



A VUELTAS CON LAS INTELIGENCIAS MÚLTIPLES

Uno de los temas más polémicos, y que sigue levantando ampollas en términos de debate sobre su autenticidad o no (ya sea como teoría de la inteligencia, ya sea como marco teórico-práctico pedagógico) es el de la Teoría de las Inteligencias Múltiples del mundialmente popular Howard Gardner.

[En esta revisión exquisita](#), de nuevo, del equipo de investigadores formado por Marta Ferrero, Miguel Ángel Vadillo y Samuel Parra León la pregunta a contestar es si tenemos a día de hoy estudios de calidad suficientes sobre esta teoría y su aplicación el aula como para implementarla y apostar por ella. De entrada, el trabajo destaca ya cuatro aspectos:

- *Se realizó un metaanálisis del impacto de las intervenciones inspiradas en la Teoría de las Inteligencias Múltiples (TIM a partir de ahora) en el aprendizaje.*
- *El análisis cualitativo mostró que los estudios primarios tienen importantes fallas metodológicas.*
- *Los tamaños del efecto informados fueron notablemente mayores que los habituales en educación.*
- *No se recomiendan las intervenciones inspiradas en el TIM para mejorar el aprendizaje del rendimiento académico.*

Pero para llegar a ello, es conveniente ver los pasos que siguieron nuestros investigadores para llegar a esa última conclusión. Transcribo empezando por el resumen que abre el trabajo:

Dado que Gardner sugirió que los seres humanos tienen inteligencias múltiples, numerosos maestros han adaptado e incorporado la teoría de las inteligencias múltiples (TIM) en su rutina diaria en el aula. Sin embargo, hasta la fecha, la eficacia de las metodologías inspiradas en la TIM sigue sin estar clara. El enfoque del presente estudio fue realizar una revisión sistemática y un metaanálisis para evaluar el impacto de estas intervenciones en el rendimiento académico a través de pruebas de lectura, matemáticas o ciencias. Los criterios de inclusión para la revisión requerían que los estudios estimaran cuantitativamente el impacto de una



intervención basada en la TIM sobre el rendimiento académico y que siguieran un diseño pre-post con un grupo de control. La muestra final incluyó 39 artículos que comprenden datos de 3009 estudiantes de preescolar y bachillerato, con diversos niveles de logro, de 14 países diferentes. Los resultados mostraron que los estudios tenían fallas metodológicas importantes, como tamaños de muestra pequeños o falta de grupos de control activos; también reportaron información insuficiente sobre elementos clave, como las herramientas empleadas para medir los resultados o las actividades específicas realizadas durante la capacitación, y revelaron signos de sesgos de publicación o reporte que impidieron una evaluación válida de la eficacia de la TIM aplicada en el aula. Se discuten las implicaciones educativas de estos resultados.

1. Introducción

Seguramente ya habéis leído mil veces la historia de cómo Howard Gardner montó su teoría, pero se hace indispensable recordarla una vez más. Así lo explican en este estudio:

En 1983 Howard Gardner publicó “Frames of Mind”, donde expuso, por primera vez, la idea de que los seres humanos poseen no una, sino múltiples inteligencias, cada una definida como la capacidad de resolver un problema específico o crear un producto que se percibe como valioso en uno o más entornos ricos en contexto. Según la propuesta inicial del autor, cada una de estas competencias cognitivas eran independientes entre sí, por lo que un mismo individuo podía ser fuerte en una inteligencia, pero débil en otra. Gardner (1993) estableció ocho criterios básicos para identificar una inteligencia: (a) aislamiento potencial por daño cerebral; (b) la existencia de sabios, prodigios y otras personas excepcionales; (c) una historia de desarrollo distintiva; (d) un conjunto definible de actuaciones o profesiones de expertos en el estado final; (e) apoyo de hallazgos psicométricos; (f) apoyo de tareas psicológicas experimentales; (g) un conjunto identificable de operaciones; y (h) susceptibilidad a la codificación en un sistema de símbolos. Siguiendo estos criterios, inicialmente propuso la existencia de siete inteligencias: lingüística, lógica matemática, musical, espacial, corporal cinestésica, interpersonal e intrapersonal. Catorce años después, la lista original se enriqueció con una nueva, la inteligencia naturalista (Gardner, 1997).



