantes experimentos de este grupo de

investigadores, se obtuvieron valores negativos muy altos, superando incluso las 2 desviaciones típicas (Thomas y Wang, 1996, Exp. 1), indicando que el desempeño del método de repetición fue significativamente superior al de la variante estándar del método de palabra clave, a pesar de que el test de recuerdo inducido tardío fue administrado tan solo 2 días después del test inmediato.

Tabla 5. Valores d de la versión estándar del método de palabra clave en los Exps. de Wang, Thomas y sus colegas y en la parte inter-participantes de los de Carney y Levin		
Autores y año de publicación	Experimento	d
	1	0,28
Wang, Thomas y Ouellette (1992)	2	0,56
	3	-1,58
Wang, Thomas, Inzana y Primicerio (1993) ^a		-0,90
Thomas y Wang (1996) ^b	1	-2,33
	2	-0,94
	1	0,36
	2	0,45
Carney y Levin (1998) ^c	4	1,07
	5	0,60

Nota. El signo (-) delante del valor d indica que el desempeño de la versión estándar del método de palabra clave fue inferior al de repetición; su ausencia significa lo contrario. ^a En este estudio se usó un paradigma de aprendizaje intencional. Sin embargo, al igual que en Wang y otros (1992, Exp. 3), se manipuló la sesión de aprendizaje, favoreciendo al grupo de repetición (10 exposiciones a razón de 3 segundos vs. 2 exposiciones a una rata de 10 segundos) para que su desempeño en el test inmediato fuera estadísticamente comparable al grupo de palabra clave. ^b Además de la variante estándar se usó una variante distinta en cada uno de estos experimentos (ver Tabla 3, Nota ^b) ; ^c Con excepción del Experimento 1 donde se utilizó únicamente la versión estándar, en los restantes experimentos registrados en esta tabla, además de dicha variante se usó una variante distinta (ver Tabla 4, Notas ^{b,d,e}).

Por su parte, los valores d computados para la versión estándar del método de palabra clave en los tests tardíos de los experimentos de Carney y Levin (1998)

fueron todos positivos, indicando que la variante estándar fue más efectiva que el método de repetición. No obstante, los valores d obtenidos fueron pequeños en su mayoría; las excepciones se produjeron bajo un paradigma de aprendizaje intencional (Exp. 5) y cuando se examinaron únicamente 8 ítems que no habían sido incluidos en los tests previos (Exp. 4). En estos casos, se obtuvieron valores d mediano y grande, respectivamente.

Diseños Intra-participantes

Comoquiera que la crítica de Wang, Thomas y sus colaboradores sobre la efectividad del método de palabra clave en la retención a largo plazo de vocabulario en L2 se fundamenta en la utilización de diseños intra-participantes que, a su juicio, sobreestiman el desempeño del método, se decidió examinar su hipótesis de que, en este tipo de diseño, la ventaja inicial del método en el test de retención inmediata se mantiene después de unos días. Por lo tanto, el análisis se enfocó en el desempeño relativo del método en los tests de retención tardía.

De los 9 experimentos desarrollados por este grupo de investigadores que fueron codificados como de retención a largo plazo de vocabulario en L2, sólo en uno (Wang y otros, 1992, Exp. 4) se usó un diseño intra-participantes; en otro (Wang y Thomas, 1995, Exp. 3) se utilizó un diseño mixto, en la que una de las 3 variables: frecuencia de presentación de las palabras experimentales (1 vez, 3 veces, 5 veces) fue manipulada como factor intra-participantes. Por su parte, y como se mencionó antes, Carney y Levin (1998) implementaron un diseño mixto (una parte inter-participantes y otra intra-participantes) en sus 5 experimentos. De allí que se calcularan los valores d del método de palabra clave en la retención a largo plazo en la parte intra-participantes de los diseños de estos experimentos. Dado el pequeño número de estudios donde se implementó este tipo de diseño, se optó por presentar sus valores d en una sola tabla, la Tabla 6.

Tabla 6. Valores *d* del método de palabra clave en la parte intra-participantes de los diseños en los Exps. de Wang, Thomas y sus colegas^a y los de Carney y Levin^b

Autores y año de publicación	Experimento	<i>d</i>
Wang, Thomas y Ouellette (1992)	4	1,05
Wang y Thomas (1995)	3	0,93
	1	0,12
	2	0,40
Carney y Levin (1998)	3	1,03
	4	1,26
	5 ^c	0,64

Nota. ^a En estos 2 experimentos, el método de palabra clave estuvo representado por la variante estándar del método. ^b Con excepción del Exp. 1 en el que la versión estándar representó al método de palabra clave, en los restantes experimentos se usaron dos variantes de método de palabra clave (ver Tabla 4, Notas ^{b-e}); ^c El valor *d* computado para la variante estándar bajo el paradigma de aprendizaje intencional fue 0,97.

Como se observa en la mencionada tabla, los valores *d* estimados para el método de palabra clave en los tests de retención tardía de los experimentos de Wang, Thomas y sus colaboradores indicaron que el método de palabra clave mantuvo una apreciable ventaja de casi una desviación estándar sobre el método de repetición luego de algunos días. Este hallazgo corroboró su hipótesis. Con respecto a los valores *d* computados para los hallazgos reportados por Carney y Levin (1998), la tabla indica que a excepción del obtenido para el Experimento 1, los demás fueron consistentes con los estimados en los experimentos de Wang, Thomas y sus colegas, especialmente cuando se utilizó la variante palabra clave + interacción (Exps. 3-4).

