

Trabajo Final

David Fernández y Franco Vicario

1. Presentación del Problema

El fenómeno conocido como “**Efecto Pigmalión**” o también conocido como profecía autocumplida en el ámbito educativo propone que las expectativas que un profesor tiene sobre el rendimiento de un estudiante pueden influir causalmente en el desempeño real de este. El estudio pionero de Rosenthal y Jacobson (1968), *Pygmalion in the Classroom*, comentó que si a los maestros se les hacía creer que ciertos alumnos eran “promesas intelectuales” (aunque fueran seleccionados al azar), esos alumnos mostraban un incremento real en su **Cociente Intelectual (CI)** al final del año escolar.

Este descubrimiento generó una controversia grande en la psicología educativa y la metodología de la investigación. Mientras algunos estudios posteriores lograron replicar los resultados, otros encontraron efectos nulos o incluso negativos. Entonces Raudenbush (1984) recopiló estos estudios para intentar explicar dicha variabilidad.

El problema que afrontar este trabajo final es la necesidad de sintetizar esta evidencia dispersa. Dado que los estudios individuales presentan resultados contradictorios, el metaanálisis es la herramienta correcta para:

1. Estimar un tamaño del efecto global robusto.
2. Determinar que características del diseño experimental (heterogeneidad) explican por qué el efecto funciona en algunos estudios y en otros no.

2. Metodología Usada

Siguiendo las pautas de Harrer et al. (2021) en *Doing Meta-Analysis with R* y la línea del curso, se procedió con la siguiente estrategia metodológica.

2.1. Métricas del Tamaño del Efecto

Debido a que los estudios originales miden el desarrollo intelectual(CI) utilizando diferentes medidas, los puntajes directos de cada estudio no son directamente comparables entre ellos. Por eso, se utilizó la **Diferencia de Medias Estandarizada (SMD)**.

Específicamente, se usará la métrica de Hedges' g . A diferencia de la d de Cohen, la g de Hedges incluye un factor de corrección J que reduce el sesgo positivo que ocurre en muestras pequeñas (comunes en este dataset). La estimación para cada estudio k se define como:

$$g_k = \frac{\bar{x}_t - \bar{x}_c}{s_{pooled}} \times J$$

Donde s_{pooled} es la desviación estándar agrupada. En el código de R, esto se implementa cuando se pone el argumento `sm = "SMD"`.

2.2. Selección del Modelo: Efectos Aleatorios

Se seleccionó un **Modelo de Efectos Aleatorios** (*Random Effects Model*).

Justificación: El modelo de efectos fijos asume que todos los estudios estiman el mismo tamaño del efecto verdadero θ y que la única fuente de variabilidad es el error de muestreo (ϵ_k). Pero, los estudios de Raudenbush difieren en diseño(semanas de contacto), población y procedimientos. Entonces asumir que hay un único efecto verdadero idéntico para todos es irreal. El modelo de efectos aleatorios asume que los efectos verdaderos θ_k siguen una distribución normal:

$$\theta_k \sim N(\mu, \tau^2)$$

Donde τ^2 representa la varianza entre estudios.

Estimación de la Heterogeneidad (τ^2)

Para estimar τ^2 , se utilizó el método de **Máxima Verosimilitud Restringida (REML)** (`method.tau = "REML"`). Este método es preferible al estimador de DerSimonian-Laird debido a que evita el sesgo negativo en la estimación de la varianza, dando resultados más robustos cuando el número de estudios es moderado, como este caso($k = 19$).

2.3. Evaluación de Datos Faltantes

Se inspeccionó el conjunto de datos `raudenbush1985`.

Se verificó si había valores perdidos (NA) en los tamaños del efecto, varianzas y moderadores (weeks, setting, tester). El dataset se encuentra completo para las variables de interés. No fue necesario aplicar técnicas de imputación múltiple o análisis de sensibilidad por datos faltantes, ya que todos los estudios reportaron los estadísticos necesarios para el cálculo del tamaño del efecto.

2.4. Pertinencia del Análisis Multivariado

Se evaluó si era necesario aplicar un modelo de Metaanálisis Multivariado. Este tipo de análisis se usa cuando los estudios traen múltiples resultados correlacionados para los mismos sujetos o múltiples puntos temporales. En este caso, cada estudio aporta un único tamaño del efecto resumen (CI Global). Dado que los efectos son independientes entre estudios, el modelo univariado de efectos aleatorios es el enfoque estadísticamente correcto.