Junto con este nuevo enfoque de la conceptualización de la inteligencia humana, Gardner también propuso una nueva forma de evaluarla. En su opinión, los métodos predominantes en ese momento se centraban exclusivamente en la medición de las capacidades lingüísticas y lógicas. Además, por lo general consistían en tareas aisladas en papel y lápiz, desvinculadas de cualquier cultura y con frecuencia desconocidas para los niños Gardner (1993). Para superar estas limitaciones, Gardner sugirió lo que llamó “medidas justas de inteligencia”. Consistieron en una serie de actividades culturalmente significativas, siempre relacionadas con profesiones particulares, que permitieron evaluar los procesos psicológicos inherentes a cada inteligencia en sesiones de una a dos horas. Esta propuesta contribuiría a demostrar la existencia de inteligencias independientes y, por tanto, a identificar las fortalezas y debilidades de los individuos (Gardner, 1983). Debido a los altos costos de las medidas justas en términos de dinero y esfuerzo, Gardner propondría más tarde una nueva herramienta para medir la inteligencia, el Inventario de campo de espectro modificado, que muestreaba varias inteligencias en dos sesiones de una hora (Gardner & Hatch, 1989).

La teoría de las inteligencias múltiples (TIM) no se concibió originalmente para ser aplicada directamente a los entornos educativos (Gardner, 1983). Sin embargo, Gardner siempre se ha declarado convencido de su potencial contribución a este campo. Tenía entendido que las escuelas han asumido tradicionalmente que todos pueden aprender los mismos contenidos de la misma manera; por lo tanto, no han intentado asociar la experiencia de aprendizaje a productos específicos en la vida de la comunidad para mejorar una comprensión rica del conocimiento. Además, las escuelas han tendido a cultivar casi exclusivamente la simbolización lingüística y lógico-matemática y lo han hecho a través del aprendizaje repetitivo y memorístico (Gardner, 1991). Como resultado, muchos niños no logran descubrir sus dones o talentos (Gardner & Hatch, 1989). Según Gardner (2011), el sistema educativo puede mejorarse notablemente animando a los profesores a considerar las diferentes preferencias y fortalezas intelectuales de los estudiantes y enseñarles de manera coherente con ellas. En este sentido, la adopción de un plan de estudios de la TIM se convierte en una opción prometedora para fomentar el aprendizaje. Aunque el autor siempre ha defendido que la TIM se puede aplicar de numerosas formas, ha sugerido algunos medios específicos para ejercitar y desarrollar las inteligencias (Gardner, 1991). Más precisamente, el autor ha resaltado la importancia de crear oportunidades para trabajar intensamente con materiales ricos y atractivos que incluyen roles sociales y mejoran la participación de diferentes inteligencias humanas. Para lograr este objetivo, el aula se



organiza en diferentes rincones o áreas donde los niños pueden explorar libremente los materiales de aprendizaje y conectarse con el respectivo conjunto de inteligencias involucradas en ellos. Por ejemplo, un salón de clases puede incluir un rincón naturalista, donde los estudiantes pueden examinar y comparar especímenes biológicos con otros materiales (Gardner, 1991). Este tipo de intervención permite a los docentes identificar los talentos y fortalezas inesperadas de los niños sin ninguna evaluación específica y promover en ellos actitudes y habilidades básicas mediante el uso de recursos en los que han demostrado interés y experiencia emergente. Frente a una plétora de propuestas pedagógicas deficientes supuestamente inspiradas en la teoría, Gardner también ha establecido una distinción firme entre enfoques positivos y no correctos para aplicar su teoría (Gardner, 1995) y ha resaltado el valor de varios proyectos educativos como ejemplo de buenas prácticas (Gardner, 1997). Además, ha colaborado en la publicación de diferentes libros sobre cómo aplicar la TIM en el aula (Chen et al., 1998a, Chen et al., 1998b, Chen et al., 1998c; Chen, Moran, & Gardner, 2009) y mostró apoyo a trabajos relacionados de otros autores (Amstrong, 2009). Finalmente, actualmente lidera Project Spectrum, una iniciativa que tiene como objetivo ofrecer un nuevo enfoque para la evaluación y el desarrollo curricular en las escuelas (Harvard Graduate School of Education, 2016).