CAPITULO V

DISCUSION Y CONCLUSIONES

El propósito de la presente investigación fue examinar el soporte empírico de la crítica de Wang, Thomas y sus colaboradores sobre la efectividad del método de palabra clave en la retención a largo plazo de vocabulario en L2 y su consistencia con los hallazgos reportados en las réplicas de Carney y Levin (1998). Tres interrogantes guiaron la presente investigación: La primera interrogante examinó hasta qué punto la prueba de significación estadística de la interacción entre la condición de aprendizaje y el intervalo de retención corroboró su crítica. La segunda interrogante examinó hasta qué punto los tamaños de los efectos del método de palabra clave en los tests tardíos, estimados a partir de los hallazgos reportados por Wang, Thomas y sus colaboradores, sustentaron su crítica. Finalmente, la tercera interrogante examinó hasta qué punto los hallazgos reportados en las réplicas de Carney y Levin (1998) fueron consistentes con los reportados en los estudios de Wang, Thomas y sus colaboradores sobre la efectividad del método de palabra clave en la retención a largo plazo de vocabulario en L2. Dada la estrecha relación entre la tercera interrogante y las dos anteriores, se examinó conjuntamente con ellas.

Con respecto a la primera interrogante, los resultados de la presente investigación corroboraron la hipótesis planteada por Wang, Thomas y sus colegas según la cual en diseños inter-participantes, la marcada superioridad del método de palabra clave en los tests inmediato de recuerdo inducido desaparece o incluso se revierte después de pocos días. En otras palabras, el método de palabra clave exhibió un porcentaje de olvido significativamente superior al de una variedad de grupos control (e.g., repetición, contexto). Este hecho se reflejó en el efecto estadísticamente significativo de la interacción entre la condición de aprendizaje (i.e., métodos comparados) y el intervalo de retención en siete de los ocho estudios primarios (ver Tabla 1). Además, en todos sus estudios, incluso en el único donde dicho efecto interactivo no fue estadísticamente significativo (Wang

y Thomas, 1995, Exp. 3), el porcentaje de olvido del método de palabra clave superó al del grupo control (i.e., contexto). Sus hallazgos fueron, en buena medida, avalados por los reportados en Carney y Levin para la parte inter-participantes de sus experimentos. Aunque el efecto de la interacción entre la condición de aprendizaje y el intervalo de retención fue estadísticamente significativo únicamente en tres de los cinco experimentos, el método de palabra clave exhibió porcentajes de olvido mayores a los del método de repetición también en esos dos experimentos (alrededor de 9% y 8% en los Experimentos 2 y 3, respectivamente). Por lo tanto, desde su perspectiva (i.e., porcentaje de olvido) resulta difícil que su crítica sea cuestionada.

La segunda interrogante abordó el soporte empírico de su crítica desde una perspectiva estrictamente práctica. Desde el punto de vista pedagógico, lo que interesa a docentes, estudiantes y encargados de las políticas educativas, es saber cuál de las estrategias de aprendizaje de vocabulario de L2s garantiza mayor retención a largo plazo. En este contexto, se examinó el desempeño del método de palabra clave en el recuerdo tardío de vocabulario en L2, tomándose como elemento de comparación el tamaño de los efectos, específicamente los valores d de Cohen (1988). Dado que Wang, Thomas y sus colaboradores no les interesó este planteamiento, no reportaron directamente información en sus análisis sobre dicho aspecto. De allí que en la presente investigación se utilizaran diversos tipos de información (e.g., gráficas, valores F o t y estadísticos descriptivos básicos) para calcular esos valores. El análisis de los valores d produjo un patrón de resultados más heterogéneo que el generado desde la perspectiva anterior. Como se indicó en la Tabla 3, mientras el método de palabra clave aventajó al método control en el test tardío en tres de sus experimentos (Wang y otros, 1992, Exps. 1-2; Wang y Thomas, 1995, Exp. 3), promediando un índice $d = 0,55$, un efecto intermedio entre pequeño y mediano (Plonsky y Oswald, 2014), en los restantes cinco, su ventaja inicial se revirtió, incluso abruptamente en algunos casos (Thomas y Wang, 1996, Exp. 1; Wang y otros, 1992, Exp. 3; Wang y Thomas, 1995, Exp. 2), luego de pocos días, llegando a promediar en esos cinco estudios un efecto grande ($d = -1,44$). En resumen, el desempeño del método de

palabra clave en el test tardío fue, en promedio, 0,70 unidades de desviación estándar inferior al de los métodos comparativos.

Por el contrario, los valores d computados para los hallazgos de la parte inter-participantes de los diseños implementados en los experimentos de Carney y Levin (1998) fueron todos positivos, oscilando entre índices pequeños ($d = 0,36$) y grandes ($d = 1,11$), con un valor d promedio de 0,55. En otras palabras, el método de palabra clave, en promedio, aventajó al método de repetición por más de media unidad de desviación estándar en los tests tardíos, lo cual evidencia la superioridad del método de palabra clave en la retención a largo plazo de vocabulario en L2. Asimismo, estos investigadores reconocieron que cuando la métrica utilizada es el número de ítems correctos (Wang, Thomas y sus colegas) o la del porcentaje de ítems correctos, como fue su caso, la ventaja del método de palabra clave es generalmente menor en un test tardío que en un test inmediato; sin embargo, al utilizar otras métricas como, por ejemplo, la diferencia relativa del método de palabra clave sobre el de repetición, reportaron que ésta se incrementó en el test tardío (p. 294). En su opinión, ambas métricas están sesgadas. Mientras la métrica absoluta está sesgada en favor de reducir los beneficios mnemónicos en el tiempo, la diferencia relativa está sesgada en la dirección opuesta, es decir, a incrementar los beneficios mnemónicos en el tiempo (p. 295).

Con respecto a los hallazgos reportados para los diseños intra-participantes, los resultados de la presente investigación parecieran corroborar el señalamiento de Wang, Thomas y sus colaboradores de que, en este tipo de diseño, el método de palabra clave mantiene una apreciable ventaja sobre los métodos comparativos en los tests tardíos. Los valores d calculados para los hallazgos reportados por este grupo de investigadores en sus dos experimentos fueron similares ($d = 1,05$ y $d = 0,93$ en Wang y otros, 1992, Exp. 4 y Wang y Thomas, 1995, Exp. 3, respectivamente). En otras palabras, el método de palabra clave mantuvo, en promedio, una ventaja sobre los métodos no mnemónicos de casi una unidad de desviación estándar ($d = 0,99$), es decir, un efecto grande. Por su parte, los valores d para la parte intra-participantes de las réplicas de Carney y

Levin (1998), con excepción del computado para el Experimento 1 que contradice el señalamiento de Wang, Thomas y sus colegas, son consistentes con el mencionado señalamiento, aunque con un valor d global relativamente menor ($d = 0,69$).