2.5. Estudio de Heterogeneidad y Sesgos

La variabilidad se evaluó mediante:

- **Estadístico I^2 :** Porcentaje de varianza debida a heterogeneidad real.
- **Análisis de Outliers:** Detección de intervalos de confianza que no solapan con el intervalo global.
- **Diagnósticos de Influencia:** Gráficos de Baujat y GOSH Plot.

Para el sesgo de publicación, se usaron métodos gráficos (*Funnel Plot*) y analíticos (Test de regresión de Egger y método *Trim-and-Fill* para imputación de estudios faltantes).

3 Resultados I: Modelo General y Heterogeneidad

A continuación, se presentan los resultados de la estimación del modelo de efectos aleatorios. Se usó el método de estimación **REML** (Máxima Verosimilitud Restringida) para el parámetro de heterogeneidad τ^2 , siendo considerado el enfoque de menor sesgo y más robusto para la estimación de la varianza inter-estudios en metaanálisis con un número chico de estudios ($k=19$).

3.1. Estimación del Efecto Global

Interpretación Estadística

La tabla resumen (`summary`) da la base estadística central del metaanálisis. Se detallan los hallazgos principales divididos por componentes estadísticos:

1. Estimación del Efecto Global (Modelo de Efectos Aleatorios)

La fila “Random effects model” presenta el resultado combinado ponderado bajo el supuesto de que hay heterogeneidad.

- **SMD (Tamaño del Efecto):** El valor estimado es 0.0837. Esto indica que, en promedio, el grupo experimental (con expectativas inducidas) superó al grupo control por solo 0.0837 desviaciones estándar. Esto se considera un efecto insignificante.
- **Intervalo de Confianza:** El intervalo [-0.0175; 0.1849] cruza el número cero. Implicando que con un 95% de confianza el verdadero tamaño del efecto podría ser tanto ligeramente negativo como positivo.
- **Estadístico z:** Este valor de prueba indica que la estimación del efecto se aleja 1.62 errores estándar del cero.
- **p-valor:** Al ser el valor $p > 0.05$, no se rechaza la hipótesis nula. No existe evidencia estadística suficiente para afirmar que las expectativas del profesor tengan un efecto sobre el Coeficiente Intelectual de los alumnos.

2 Intervalo de Predicción

El intervalo de predicción estimado es [-0.2243; 0.3917]. A diferencia del intervalo de confianza, este intervalo predice dónde caería el tamaño del efecto de un **futuro estudio**. Que este intervalo sea amplio y que incluya valores negativos demarca la incertidumbre y la dependencia del contexto del efecto Pigmalión: un nuevo experimento podría dar resultados perjudiciales o beneficiosos para los alumnos.

Visualización: Forest Plot

El *Forest Plot* es la representación gráfica que da una mejor visualización. Cada cuadrado representa el efecto de un estudio (su tamaño es proporcional al peso del estudio) y el rombo final representa el efecto global.

```
forest(m.gen, leftcols = c("studlab"), rightcols = c("effect","ci"))
```

Análisis Visual: Se ve una clara dispersión. Mientras que la mayoría de los estudios (ej: *Rosenthal et al*, *Fielder et al*) se agrupan cerca de la línea central de no efecto(cero), existen casos notables como *Pellegrini & Hicks* (Estudio 4) y *Marxwell* (Estudio 10), donde los intervalos de confianza estan significativamente hacia la derecha con efectos positivos grandes. Esta variedad de efectos que se ven es señal de heterogeneidad.

3.2 Análisis de Heterogeneidad

La heterogeneidad cuantifica la variabilidad en los resultados que no puede explicarse solo por el error de muestreo. Los estadísticos obtenidos son:

2. Heterogeneidad

Se evaluó la variabilidad entre los estudios.

- τ^2 (tau²): Representa la varianza total entre los efectos verdaderos de los estudios. Un valor distinto de cero confirma la presencia de variabilidad entre estudios.
- I^2 : Este estadístico dice qué porcentaje de la variación total en los estudios se debe a la heterogeneidad y no al azar muestral. El valor 49.8% se clasifica como heterogeneidad moderada, indicando que casi la mitad de la dispersión observada se debe a diferencias reales en los diseños o contextos de los estudios, pudiendo justificando el uso del modelo de efectos aleatorios.
- **Interpretación:** Dado que $p < 0.01$, se rechaza la hipótesis nula de homogeneidad. Esto confirma que los estudios no comparten un único efecto verdadero entre todos, volviendo a confirmar la decisión de usar un modelo de efectos aleatorios (`method.tau = "REML"`) en lugar de uno de efectos fijos.