Hasta el día de hoy, la evidencia empírica que respalda la existencia de múltiples inteligencias no relacionadas entre sí es débil. Los elementos centrales de la TIM han sido criticados en innumerables ocasiones (Geake, 2008; Waterhouse, 2006; White, 2004; Willingham, 2004; pero ver Gardner & Morgan, 2006). En oposición a la afirmación de Gardner sobre la existencia de ocho inteligencias independientes, la comunidad científica cierra filas en torno a la naturaleza jerárquica de la inteligencia y la existencia de un factor general que explica un porcentaje notable de diferencias individuales (Colom, 2018; Hunt, 2001; Jensen, 1998; Lubinski, 2004; Visser, Asthon, & Vernon, 2006a). Además, un estudio reciente ha demostrado que muchas de las pruebas destinadas a medir los diferentes dominios en el marco de Gardner no solo están fuertemente intercorrelacionadas entre sí, sino también con pruebas externas de inteligencia general ((Visser, Asthon, & Vernon, 2006b; pero ver Gardner, 2006).

Más allá de la plausibilidad de la TIM, el objetivo del presente trabajo es evaluar su impacto en las escuelas. Desde su formulación, la TIM ha atraído el entusiasmo y el interés de un número creciente de profesores en todo el mundo (White, 2004). Aunque el propio Gardner (1983) reconoció que



había empleado el término inteligencia, y no talento o habilidad, para captar la atención de la audiencia, la idea de que todas las personas pueden ser inteligentes de alguna manera y que los educadores deben asegurarse de que todas las inteligencias sean igualmente inteligentes ejercida entre los estudiantes, ha sido acogida de todo corazón por gran parte de la comunidad educativa e inspiró la creación e introducción de un número considerable de metodologías y recursos en las escuelas (Amstrong, 2009; Collin, 2001; Gardner, 1997). Al mismo tiempo, la TIM ha estimulado la realización de varios estudios encaminados a medir el impacto de la teoría en el rendimiento académico de los estudiantes (Bas, 2016; Batdi, 2017). En general, estos estudios han analizado el efecto de la intervención de la TIM mediante el uso de actividades y materiales etiquetados por la TIM en resultados como ciencia, lectura o matemáticas, en comparación con métodos más tradicionales. En la mayoría de los casos, todas las inteligencias se han abordado en cada sesión de aprendizaje. Los resultados recogidos hasta la fecha sugieren que, en general, los estudiantes formados con la intervención de la TIM superan a los grupos de control. Sin embargo, como se detalla a continuación, no todos los estudios han encontrado un beneficio estadísticamente significativo para las intervenciones basadas en la TIM. Estos resultados contradictorios podrían explicarse parcialmente por diferencias metodológicas entre los estudios, como el tamaño de los grupos o la inclusión de un grupo de control. Sobre este último, solo un estudio incluyó un grupo de control activo (Modirkhamene & Azhiri, 2012). Los autores del estudio comprobaron el efecto de la intervención basada en la TIM en la comprensión lectora de una muestra de 70 estudiantes de secundaria durante dos meses. A pesar de los buenos resultados obtenidos a favor del grupo experimental, la inclusión de diversas prácticas etiquetadas como no válidas por el propio Gardner (1995) hace inviable considerar este trabajo como una buena estimación de lo que se puede esperar en esta área. Dada la importancia de basar las prácticas educativas en evidencia empírica sólida, el enfoque del presente artículo es precisamente realizar una revisión sistemática y un metaanálisis que permitan evaluar el impacto de las metodologías instruccionales inspiradas en la TIM en el rendimiento académico de los alumnos después de un cuidadoso examen de cualquier sesgo o problema metodológico.