Ahora bien, en la mayoría de los experimentos examinados en la presente investigación, se observó un apreciable decrecimiento en el desempeño del método de palabra clave en los tests tardíos; más pronunciado en los estudios con diseños inter-participantes que en los que implementaron diseños intra-participantes. La explicación a ese manifiesto decrecimiento se ha buscado en la naturaleza del método. Como se sabe, el método de palabra clave consiste en la activación de dos enlaces: el enlace acústico, representado por la palabra clave (palabra en L1), y el enlace mental, representado por la imagen interactiva entre el referente de la palabra clave y el del equivalente en L1. En seis de los siete experimentos con diseños inter-participantes realizados por Wang, Thomas y sus colaboradores (Thomas y Wang, 1996, Exp. 2 fue la excepción), los participantes asignados al método de palabra clave, además de los equivalentes de las palabras experimentales, debieron escribir las palabras clave en los tests de recuerdo inducido. Los resultados reportados en los tests tardíos de los mencionados estudios indicaron que los estudiantes que usaron el método mnemónico fueron capaces de recordar un poco más del 80% de las palabras clave. Un porcentaje similar fue reportado en el Experimento 1 de Carney y Levin (1998), el único de sus estudios donde se le pidió al grupo de la estrategia mnemónica escribir las palabras clave; por lo tanto, el pobre desempeño del método mnemónico en la retención a largo plazo de vocabulario en L2 no podría atribuírsele al olvido de las palabras clave.

Descartado el olvido de las palabras clave como factor causante de este desempeño relativamente pobre del método de palabra clave, Wang, Thomas y sus colaboradores dirigieron su atención hacia el enlace mental, es decir, las imágenes interactivas. Una de sus explicaciones más atractivas involucró las asociaciones pre-experimentales de las palabras clave y sus probabilidades de

interferir en el recuerdo de las imágenes interactivas presentadas en la fase de estudio. Específicamente señalaron:

Aunque la palabra clave sirve inicialmente como una ayuda efectiva para recordar, las asociaciones pre-experimentales pueden retomar su importancia con el tiempo y obstaculizar el recuerdo de la imagen mnemónica. Por ejemplo, durante el aprendizaje de la palabra francesa *eglise* (iglesia), la palabra clave *egg* pudo haber sido interpretada como un huevo de gallina ovalado blanco. Sin embargo, una semana más tarde, el recuerdo de la palabra clave *egg* pudo haber generado una interpretación distinta (e.g., huevos revueltos) que no estuvo incluida en la imagen interactiva original (Wang, Thomas y Ouellette, 1992, p. 527).

Posteriormente, Wang y Thomas (1999), en su respuesta a los comentarios de Gruneberg (1998), reafirmaron dicha explicación, usando como ejemplo, la palabra tagalog *salamin* (eyeglasses, lentes):

Nuestra explicación (corroborada por los hallazgos reportados en Thomas y Wang, 1996) es que las asociaciones originales de la palabra clave son frágiles porque las representaciones alternativas de la palabra clave pueden evocarse, particularmente después de un tiempo. Por ejemplo, después de un tiempo, un estudiante puede interpretar la palabra clave *salmon* como un filete anaranjado de pescado y así ser poco probable que localice una representación almacenada de “un pescado usando lentes” (p. 284).

Adicionalmente, se examinaron otros aspectos de interés estrechamente relacionados con el tema central abordado en el presente estudio, incluyendo el efecto de la práctica en la retención a largo plazo del método de palabra clave, la visión práctica de Wang, Thomas y sus colegas sobre sus hallazgos y el debate que protagonizaron Gruneberg (1998) y Wang y Thomas (1999).

La práctica, entendida como las oportunidades para estudiar las palabras experimentales o el número de veces (frecuencia) que éstas son presentadas en la fase de estudio, fue examinada únicamente en dos de los estudios primarios incluidos en la presente investigación: Wang y Thomas (1995, Exp. 3) y Carney y Levin (1998, Exp. 4). A pesar de las notables diferencias metodológicas entre los mencionados estudios (e.g., 8 palabras presentadas una vez, 8 presentadas 3

veces y 8 presentadas 5 veces vs. 8 palabras examinadas una vez, 8 examinadas 2 veces y 8 examinadas 3 veces), compartieron la interpretación de los hallazgos relativos a la frecuencia. En ambos casos, el porcentaje de olvido del grupo mnemónico fue estadísticamente comparable con el del grupo no mnemónico (contexto y repetición, respectivamente) cuando se les permitió practicar las palabras experimentales tres y cinco veces o cuando éstas fueron examinadas dos y tres veces.

De lo anterior, podría especularse teóricamente que suficiente práctica tendría efectos similares en los métodos de aprendizaje, generando en ellos porcentajes de olvido estadísticamente comparables. Este hallazgo se reflejaría en que el efecto entre la condición de aprendizaje y el intervalo de retención no alcanzaría significación estadística, contradiciendo así la crítica de Wang, Thomas y sus colaboradores sobre la efectividad del método de palabra clave en la retención a largo plazo de vocabulario en L2. Este planteamiento se pondría a prueba de la siguiente manera: los ítems experimentales serían presentados varias veces (e.g., seis veces) y de manera *espaciada* en el tiempo (e.g., dos veces un día, dos veces tres días después y dos veces una semana más tarde), fortaleciendo la asociación del referente de la palabra clave con el del significado de la palabra por aprender (i.e., la imagen interactiva). Dicho fortalecimiento incrementaría la probabilidad de acceder a la imagen interactiva almacenada. El mencionado planteamiento debería examinarse en futuras investigaciones.