Evaluación Inicial de Sesgo (Funnel Plot Preliminar)

Antes de profundizar en las causas de la heterogeneidad, vemos gráfico de la simetría de los datos y una prueba formal de asimetría.

```
#grafico de embudo  
funnel(m.gen)
```

El gráfico de embudo muestra visualmente una falta de estudios pequeños con resultados negativos. El test de regresión de Egger afirma esta asimetría positiva($p = 0.0115$). Esto puede sugerir en un **Sesgo de Publicación**: el efecto global de 0.08 observado está probablemente sobreestimado porque se ha ignorado continuamente los estudios pequeños que no encontraron el efecto Pigmalión.

4. Análisis de Valores Atípicos e Influencia

Debido la heterogeneidad moderada-alta ($I^2 \approx 50\%$), es imprescindible identificar si existen estudios específicos que estén distorsionando el promedio global (**Outliers**).

4.1 Detección de los Outliers

Se utilizó la función `find.outliers`, que identifica estudios donde sus intervalos de confianza no se superponen con el intervalo de confianza del efecto combinado.

Hallazgos: La función identificó dos estudios atípicos:

1. Pellegrini & Hicks (Estudio 4)
2. Maxwell (Estudio 10)

Análisis de Sensibilidad: La salida muestra que quitar estos dos estudios, la heterogeneidad I^2 baja del 49.8% al 9.1% y la prueba Q deja de ser significativa ($p = 0.34$). Demostrando que la “inconsistencia” del metaanálisis se debe en gran proporción por estos dos casos extremos. Sin embargo, se puede notar que el efecto global recalculado sin esos outliers ($SMD=0.03$) sigue siendo no significativo.

4.2. Diagnósticos de Influencia (Influence Analysis)

Para confirmar visualmente la influencia de estos estudios, utilizamos un conjunto de diagnósticos gráficos.

```
m.gen.inf <- InfluenceAnalysis(m.gen, random = TRUE)

# grafico baujat
plot(m.gen.inf, "baujat")
```

Interpretación del Gráfico de Baujat: El gráfico de Baujat mapea la contribución de cada estudio a la heterogeneidad global (eje X) frente a su influencia en el resultado del tamaño del efecto global (eje Y).

- Los estudios situados en la esquina **superior derecha** son los más problemáticos.
- Claramente, los estudios identificados previamente (etiquetados numéricamente) se destacan del resto, confirmando que aportan una cantidad desproporcionada de varianza al modelo.
- **Gráfico de Influencia:** Barras altas indican estudios que si se eliminaran cambiarían fuertemente los parámetros del modelo.

- **Gráfico I2:** Muestra como cambiaría el I^2 global si se eliminara cada estudio uno por uno. Se observa una caída notable si se eliminan los estudios 4 y 10.

4.3 Análisis GOSH (Graphic Display of Study Heterogeneity)

El análisis GOSH ajusta el modelo de metaanálisis a todos los subconjuntos posibles de estudios (combinatoria de $2^k - 1$ modelos). Dejando ver si los datos provienen de una única población homogénea o si existen clústeres o patrones distintos.

```
# Ajuste del modelo con metafor para pasar a GOSH
m.rma <- rma(yi = dat$yi, sei = dat$seTE, method = "REML", test = "knha")

# Ejecución del algoritmo GOSH
set.seed(123)
#res.gosh <- gosh(m.rma)
#plot(res.gosh, alpha = 0.02)
```

Interpretación del GOSH Plot: Cada punto en el gráfico representa un metaanálisis de un subconjunto de los datos.

- En el eje X: Tamaño del efecto.
- En el eje Y: Heterogeneidad (I^2 o H^2).
- Una distribución unimodal y compacta indicaría homogeneidad.
- Si observamos una distribución dispersa, alargada o con múltiples núcleos (nubes de puntos separadas), indica que existen subpoblaciones dentro de los datos que se comportan de manera distinta. En nuestro caso, la dispersión observada refuerza la necesidad de buscar variables moderadoras (subgrupos y regresión) para explicar estos patrones.