Bien, una vez hecho este excelente repaso de historia de la TIM, pasamos a la parte del estudio donde se justifica el porqué de la necesidad de un nuevo metaanálisis sobre esta cuestión, existiendo ya dos anteriores que trataban el tema de la TIM. Según el punto de vista de los investigadores de nuestra revisión actual, en los dos metaanálisis anteriores existieron importantes deficiencias metodológicas, aplicando criterios de calidad mínimos a la hora de seleccionar los estudios y ni tan siquiera intentaron medir o controlar el riesgo de sesgo inducido por el diseño y procedimientos de estos.

En la misma línea, ninguno de ellos analizó el impacto potencial del sesgo de publicación o del informe selectivo sobre los resultados. Dadas estas deficiencias, quizás no sea sorprendente que ambos metaanálisis obtuvieran efectos promedio notablemente grandes ($d = 1.077$ y 0.95 , respectivamente). Con base en pruebas tan limitadas, es imposible evaluar si estos grandes efectos deben atribuirse a intervenciones genuinas basadas en la TIM o en las múltiples fuentes de sesgo que podrían influir en los resultados de los estudios individuales incluidos en los metaanálisis. Además, el procedimiento y los criterios utilizados para buscar estudios primarios no se definieron con suficiente detalle para reproducir los resultados y ninguna de las revisiones ofreció una lista de estudios incluidos y excluidos que pudieran utilizarse para confirmar y ampliar sus análisis.

Para diferenciarse de estos dos anteriores metaanálisis y mejorar los criterios de selección

“la estrategia de búsqueda bibliográfica adoptada en el presente estudio tuvo como objetivo localizar tanto los estudios publicados en revistas revisadas por pares a través de Web of Science como la literatura gris a través de ProQuest y Google Scholar, mediante un procedimiento bien definido que cualquier lector con acceso a estas bases de datos podrá reproducir. En segundo lugar, uno de nuestros principales objetivos era evaluar la calidad de cada estudio individual e identificar posibles fuentes de sesgo. Por último, a diferencia de los metaanálisis anteriores en este dominio, intentamos detectar y medir el impacto de los sesgos de publicación e informe.



El siguiente punto del estudio es uno de los que más me ha fascinado de todo este trabajo de investigación, por la escrupulosa búsqueda que se detalla en él y los criterios que se siguieron a la hora de escoger los de más calidad, todo un ejemplo de trabajo concienzudo y de excelencia.

Esta revisión sistemática sigue las recomendaciones de (Moher, Liberati, Tetzlaff, & Altman, 2009) y los estándares de informes de APA (Appelbaum et al., 2018) para revisiones sistemáticas y metaanálisis. El 1 de mayo de 2019, la primera autora (MF) realizó una búsqueda en Web of Science con el término inteligencias múltiples y el 20 de agosto de 2020 repitió la búsqueda en ProQuest y Google Scholar con el software gratuito Publish or Perish (Harzing, 2007). Estas búsquedas se limitaron a artículos en inglés publicados después de 1983 (año en el que Gardner publicó Frames of Mind). Después de eliminar 18 duplicados, la primera búsqueda arrojó 937 estudios en Web of Science, mientras que la segunda búsqueda arrojó 1642 estudios en ProQuest y 944 estudios en Google Scholar.