Con respecto a la visión práctica sobre sus hallazgos, Wang y Thomas (1995, p. 474) reconocieron que mientras el uso de diseños inter-participantes corre el riesgo de no reflejar el paradigma de aprendizaje establecido en la típica aula de clase, el uso de diseños intra-participantes producen aprendizajes que son más propensos a reflejar los contextos educativos. Asimismo, reconocieron que aunque el paradigma de aprendizaje incidental es apropiado para investigaciones empíricas de olvido a largo plazo, puede no aplicar a la típica práctica educativa. De acuerdo con esta práctica, los estudiantes que hacen su esfuerzo para aprender un nuevo vocabulario generalmente lo hacen con la expectativa de que

ese conocimiento será evaluado. El paradigma de aprendizaje intencional, por lo tanto, promueve situaciones en las que los estudiantes son expuestos nuevamente al material de estudio a través de técnicas como la auto-evaluación y la práctica espontánea (ver debate para complementar esta visión práctica).

Finalmente, se examinó el debate sostenido entre Gruneberg (1998) y Wang y Thomas (1999), publicado en la revista *Applied Cognitive Psychology*. Aunque Gruneberg no objeta la data experimental reportada por Wang, Thomas y sus colaboradores, cuestiona su conclusión de que proporcionar a los estudiantes imágenes derivadas mnemónicamente no produce ventajas en la retención a largo plazo (p. 530). En su opinión, existe considerable evidencia que indica que la instrucción mnemónica facilita la retención cuando a los aprendices se les examina inmediatamente y en una fecha posterior. *Este caso corresponde a un diseño intra-participantes* (agregado de la autora de la presente investigación). En ese sentido, criticó el que estos investigadores no discutieran las implicaciones prácticas de los efectos beneficiosos de los estudios intra-participantes y describió una serie de implicaciones prácticas, algunas de las cuales se presentan a continuación:

Desde un punto de vista práctico, no importa que el desempeño del método de palabra clave después de un tiempo pueda ser inferior al de repetición en ausencia de un test inmediato si puede demostrarse que es superior cuando éste se ha permitido (p. 530).

En términos prácticos, parece pedagógicamente perverso negar un test de recuerdo inmediato después de haberse presentado el material cuando se sabe que esta manipulación claramente promueve una retención mayor a la de la repetición tanto inmediatamente como después de un tiempo (p. 530).

Desde el punto de vista del aprendiz, poco importa el por qué una manipulación es efectiva y con seguridad no importa que el mecanismo subyacente involucre una interacción del método de palabra clave con el test inmediato. Lo que muestra la evidencia de Wang y otros (1992) es que el recuerdo inmediato es necesario para asegurar la consolidación de las imágenes interactivas, y que esto produce mayor retención a largo plazo que la interacción del método de repetición con el test inmediato (pp. 530-531).

Además, cuestionó la validez ecológica de los estudios de Wang, Thomas y sus colegas y resaltó el beneficio actitudinal y motivacional del método de palabra clave. El cuestionamiento a la validez ecológica la sustentó en la realización de dichos experimentos en laboratorios y no en aulas reales de L2s. En ese sentido, presentó evidencias de un número importante de estudios de campo, liderados por él y sus asociados, que demuestran la efectividad del método de palabra clave en la retención a largo plazo de vocabulario en L2s. Con respecto a los beneficios actitudinales y motivacionales del método de palabra clave, citó un artículo suyo en el que se sugiere que los usuarios consideran al método de palabra clave mucho más agradable y motivante que el método de repetición.

Por su parte, Wang y Thomas (1999), sustentaron su respuesta a los comentarios de Gruneberg (1998) fundamentalmente en la validez interna de sus experimentos. Destacaron su relevancia y su interrelación con la validez externa o ecológica. A su juicio, sus hallazgos sobre la interacción entre la efectividad del método de palabra clave y el intervalo de retención revelaron información válida e importante sobre la relación causal subyacente entre el método de estudio y la retención a largo plazo (p. 284). Asimismo, señalaron que a partir de sus estudios de laboratorio, descubrieron una limitación importante en el aprendizaje basado en imágenes: “a largo plazo, las imágenes pueden ser extremadamente difíciles de recordar a menos que hayan sido bien practicadas inicialmente” (p. 284).

Creen firmemente que, precisamente debido a validez interna de sus experimentos, sus hallazgos sobre la aplicación del método de palabra clave en el mundo real son relevantes. Afirmaron que aunque pudieran estar de acuerdo con Gruneberg en el sentido de que a los estudiantes pudiera no importarles el por qué un método funciona, es imperativo que los diseñadores de las políticas educativas comprendan los por qué (pp. 284-285). En ese mismo orden de ideas, cuestionaron tanto los estudios de campo mencionados por Gruneberg que evidencian la efectividad del método de palabra clave como los que sugieren que el método mnemónico es más agradable y motivante que otros métodos por carecer, precisamente, de validez interna. En su opinión, “es únicamente a través de estudios internamente validos que podemos descubrir las explicaciones de las relaciones observadas” (p. 285).

Finalizaron su respuesta señalando que su intención no es que se abandone la noción de que el uso del método de palabra clave produce beneficios a largo plazo. Por el contrario, sostuvieron que el mejor momento para diseñar e implementar un programa efectivo basado en el método de palabra clave es cuando los instructores sepan bajo qué condiciones hay mayores probabilidades de que este programa tenga éxito (p. 286).

Revisado el contenido de ambos manuscritos, y analizada su relación con el tema abordado en el presente estudio, dos señalamientos fueron pertinentes: el primero se refiere a la ausencia en ambos manuscritos de alguna mención de los hallazgos reportados en Carney y Levin (1998). La explicación lógica pareciera ser que los autores de los manuscritos del debate desconocieran el contenido del manuscrito de Carney y Levin y que los tres manuscritos hubiesen sido consignados para publicación en fechas muy cercanas. El segundo señalamiento tiene que ver con la visión práctica de los hallazgos de Wang, Thomas y sus colaboradores. A pesar del reconocimiento explícito en Wang y Thomas (1995) de la contradicción entre los diseños de sus estudios y los hallazgos correspondientes, por una parte y las condiciones de aprendizaje del típico salón de clase, por la otra, extraña que Gruneberg haya centrado su cuestionamiento en

esta perspectiva. Asimismo, extraña que, en su respuesta, Wang y Thomas (1999) no hiciesen mención al mencionado reconocimiento.