5. Resultados II: Análisis de Moderadores

Dada la heterogeneidad significativa que se obtuvo ($I^2 \approx 50\%$) y la presencia de *outliers*, no es suficiente quedarse con el efecto global. Es necesario averiguar si ciertas características del diseño de los estudios actúan como variables moderadoras.

5.1 Análisis de Subgrupos

Se analizaron dos variables categóricas: el entorno de instrucción (**setting**) y el conocimiento del evaluador (**tester**). Se compararon modelos asumiendo una varianza entre estudios (τ^2) separada para cada grupo versus una varianza común.

A) Subgrupo por Entorno

Se investigó si el efecto varía dependiendo de si la generación de expectativas se realizó en grupo (*Group*) o individualmente (*Indiv*).

Interpretación del Análisis de Subgrupos: Entorno

Dado que el subgrupo “individual” cuenta con un número muy bajo de estudios ($k=3$), se opta por interpretar los resultados del modelo con estimación de varianza común ($\text{tau.common} = \text{TRUE}$), ya que ofrece estimaciones más estables que el modelo de varianzas separadas.

Resultados por Subgrupo (Modelo τ^2 común): 1. **Entorno Grupal** ($k = 16$): El tamaño del efecto es irrelevante ($SMD = 0.05$; IC 95%: -0.04 a 0.14). 2. **Entorno Individual** ($k = 3$): El tamaño del efecto es moderado y estadísticamente significativo. ($SMD = 0.38$; IC 95%: 0.05 a 0.71)

Prueba de Diferencias entre Subgrupos: El test de heterogeneidad entre grupos (que evalúa la Hipótesis Nula de que los efectos en el grupo “Grupal” e “Individual” son idénticos) arrojó un valor $Q = 3.58$ con un p-valor = 0.0586. **Conclusión:** El resultado es lo suficientemente bajo como para sugerir una fuerte tendencia y una diferencia esencial en la práctica. El p-valor indica que solo hay un 5.86% de probabilidad de que una diferencia tan marcada (entre $SMD \approx 0.05$ y $SMD \approx 0.38$) haya ocurrido puramente por azar. Los datos indican que el efecto de las expectativas docentes se pierde casi por completo en entornos grupales, mientras que en interacciones profesor alumno “uno a uno” (individuales), el efecto Pigmalión parece ser mayor y positivo.

```
#forest plot por subgrupos
forest(m.setting_separateTau,
      xlab = "Diferencia de Media Estandarizada",
      print.subgroup.name = TRUE)
```

B) Subgrupo por Evaluador (Tester)

Se analizó si el hecho de que el evaluador conociera las expectativas (*aware*) o fuera ciego a ellas (*blind*) influía en los resultados (posible sesgo del observador).

Interpretación del Análisis de Subgrupos: Conocimiento del Evaluador (Tester)

Se investigó la hipótesis de que el “Sesgo del Observador” podría estar inflando los resultados. En teoría, si los evaluadores conocen que estudiantes supuestamente tienen un alto potencial

“Aware”), podrían evaluar sus tests de manera más generosa que aquellos que desconocen la condición experimental (“Blind”).

Para este análisis, los grupos están relativamente equilibrados en tamaño ($k = 10$ para conscientes vs. $k = 9$ para ciegos), por lo que los resultados son robustos. A continuación se presentan los hallazgos basados en el modelo de efectos aleatorios:

Resultados Descriptivos por Subgrupo: 1. **Evaluadores Conscientes (Aware):** El tamaño del efecto es trivial y cercano a cero ($SMD = 0.046$; IC 95%: -0.09 a 0.19). 2. **Evaluadores Ciegos (Blind):** Sorprendentemente, el efecto estimado es mayor ($SMD = 0.149$; IC 95%: -0.02 a 0.32), aunque sigue siendo estadísticamente no significativo.

Prueba de Diferencias entre Subgrupos: El test de heterogeneidad entre grupos dió un estadístico un p -valor = 0.3606.