MF examinó los títulos y resúmenes de los estudios para determinar su elegibilidad. Los estudios solo se incluyeron si cumplían con los siguientes criterios de inclusión: (c1) el objetivo era medir cuantitativamente el impacto de una intervención basada en la TIM en el rendimiento académico de los estudiantes, y (c2) el estudio siguió un diseño pre-post con control grupo. Este paso resultó en un total de 38 artículos de texto completo en la primera búsqueda (Web of Science) y 232 artículos de texto completo en la segunda búsqueda (197 de ProQuest y 35 de Google Scholar). Luego, MF y SPL leyeron estos artículos de forma independiente para verificar que cumplieran con los criterios de inclusión. Como resultado de esta selección, se seleccionó un conjunto de 15 artículos en la primera búsqueda. A continuación, se realizaron búsquedas de descendencia de los artículos que citaban estos 15 artículos y se identificaron 129 nuevos estudios. Como en la etapa anterior, MF revisó los títulos y resúmenes de estos artículos, lo que resultó en 26 publicaciones adicionales de texto completo que fueron leídas de forma independiente por MF y SPL. Once de ellos cumplieron con los criterios de inclusión. Por lo tanto, la muestra de artículos revisados para su inclusión en la primera búsqueda consistió en 26 estudios (15 + 11). Asimismo, la selección de las 232 publicaciones a texto completo obtenidas en la segunda búsqueda resultó en la selección de 13 artículos (1 de ProQuest y 12 de Google Scholar). Por tanto, la muestra final de artículos revisados para su inclusión constó de 39 estudios (26 + 13). (...) En todos los artículos de texto completo evaluados para su inclusión, MF y SPL alcanzaron un acuerdo inicial entre evaluadores del 94,64%. Los desacuerdos se discutieron hasta alcanzar un consenso del 100%.



Una vez escogidos los artículos siguiendo esta metodología tan bien detallada, llega el momento de la extracción y codificación de datos.

MF y SPL codificaron de forma independiente cada uno de los 39 estudios seleccionados, incluyendo el título de la revista, tamaño y características de la muestra, tipo de experimentador (investigador o docente), duración de la intervención, variables dependientes y pruebas empleadas para medirlas.

Pero el esfuerzo por mejorar la calidad del resultado final no acaba aquí; una vez tuvieron los 39 estudios seleccionados, aún había de hacer una evaluación de los mismos sobre su calidad e identificación de posibles fuentes de sesgo. Para ello se elaboró una escala de calidad de 17 ítems (que se pueden consultar en el artículo original mediante una tabla).

Para cada estudio, MF y SPL asignaron de forma independiente un valor a cada ítem, alcanzando una concordancia inicial del 88,5%. Cuando hubo desacuerdos, estos se resolvieron mediante discusión hasta que se alcanzó un consenso del 100%.

A cada ítem se le podrían asignar tres valores: (a) positivo, cuando el estudio cumplió con el criterio; (b) negativo, cuando el estudio no cumplió con el criterio; y (c) desconocido, cuando el estudio no proporcionó información suficiente.

Nos encontramos pues ante estudios muy heterogéneos en diferentes términos: tipo de muestra (rendimiento académico o nivel socioeconómico), tamaño de la muestra (de 14 a 410 participantes), nivel educativo (desde preescolar hasta pregrado), duración de las intervenciones (de 2 a 10 semanas) o tipo de resultados (logros en lectura, matemáticas o ciencias).

¿Vamos pues a los resultados? Por lo que respecta a la evaluación cualitativa...

La mayoría de los artículos no reportan suficiente información para evaluar la calidad de la investigación. Para mencionar solo algunos ejemplos notables, el ítem 4, que se refiere al cegamiento de los participantes, solo se describe explícitamente en cinco estudios (12,82%), mientras que el ítem 15, relacionado con la validez de las pruebas empleadas para medir la variable dependiente, no se describe explícitamente en ningún estudio. Además, la



mayoría de los estudios no cumplen con una cantidad sustancial de criterios de calidad. Por ejemplo, el ítem 11, que se refiere al uso de un grupo de control activo, solo lo cumple un estudio (2,56%), mientras que los estudios restantes utilizan un grupo de control pasivo que está entrenado con un método tradicional (87,18%) o no aportan información sobre este tema (10,26%). Y el ítem 12, que evalúa si los artículos reportan suficiente información para replicar la intervención, solo se aborda en tres estudios (7,69%). De especial interés es el ítem 10, relacionado con la fidelidad de la intervención, donde la mayoría de los estudios (69,23%) no dan información sobre las actividades incluidas en la intervención. Seis estudios (15,38%) detallaron propuestas bien descritas, pero algunos de ellos pertenecen a las categorías etiquetadas como incorrectas por el propio Gardner. Solo 6 (15,38%) estudios cumplieron ambas condiciones, es decir, reportar información suficiente y comprender actividades sólidas, de acuerdo con los criterios de Gardner. Solo cuatro criterios de calidad fueron cumplidos por más del 50% de los estudios: ítem 7, referido a la igualdad de variables socioeconómicas entre grupos; Ítem 8, referido al análisis de los puntajes previos a la prueba; El ítem 14, referido a la confiabilidad de las pruebas empleadas para medir la variable dependiente, y el ítem 16, relacionado con la disponibilidad de información sobre al menos una de las comparaciones clave reportadas en el análisis de resultados.