CONCLUSIONES

La presente investigación examinó el soporte empírico de la crítica de Wang, Thomas y sus colaboradores sobre la efectividad del método de palabra clave en la retención a largo plazo de vocabulario en L2 y su nivel de coincidencia con los hallazgos reportados en los estudios de Carney y Levin (1998). El soporte empírico fue examinado desde dos perspectivas complementarias: el porcentaje de olvido (adoptada por el grupo de Wang, Thomas y sus asociados) y el desempeño en los tests tardíos.

Desde la perspectiva del porcentaje de olvido, los hallazgos reportados por Wang, Thomas y sus colegas confirmaron su crítica y fueron parcialmente avalados por los reportados en las réplicas de Carney y Levin (1998). Sin embargo, desde la perspectiva del desempeño relativo del método mnemónico en los tests tardíos, los resultados fueron mucho menos homogéneos y más difíciles de interpretar. Los valores d calculados a partir de los hallazgos reportados por Wang, Thomas y sus colaboradores se caracterizaron por su inconsistencia en términos tanto de dirección como de magnitud; mientras en un número importante de estudios el método control aventajó al método de palabra clave, en otros, la dirección se dio en sentido contrario. Asimismo, el rango de los mencionados valores fue extremadamente amplio, oscilando entre efectos pequeños y muy grandes. Por el contrario, los valores d calculados para la parte inter-participantes de los diseños de Carney y Levin exhibieron un rango más estrecho y una marcada consistencia direccional (todos positivos), indicando que el método de palabra clave superó al método de repetición en los cinco experimentos, lo cual contradice la crítica de Wang, Thomas y sus colegas. Finalmente, los valores d calculados para los diseños intra-participantes fueron consistentes y corroboraron el señalamiento de este grupo de investigadores. Además, se examinaron aspectos secundarios, estrechamente relacionados con el tema central abordado en el presente estudio, a saber: la explicación sobre el desempeño relativamente pobre del método mnemónico en los tests tardíos, la potencialidad del efecto de la práctica en la retención a largo plazo de este método, las implicaciones prácticas

de los hallazgos reportados por Wang, Thomas y sus colegas, y de cómo algunos de estos aspectos fueron abordados en el debate entre Gruneberg (1998) y Wang y Thomas (1999).

LIMITACIONES

Aunque en el presente estudio se ofreció una visión más completa del soporte empírico de la crítica de Wang, Thomas y sus colaboradores sobre la efectividad del método de palabra clave en la retención a largo plazo de vocabulario en L2 al examinar la noción de “ventaja a largo plazo” desde dos perspectivas diferentes, se evidenciaron algunas limitaciones que deben ser abordadas en investigaciones futuras sobre el tema. Tres de ellas destacan abiertamente. En primer lugar, no se examinó hasta qué punto las comparaciones entre el método de palabra clave y los grupos control fueron metodológicamente válidas. En segundo lugar, la ausencia de estudios ecológicamente válidos; ninguno de los estudios incluidos en la presente investigación fue realizado en aulas de clases reales con aprendices de la lengua target. Finalmente, los criterios de inclusión debieron flexibilizarse para dar cabida a un mayor número de estudios sobre la retención a largo plazo de vocabulario en L2.

REFERENCIAS

- Atkinson, R. C. (1975). Mnemotechnics in second-language learning. *American Psychologist*, 30, 821-828.
- *Atkinson, R. C., y Raugh, M. R. (1975). An application of the mnemonic keyword to the acquisition of Russian vocabulary. *Journal of Experimental Psychology: Human Learning and Memory*, 104, 126-133.
- *Ávila, E., y Sadoski, M. (1996). Exploring new applications of the keyword method to acquire English vocabulary. *Language Learning*, 46, 379-395.
- Beaton, A., A., Gruneberg, M. M., y Ellis, N. C. (1995). Retention of foreign vocabulary learned using the keyword method: A ten-year follow-up. *Second Language Research*, 11, 112-120.
- *Bird, S. A., y Jacobs, G. M. (1999). An examination of the keyword method: how effective is it for native speakers of Chinese Learning English? *Asian Journal of English Language Teaching*, 9, 75-97.
- *Brown, T. S., y Perry, F. L, Jr. (1991). A comparison of three learning strategies for ESL vocabulary acquisition. *TESOL Quarterly*, 25, 655-670.
- Campos, A., Amor, A., y González, M. A. (2002). Presentation of keywords by means of interactive drawings. *The Spanish Journal of Psychology*, 5, 102-109.
- *Campos, A., Amor, A., y González, M. A. (2004). The importance of the keyword-generation method in keyword mnemonics. *Experimental Psychology*, 51, 125-131.
- *Campos, A., González, M. A., y Amor, A. (2003). Limitations of the mnemonic-keyword method. *The Journal of General Psychology*, 130, 399-413.

- *Campos, A., Pérez- Fabello, M. J., y Camino, E. (2010). Eficacia de la mnemotecnia de la palabra clave en personas adultas. *Psicotema*, 22, 752-757.
- *Colón, M. y Rodríguez, M. (2005). The effects of self-generated versus supplied interactive images on the long-term retention of EFL vocabulary by college students. *Lenguas Modernas*, 30, 191-200.
- *Carney, R. N., y Levin, J. R. (1998). Do mnemonic memories fade as time goes by? Here's looking anew! *Contemporary Educational Psychology*, 23, 276-297.
- Ellis, N. C., y Beaton, A. A. (1993a). Psycholinguistic determinants of foreign language vocabulary learning. *Language Learning*, 43, 559-617.
- *Ellis, N. C., y Beaton, A. A. (1993b). Factors affecting the learning of foreign language vocabulary: Imagery keyword mediators and phonological short-term memory. *Quarterly Journal of Experimental Psychology*, 46, 533-558.
- Gruneberg, M. M. (1998). A commentary on criticism of the keyword method of learning foreign languages. *Applied Cognitive Psychology*, 12, 529-532.
- Gruneberg, M. M., y Pascoe, K. (1996). The effectiveness of the keyword method for receptive and productive foreign vocabulary learning in the elderly. *Contemporary Educational Psychology*, 21, 102-109.
- Gruneberg, M. M., y Sykes, R. (1991). Individual differences and attitudes to the keyword method of foreign language learning. *Language Learning Journal*, 4, 60-62.
- *Hall, J. W., Wilson, K. P., y Patterson, R. (1981). Mnemotechnics: Some limitations of the mnemonic keyword method for the study of foreign language vocabulary. *Journal of Educational Psychology*, 73, 345-357.