Conclusión: El valor $p > 0.05$ indica que no existe una diferencia estadísticamente significativa entre los estudios donde los evaluadores sabían qué alumnos eran “Pígmalión” y los estudios donde no lo sabían. Esto conlleva a dos implicaciones importantes: 1. **Rechazo de la hipótesis del sesgo del observador:** No hay evidencia de que el conocimiento de las expectativas por parte del evaluador haya inflado los puntajes de Conocimiento Intelectual en estos estudios. 2. **Heterogeneidad no explicada:** A diferencia de la variable “Entorno”, la condición del evaluador no ayuda a explicar la variabilidad de los resultados (la heterogeneidad residual I^2 permanece alta en 51.6%). Por eso, el conocimiento del evaluador no es un factor relevante en este metaanálisis.

5.2. Metaregresión

Aparte de las categorías, la duración del experimento es una variable continua clave. Se ajustó un modelo de metaregresión utilizando la variable **weeks** (semanas de contacto).

$$\theta_k = \beta_0 + \beta_1 \times \text{Weeks}_k + u_k + \epsilon_k$$

```
#ajuste del modelo de metaregresión
mr.weeks <- metareg(m.gen, ~ weeks)
summary(mr.weeks)
```

Análisis de la Regresión:

- **Coefficiente (weeks):** El test de moderadores ($Q_M = 7.51$) confirma una relación significativa. La dirección negativa del coeficiente indica que por cada semana adicional de contacto entre profesor y alumno, el tamaño del “Efecto Pígmalión” cae. Esto sugiere que el efecto es fuerte al inicio pero se apaga con el tiempo.

- **Significancia:** Siendo la hipótesis nula: la duración del contacto (weeks) no tiene ningún efecto sobre el tamaño del efecto Pigmalión, el p-valor obtenido es 0.0061. Al ser este valor muy inferior a 0.05, podemos afirmar que la duración del experimento explica significativamente la heterogeneidad detectada. La variable weeks es un predictor determinante.
- R^2 : El modelo muestra un $R^2 = 40.63\%$. Esto indica que la variable “duración” es capaz de explicar por sí sola más del 40% de la varianza entre estudios que al principio atribuíamos a heterogeneidad desconocida. Como consecuencia, la heterogeneidad residual (I^2) desciende del 49.8% al 29.40%.

Gráficos de Burbujas (Bubble Plots)

Estos gráficos permiten visualizar la relación lineal entre la duración y el efecto. El tamaño de cada burbuja es proporcional al peso del estudio (inverso de la varianza).

```
#burbujas etiquetadas por nombre
bubble(mr.weeks,
       xlab = "Semanas de Contacto",
       ylab = "Tamaño del Efecto (SMD)",
       studlab = TRUE)
```

Conclusión: Los estudios cortos tienden a reportar efectos grandes, mientras que los estudios largos reportan efectos nulos. Esto indica que el efecto de las expectativas inducidas es pasajero.

6 Evaluación de Sesgos de Publicación

El sesgo de publicación pasa cuando los estudios con resultados significativos tienen mayor probabilidad de ser publicados. Se evaluó este fenómeno mediante gráficos y pruebas estadísticas formales.

6.1 Revisión Gráfica

En el Funnel Plot anteriormente ya visto, la asimetría que tiene sugiere que faltan estudios pequeños con efectos negativos.

6.2. Prueba de Regresión de Egger

Verificar la relación entre el efecto estandarizado y su precisión.

```
# radial base
radial(m.gen, main = "Radial plot con línea de Egger")

#creación de variables auxiliares para la regresion
yi_over_se <- m.gen$TE / m.gen$seTE      # variable dependiente Z-score
inv_se      <- 1 / m.gen$seTE            #Variable independiente precisión

#ajuste del modelo lineal (Egger)
egger_model <- lm(yi_over_se ~ inv_se)

# visualizacion de la linea de regresion sobre el radial plot
abline(egger_model, lwd = 2, col = "red")

#resumen estadístico del modelo manual
summary(egger_model)
```

Para visualizar y confirmar analíticamente la asimetría detectada, se calculó la regresión de Egger sobre las coordenadas del Gráfico Radial. Este procedimiento transforma los datos para evaluar la relación entre la magnitud del efecto estandarizado y su precisión.

Procedimiento: 1. **Variables:** Se generaron las coordenadas del Radial Plot: * Variable Dependiente (`yi_over_se`): Es el tamaño del efecto estandarizado ($z\text{-score} = \theta_k/SE_k$). * Variable Independiente (`inv_se`): Es la precisión del estudio ($1/SE_k$). 2. **Modelo:** Se ajustó un modelo de regresión lineal simple (`lm`). En este espacio transformado, la pendiente de la línea de regresión estima el efecto global, mientras que el **intercepto** estima el sesgo.