Si nos fijamos en el metaanálisis cuantitativo, los resultados a nivel numérico son espectaculares. Son tales las cifras que salen en cuanto al tamaño del efecto (en algunos casos 7'16 o 3,76) que, analizando las causas de tal significatividad estadística, se encontró que en general, los tamaños de efecto más grandes provienen de estudios con los errores estándar más grandes (es decir, con las muestras más pequeñas). Todo parece indicar que los resultados del metaanálisis deben interpretarse con extrema precaución porque los tamaños del efecto informados pueden sobreestimar el tamaño de los efectos reales (Egger, Davey Smith, Schneider, & Minder, 1997).

Y eso no acabó aquí, ¿podían nuestros investigadores aplicar alguna fórmula o método que corrigiese estas cifras tan infladas y repuntuarlas de una manera más fidedigna a la realidad? De entrada, aplicaron 4 métodos diferentes para la detección y corrección del sesgo: el PET, el PEESE, el trim-and-fill (recortar y rellenar) y el modelo de selección de **Vevea and Hedges' (1995)**. ¿Qué conclusiones sacaron? Pues que



tres de las cuatro pruebas sugieren que las estimaciones promedio de nuestros metanálisis podrían estar infladas por sesgos de publicación o informe. Uno de ellos, PET, sugiere que los promedios corregidos por sesgo podrían no ser significativos. PEESE y trim-and-fill, por el contrario, sugieren que el efecto real podría ser diferente de cero, aunque sustancialmente menor de lo sugerido por las estimaciones metaanalíticas no corregidas.

Por poner ejemplos, había tamaños del efecto que originalmente marcaban 4,57 o 3'99 y aplicando el PET, obtenían una puntuación de 0'28 y 0'26 respectivamente. Hay más ejemplos numéricos aplicando los otros métodos que se pueden consultar en el artículo original, pero casi todos hacen una rectificación a la baja muy notable de unas cifras tan optimistas.

Resumen final

Desde que Gardner desarrolló su teoría sobre la existencia de inteligencias múltiples, un número creciente de profesores ha adaptado e incorporado la teoría a su rutina diaria en el (White, 2004). A pesar de este éxito inesperado, como el mismo Gardner ha reconocido de forma recurrente, no hay datos sólidos sobre la eficacia de la aplicación de intervenciones inspiradas en la TIM en el rendimiento académico de los estudiantes. Hasta la fecha solo existen dos metanálisis sobre este tema y, como hemos comentado anteriormente, ambos presentan importantes deficiencias metodológicas, como la ausencia de una evaluación de la calidad de los estudios incluidos o la falta de control del sesgo de publicación. El objetivo de la presente revisión sistemática fue evaluar la calidad de los estudios que prueban el impacto de las metodologías de instrucción inspiradas en la TIM en el rendimiento académico de los estudiantes, superando las fallas existentes de revisiones anteriores tanto como fuera posible.

En general, el análisis cualitativo de los resultados mostró que los estudios incluidos en esta revisión tienen fallas metodológicas importantes y reportan información insuficiente sobre elementos esenciales para hacer una valoración crítica de los métodos, como si los participantes e instructores eran ciegos a la manipulación experimental, o si las medidas empleadas eran fiables y válidas. Quizás lo más importante es que solo unos pocos estudios describieron la intervención realizada con suficiente detalle para permitir su reproducción. Es decir, no hay forma de saber en qué consistieron las intervenciones y cómo se midió la variable dependiente. Cuando se proporcionó información metodológica, muchos de los estudios no cumplieron con criterios de calidad importantes, como la asignación al azar de los participantes o la inclusión de un grupo de control activo. De hecho, la



mayoría de los estudios cumplieron claramente solo un par de criterios de calidad.