- *Hogben, D., y Lawson, M. J. (1994). Keyword and multiple elaboration strategies for vocabulary acquisition in foreign language learning. *Contemporary Educational Psychology, 19*, 367-376.
- Johnson, C. W., Adams, M., y Bruning, R. (1985). Keyword and vocabulary acquisition: Some words of caution about words of assistance. *Educational Communication and Technology Journal, 33*, 125-138.
- Kasper, L. F. (1993). The keyword method and foreign language vocabulary learning: A rationale for its use. *Foreign Language Annals, 26*, 244-251.
- *Kasper, L. F., y Glass, A. L. (1988). An extension of the keyword method facilitates the acquisition of simple Spanish sentences. *Applied Cognitive Psychology, 2*, 137-146.
- King-Sears, M. E., Mercer, C. D., y Sindelar, P. T. (1992). Toward independence with keyword mnemonics: A strategy for science vocabulary instruction. *Remedial and Special Education, 13*, 22-33.
- *Lawson, M. J., y Hogben, D. (1998). Learning and recall of foreign language vocabulary: Effects of keyword strategy for immediate and delayed recall. *Learning and Instruction, 8*, 179-194.
- Levin, J. R. (1983). Pictorial strategies for school learning: Practical illustrations. En M. Pressley y J. R Levin (Eds.), *Cognitive strategy research: Educational applications* (pp. 213-237). New York: Springer-Verlag.
- Levin, J. R., Levin, M. E., Glasman, L. D., y Nordwall, M. B. (1992). Mnemonic vocabulary instruction: Additional effectiveness evidence. *Contemporary Educational Psychology, 17*, 156-174.
- Levin, J. R., McCormick, C. B., Miller, G. E., Berry, J. K., y Pressley, M. (1982). Mnemonic versus nonmnemonic vocabulary learning strategies for children. *American Educational Research Journal, 19*, 121-136.

- Levin, J. R., y Pressley, M. (1985). Mnemonic vocabulary instruction: What's fact, what's fiction. En R. F. Dillon y R. R. Schmeck (Eds.), *Individual differences in cognition* (pp. 145-172). New York: Academic Press.
- *Levin, J. R., Pressley, M., McCormick, C. B., Miller, G. E., y Shriberg, L. K. (1979). Assessing the classroom potential of the keyword method. *Journal of Educational Psychology, 71*, 583-594.
- Levin, J. R., Shriberg, L. K., y Berry, J. K. (1983). A concrete strategy for remembering abstract prose. *American Educational Research Journal, 20*, 277-290.
- Levin, J. R., Shriberg, L. K., Miller, G. E., McCormick, C. B., y Levin, B. (1980). The keyword method in the classroom: How to remember the states and their capitals. *The Elementary School Journal, 80*, 185-191.
- Liu, Q., y Brown, D. (2015). Methodological synthesis of research on the effectiveness of corrective feedback in L2 writing. *Journal of Second Language Writing, 30*, 66-81.
- Marsden, E., Thompson, S., y Plonsky, L. (2018). A methodological synthesis of self-paced reading tests in second language research. *Applied Psycholinguistics, 39*, 861-984.
- *Merry, R. (1980). The keyword method and children's vocabulary learning in the classroom. *British Journal of Educational Psychology, 50*, 123-136.
- Miller, G. E., Levin, J. R., y Pressley, M. (1980). An adaptation of the keyword method to children's learning of foreign verbs. *Journal of Mental Imagery, 4*, 57-61.
- *Moore, J. C., y Surber, J. R. (1992). Effects of context and keyword methods on second language vocabulary acquisition. *Contemporary Educational Psychology, 17*, 286-292.

- Norris, J. M., y Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, 50, 417-528.
- Plonsky, L. (2013). Study quality in SLA: An assessment of designs, analyses, and reporting practices in quantitative L2 research. *Studies in Second Language Acquisition*, 35, 655-687.
- Plonsky, L. (2014). Study quality in quantitative L2 research (1990-2010): A methodological synthesis and a call for reform. *Modern Language Journal*, 98, 450-470.
- Plonsky, L., y Derrick, D. J. (2016). A meta-analysis of reliability coefficients in second language research. *The Modern Language Journal*, 100, 538-553.
- Plonsky, L., y Gass, S. (2011). Study quality in interactionist research. *Language Learning*, 61, 325-366.
- Plonsky, L., y Ghanbar, H. (2018). Multiple regression in L2 research: A methodological synthesis and guide to interpreting R^2 values. *Modern Language Journal*, 102, 713-731.
- Plonsky, L., y Gonulal, T. (2015). Methodological synthesis in quantitative L2 research: A review of reviews and a case study of exploratory factor analysis. *Language Learning*, 65, 9-36.
- Plonsky, L., y Kim, Y. (2016). Task-based learner production: A substantive and methodological review. *Annual Review of Applied Linguistics*, 36, 73-97.
- Plonsky, L., Marsden, E., Crowther, D., Gass, S., y Spinner, P. (2019). A methodological synthesis and meta-analysis of judgment tasks in second language research. *Second Language Research*, 35, 1-39.
- Plonsky, L., y Oswald, F. L. (2014). How big is 'big'? Interpreting effect sizes in L2 research. *Language Learning*, 64, 878-912.

- Plonsky, L., y Oswald, F. L. (2015). Meta-analysing second language research. In L. Plonsky. (Ed.), *Advancing quantitative methods in second language research* (pp. 106-128). New York: Routledge.
- Pressley, M. (1977). Children's use of the keyword method to learn simple Spanish vocabulary words. *Journal of Educational Psychology, 69*, 465-472.
- Pressley, M., Levin, J. R., y Delaney, H. D. (1982). The mnemonic keyword method. *Review of Educational Research, 52*, 61-91.
- Pressley, M., Levin, J. R., Digdon, N., Bryant, S. L., y Ray, K. (1983). Does method of item presentation affect keyword method effectiveness? *Journal of Educational Psychology, 75*, 686-691.
- Pressley, M., Levin, J. R., Kuiper, N. A., Bryant, S. L., y Michener, S. (1982). Mnemonic versus nonmnemonic vocabulary learning strategies: Additional comparisons. *Journal of Educational Psychology, 74*, 693-707.
- Pressley, M., Levin, J. R., y McDaniel, M. A. (1987). Remembering versus inferring what a word means: Mnemonic and contextual approaches. En M. G. McKeown y M. E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 107-127). Hillsdale, NJ: Erlbaum.
- Pressley, M., Levin, J. R., y Miller, G. E. (1981a). How does the keyword method affect vocabulary comprehension and usage? *Reading Research Quarterly, 16*, 213-226.
- Pressley, M., Levin, J. R., y Miller, G. E. (1981b). The keyword method and children's learning of foreign vocabulary with abstract meanings. *Canadian Journal of Psychology, 35*, 283-287
- Pressley, M., Levin, J. R., Nakamura, G. V., Hope, D. J., Bispo, J. G., y Toye, A. R. (1980). The keyword method of foreign vocabulary learning: An investigation of its generalizability. *Journal of Applied Psychology, 65*, 635-644.

- *Raugh, M. R., y Atkinson, R. C. (1975). A mnemonic method for learning a second-language vocabulary. *Journal of Educational Psychology*, 67, 1-16.
- Raugh, M. R., Schupbach, R. D., y Atkinson, R. C. (1977). Teaching a large Russian language vocabulary by the mnemonic keyword method. *Instructional Science*, 6, 199-221.
- Rivas, J., y Rodríguez, M. (2003). The long-term effects of the context/keyword and rote methods on the receptive and productive learning of EFL vocabulary. *Lenguas Modernas*, 29, 205-211.
- *Rodríguez, M., y Sadoski, M. (2000). Effects of rote, context, keyword and context/keyword methods on retention of vocabulary in EFL classrooms. *Language Learning*, 50, 385-412.
- Rosenheck, M. B., Levin, M. E., y Levin, J. R. (1989). Learning botany concepts mnemonically: Seeing the forest and the trees. *Journal of Educational Psychology*, 81, 196-203.
- *Sagarra, N. y Alba, M. (2006). The key is in the keyword: L2 vocabulary learning methods with beginning learners of Spanish. *The Modern Language Journal*, 90, 228-243.
- Shriberg, L. K., Levin, J. R., McCormick, C. B., y Pressley, M. (1982). Learning about 'famous' people via the keyword method. *Journal of Educational Psychology*, 74, 238-247.
- Sok, S., Kang, E. Y., y Han, Z. H. (2019). Thirty-five years of ISLA on form-focused instruction: A methodological synthesis. *Language Teaching Research*, 23, 403-427.
- Sweeney, C. A., y Belleza, F. (1982). Use of the keyword mnemonic in learning English vocabulary. *Human Learning*, 1, 155-163.

- *Thomas, M. H., y Wang, A. Y. (1996). Learning by the keyword mnemonic: Looking for long-term benefits. *Journal of Experimental Psychology: Applied* 2, 330-342.
- Troutt-Ervin, E. D. (1990). Application of keyword mnemonics to learning terminology in the college classroom. *Journal of Experimental Education*, 59, 31-41.
- *Van Hell, J. G., y Candia- Mahn, A. (1997). Keyword mnemonics versus rote rehearsal: Learning concrete and abstract foreign words by experienced and inexperienced learners. *Language Learning*, 47, 507-546.
- *Wang, A. Y., y Thomas, M. H. (1995). Effect of keywords on long-term retention: Help or hindrance? *Journal of Educational Psychology*, 87, 468-475.
- *Wang, A. Y., Thomas, M. H., Inzana, C. M., y Primicerio, L. J. (1993). Long-term retention under conditions of intentional learning and the keyword mnemonic. *Bulletin of Psychonomic Society*, 31, 545-547.
- *Wang, A. Y., Thomas, M. H., y Ouellette, J. A. (1992). Keyword mnemonic and retention of second-language vocabulary words. *Journal of Educational Psychology*, 84, 520-528.
- *Wei, Z. (2015). Does teaching mnemonics for vocabulary learning make a difference? Putting the keyword method and the word part technique to the test. *Language Teaching Research*, 19, 43-69.
- *Wieczynski, D. M., y Blick, K. A. (1996). Self-referencing versus the keyword method in learning vocabulary words. *Psychological Reports*, 79, 1391-1394.
- *Willerman, B., y Melvin, B. (1979). Reservations about the keyword method. *Canadian Modern Language Review*, 35, 443-453.