El análisis cuantitativo de los datos replica los resultados de metaanálisis anteriores, pero con importantes salvedades. (...) De acuerdo con ellos, encontramos tamaños de efecto notablemente grandes de $g\Delta = 1,49$ y $g_p = 1,15$. La magnitud de estos efectos debería, por sí sola, ser motivo suficiente para el escepticismo (Pashler, Rohrer, Abramson, Wolfson, & Harris, 2016). (...) Está claro que los efectos informados para las intervenciones basadas en la TIM revisadas aquí son mucho más grandes que los tamaños de efecto típicos reportados en la investigación psicológica (Funder & Ozer, 2019; Rubio-Aparicio, Marín-Martínez, Sánchez-Meca, & López-López, 2018).

¿Qué factores podrían explicar la notable diferencia entre los tamaños del efecto encontrados en los estudios actuales y los informados en otras áreas de la investigación educativa? Los gráficos en embudo representados en la Fig. 6 (en el artículo original) ofrecen una respuesta plausible a esta pregunta. Como puede verse, los tamaños de efecto más grandes provienen de los estudios con la precisión más baja, es decir, con el número más pequeño de participantes. Este patrón de resultados sugiere que el tamaño medio del efecto probablemente esté inflado por los (grandes) resultados de los estudios de menor calidad.

Además, todos los estudios encargados por la EEF y la NCEE deben cumplir con los más altos estándares metodológicos (...) Solo un puñado de los estudios revisados aquí cumplieron con estos estándares. Solo uno de los estudios incluyó un grupo de control activo. Esto es lamentable, porque la evidencia disponible muestra que los estudios educativos que se basan en grupos de control pasivo arrojan tamaños de efecto muy sobreestimados (Sala & Gobet, 2017). De hecho, la inclusión de un grupo de control activo se ha considerado una medida decisiva para probar la eficacia de las intervenciones educativas (por ejemplo, Datta, 2007), siempre que se garantice que las expectativas de los estudiantes en un grupo de control activo sean las mismas que las del grupo de los del grupo experimental (Boot, Simons, Stothart, & Stutts, 2013).

Ninguno de los estudios fue prerregistrado, lo que, nuevamente, es una protección esencial contra los sesgos en la investigación (Kaplan & Irvin, 2015; Warren, 2018) ya que reduce el grado de libertad de los investigadores y las prácticas de investigación cuestionables, como la publicación selectiva de análisis que "funcionaron" (Simmons, Nelson, & Simonsohn, 2011). De manera similar, el error de medición puede inflar los tamaños del efecto cuando se estima el tamaño del efecto de una población en tamaños de muestra pequeños, un sesgo cuyo impacto en los estudios



actuales es difícil de estimar porque la mayoría de ellos no reportó información psicométrica sobre las medidas dependientes. La figura 2 (en el artículo original) también muestra que ninguno de los artículos revisados declaró explícitamente que los participantes y los instructores eran ciegos a la manipulación experimental, lo que significa que los resultados de las intervenciones podrían deberse en su totalidad a las expectativas positivas de los participantes, como se mencionó anteriormente (Boot et al., 2013). Aunque es difícil, es posible cegar a los participantes e instructores mediante el uso de grupos de control activo donde los actores involucrados no saben si están siendo capacitados por la intervención en estudio o por una alternativa.

Dadas estas advertencias (y otros problemas resaltados en la Figura 2), el hecho de que los tamaños del efecto reportados en esta literatura sean grandes no es sorprendente. En nuestra opinión, esta literatura no debe tomarse como evidencia de que las intervenciones basadas en TIM funcionan. Con todo, aunque la mayoría de los estudios incluidos en el presente trabajo sugirieron que las intervenciones inspiradas en la TIM produjeron mejoras significativas en el rendimiento académico de los estudiantes, es imperativo interpretar estos resultados a la luz de las deficiencias críticas que han surgido en el campo cualitativo y análisis cuantitativos de los datos.

Para poner estos resultados en contexto, también es importante señalar que el principio principal de la TIM sobre la existencia de inteligencias múltiples no está respaldado por la comunidad científica. La investigación en psicología cognitiva ha señalado sistemáticamente la existencia de una inteligencia única, o factor general, que explica la mayor parte de la variación en el desempeño cognitivo en diferentes tareas (Lubinski, 2004; Visser et al., 2006a). Más relevante para este estudio, la afirmación central con respecto a la aplicación de la TIM en las escuelas carece de evidencia sólida. Presumiblemente, todas las inteligencias deben usarse como canales al presentar nuevos materiales para que los estudiantes experimenten el material a través de su mejor inteligencia y así se promueva la comprensión. Sin embargo, estudios en el campo de la psicología del aprendizaje han demostrado que la mejor manera de aprender algo suele estar definida por el contenido en sí, y no por las habilidades particulares o, en términos de Gardner, el perfil de inteligencias específicas de los estudiantes (Willingham, 2004). En otras palabras, según la mejor evidencia disponible hasta el momento, la enseñanza debe subordinarse al objeto de aprendizaje, no a las características de los alumnos individuales.



Aparte de estas importantes lagunas en la teoría y su traducción a la práctica en el aula, cualquier intento de probar la eficacia de las intervenciones inspiradas en la TIM en el futuro debería abordar las fallas metodológicas de la literatura existente que hemos destacado en la presente revisión. Idealmente, estos estudios deberían adoptar diseños experimentales, utilizar muestras grandes, garantizar el cegamiento de los participantes e instructores, incluir un grupo de control activo y seguir pautas detalladas de presentación de informes, incluida información precisa sobre la muestra, el procedimiento y los materiales empleados en el estudio, de modo que los resultados pueden ser replicados por investigadores independientes.

La TIM podría haber contribuido a repensar algunas cuestiones importantes entre los educadores, como el hecho de que los niños son únicos y valiosos independientemente de sus capacidades y que las escuelas son responsables de ayudarlos a todos a sacar lo mejor de sí mismos y a encontrar sus verdaderos intereses y fortalezas. O el hecho de que, con demasiada frecuencia, las escuelas se han centrado exclusivamente en habilidades puramente académicas, como la lectura o las matemáticas, a expensas de otras habilidades, como la música o la expresión corporal, lo que lleva a muchos niños a no encontrar sus verdaderos intereses y fortalezas. Teniendo en cuenta esta innegable contribución a la educación, es comprensible que muchos profesores hayan adoptado las intervenciones inspiradas en la TIM en el aula con gran entusiasmo. Sin embargo, como se muestra en el presente estudio, la evidencia recopilada hasta la fecha sobre la efectividad de estas acciones educativas no permite una evaluación válida de su impacto en el aprendizaje. Debido a la importancia de implementar métodos de instrucción bien fundamentados en la clase (Cook & Cook, 2004), es imperativo realizar investigaciones de alta calidad sobre la efectividad de la intervención basada en la TIM antes de que se pueda recomendar o promover su uso en el aula.

Ni se puede decir mejor, ni más claro. Este estudio pasa a ser automáticamente referente indispensable en la (excesivamente eterna) discusión sobre Gardner, su teoría y su aplicación en el aula. Una vez más, Marta Ferrero, Miguel Ángel Vadillo y Samuel Parra León dan muestra de la gran calidad de sus investigaciones. Estamos en buenas manos, no lo duden.

PODÉIS LEEROS ESTA REVISIÓN EN SU VERSIÓN COMPLETA Y ORIGINAL EN INGLÉS EN EL SIGUIENTE ENLACE:

[https://www.sciencedirect.com/science/article/pii/S016028962100050](https://www.sciencedirect.com/science/article/pii/S0160289621000507)