HOJAS DE METADATOS

Hoja de Metadatos para Tesis y Trabajos de Ascenso – 1/6

Título	Soporte Empírico de la Crítica de Wang, Thomas y sus Colaboradores sobre la Efectividad del Método de Palabra Clave en la Retención a Largo Plazo de Vocabulario en L2
Subtítulo	

Autor(es)

Apellidos y Nombres	Código CVLAC / e-mail	
RODRIGUEZ, YARIMAX	CVLAC	14.596.925
	e-mail	yarimaxr@gmail.com
	e-mail	
	CVLAC	
	e-mail	
	e-mail	
	CVLAC	
	e-mail	
	e-mail	

Palabras o frases claves:

Método de palabra clave, vocabulario en L2, retención a largo plazo, porcentaje de olvido, diseños inter-participantes, diseños intra-participantes.
--

Hoja de Metadatos para Tesis y Trabajos de Ascenso – 2/6

Líneas y sublíneas de investigación:

Área	Sub área
HUMANIDADES Y EDUCACIÓN	INGLÉS

Resumen (abstract):

La presente investigación examinó el soporte empírico de la crítica de Wang, Thomas y sus colaboradores sobre la efectividad del método de palabra clave en la retención a largo plazo de vocabulario en L2 y su coincidencia con los hallazgos reportados en Carney y Levin (1998). La muestra la conformaron 14 estudios primarios (nueve de Wang, Thomas y sus colegas y cinco de Carney y Levin). Los hallazgos reportados en dichos estudios se examinaron desde dos perspectivas: el porcentaje de olvido y el desempeño en los tests tardíos. Desde la perspectiva del porcentaje de olvido, los resultados confirmaron el planteamiento de Wang, Thomas y sus colegas de que en diseños inter-participantes, el método de palabra clave exhibe porcentajes de olvido significativamente mayores que los de las estrategias control. Estos hallazgos fueron corroborados parcialmente por los reportados en Carney y Levin. Desde la perspectiva del desempeño en los tests tardíos, los valores *d* computados a los hallazgos reportados en Wang, Thomas y sus colaboradores fueron inconsistentes; mientras en algunos estudios el grupo de palabra clave aventajó al grupo control, en otros, sucedió lo contrario. En cambio, los valores *d* de los hallazgos reportados en Carney y Levin difirieron de los de Wang, Thomas y sus colaboradores; el método de palabra clave superó al método de repetición en todos los cinco experimentos. Con respecto a los diseños intra-participantes, los hallazgos corroboraron su señalamiento de que en ese tipo de diseño, la ventaja inicial del método de palabra clave se mantiene por unos días. Adicionalmente, se examinaron otros temas relacionados.

Hoja de Metadatos para Tesis y Trabajos de Ascenso – 3/6

Contribuidores:

Apellidos y Nombres	ROL / Código CVLAC / e-mail	
	ROL	C <input type="text"/> A <input type="text"/> S <input type="text"/> T <input type="text"/> U <input type="text"/> JU <input type="text"/>
	CVLAC	
	e-mail	
	e-mail	
	ROL	C <input type="text"/> A <input type="text"/> S <input type="text"/> T <input type="text"/> U <input type="text"/> JU <input type="text"/>
	CVLAC	
	e-mail	
	e-mail	
	ROL	C <input type="text"/> A <input type="text"/> S <input type="text"/> T <input type="text"/> U <input type="text"/> JU <input type="text"/>
	CVLAC	
	e-mail	
	e-mail	

Fecha de discusión y aprobación:

Año	Mes	Día
2022	00	00

Lenguaje: SPA

Hoja de Metadatos para Tesis y Trabajos de Ascenso – 4/6

Archivo(s):

Nombre de archivo	Tipo MIME
Trabajo de Ascenso-RodríguezY.docx	Aplication/word

Alcance:

Espacial: _____ (Opcional)

Temporal: _____ (Opcional)

Título o Grado asociado con el trabajo:

Magister Scientiarum

Nivel Asociado con el Trabajo: Magister

Área de Estudio: Humanidades

Institución(es) que garantiza(n) el Título o grado:

UNIVERSIDAD DE ORIENTE-VENEZUELA

Hoja de Metadatos para Tesis y Trabajos de Ascenso – 5/6



UNIVERSIDAD DE ORIENTE
CONSEJO UNIVERSITARIO
RECTORADO

CU N° 0975

Cumaná, 04 AGO 2009

Ciudadano
Prof. JESÚS MARTÍNEZ YÉPEZ
Vicerrector Académico
Universidad de Oriente
Su Despacho

Estimado Profesor Martínez:

Cumplo en notificarle que el Consejo Universitario, en Reunión Ordinaria celebrada en Centro de Convenciones de Cantaura, los días 28 y 29 de julio de 2009, conoció el punto de agenda **"SOLICITUD DE AUTORIZACIÓN PARA PUBLICAR TODA LA PRODUCCIÓN INTELECTUAL DE LA UNIVERSIDAD DE ORIENTE EN EL REPOSITORIO INSTITUCIONAL DE LA UDO, SEGÚN VRAC N° 696/2009"**.

Leído el oficio SIBI – 139/2009 de fecha 09-07-2009, suscrita por el Dr. Abul K. Bashirullah, Director de Bibliotecas, este Cuerpo Colegiado decidió, por unanimidad, autorizar la publicación de toda la producción intelectual de la Universidad de Oriente en el Repositorio en cuestión.



Comunicación que hago a usted a los fines consiguientes.

Cordialmente,

JUAN A. BOLAÑOS CURVELO
Secretario



C.C: Rectora, Vicerrectora Administrativa, Decanos de los Núcleos, Coordinador General de Administración, Director de Personal, Dirección de Finanzas, Dirección de Presupuesto, Contraloría Interna, Consultoría Jurídica, Director de Bibliotecas, Dirección de Publicaciones, Dirección de Computación, Coordinación de Telemática, Coordinación General de Postgrado.

JABC/YGC/manuja

Apartado Correos 094 / Telfs: 4008042 - 4008044 / 8008045 Telefax: 4008043 / Cumaná - Venezuela

Hoja de Metadatos para Tesis y Trabajos de Ascenso- 6/6

Artículo 41 del REGLAMENTO DE TRABAJO DE PREGRADO (vigente a partir del II Semestre 2009, según comunicación CU-034-2009) : “los Trabajos de Grado son de la exclusiva propiedad de la Universidad de Oriente, y sólo podrán ser utilizados para otros fines con el consentimiento del Consejo de Núcleo respectivo, quien deberá participarlo previamente al Consejo Universitario para su autorización”.



YARIMAX RODRIGUEZ
AUTORA