Interpretación de los Resultados del Modelo:

- **Intercepto (Bias):** El coeficiente del intercepto es **1.6243**. En ausencia de sesgo de publicación, esta línea debería pasar por el origen (intercepto = 0). Un valor de 1.62 sugiere que los estudios pequeños (con menor precisión) tienden a reportar efectos sistemáticamente mayores.
- **Significancia Estadística:** El p-valor asociado al intercepto es **0.0574**.
 - Este valor se encuentra en el límite de la significancia estadística (ligeramente superior a 0.05).
 - **Conclusión:** Aunque estrictamente no se rechaza la hipótesis nula de simetría al nivel del 5% con este método específico, el resultado es **marginalmente significativo**. Esto aporta evidencia sugestiva (aunque no definitiva por sí sola bajo este método) de que existe asimetría en la literatura, reforzando la necesidad de

aplicar métodos de corrección como *Trim-and-Fill* para asegurar la robustez de las conclusiones.

6.3. Pruebas Formales de Asimetría (Metabias)

Se aplicaron las variantes robustas de las pruebas de asimetría disponibles en la librería `meta`.

```
# 1. Prueba de correlación de rangos (Begg & Mazumdar)
metabias(m.gen, method = "rank")

# 2. Prueba de regresión lineal (Egger clásico)
metabias(m.gen, method = "linreg")

# 3. Prueba con estimador MM (Thompson & Sharp)
metabias(m.gen, method = "mm")
```

La consistencia entre estos métodos (especialmente si los p-valores son < 0.05 o < 0.10) refuerza la conclusión de que existe un sesgo de publicación en este conjunto de datos.

6.4. Método Trim-and-Fill (Recortar y Rellenar)

Para corregir el impacto del sesgo de publicación, se utilizó el método *Trim-and-Fill*. Este algoritmo imputa los estudios teóricamente “faltantes” para restaurar la simetría del *Funnel Plot* y recalcula el efecto global ajustado.

```
# Aplicación del algoritmo
tf <- trimfill(m.gen)

# Resumen del modelo ajustado (con estudios imputados)
summary(tf)

# Visualización del Funnel Plot con estudios imputados (puntos rellenos)
funnel(tf, main = "Trim and fill - Raudenbush 1985")
```

Interpretación Final del Sesgo: Observamos cuántos estudios fueron añadidos (imputados) y cómo cambia el efecto global (SMD). Si el efecto ajustado es mucho menor que el original (o cruza el cero), confirmamos que el sesgo de publicación estaba inflando artificialmente los resultados iniciales.

7. Estudio de Tamaño de Muestra y Potencia

Siguiendo las pautas del trabajo final (“Estudio de tamaño de muestra y potencia”) y utilizando la metodología descrita en el material complementario del curso (*Metapower.pdf*), se realizó un análisis de potencia *post-hoc* (retrospectivo).

El objetivo es determinar si el metaanálisis tenía la sensibilidad estadística suficiente para detectar un efecto, dado el número de estudios ($k = 19$) y la heterogeneidad observada.

```
# Carga de la librería para análisis de potencia (según PDF adjunto)
library(metapower)

# Parámetros basados en los resultados observados en la Parte 1 y 2:
# - effect_size: 0.08 (El efecto global que encontramos)
# - study_size: 60 (Promedio aproximado de participantes por estudio en el dataset)
# - k: 19 (Número de estudios)
# - i2: 0.50 (Heterogeneidad aproximada del 50%)

power_result <- mpower(effect_size = 0.08,
                      study_size = 60,
                      k = 19,
                      i2 = 0.50,
                      es_type = "d")

print(power_result)
```

Interpretación de la Potencia

El análisis revela que la potencia estadística para detectar un efecto tan pequeño ($d = 0.08$) con la heterogeneidad existente es **baja** (inferior al 20%). Esto significa que existe una alta probabilidad de cometer un **Error Tipo II** (falso negativo). Sin embargo, esto también sugiere que si el efecto fuera realmente grande o moderado (ej. $d = 0.5$), el metaanálisis lo habría detectado. El hecho de no haber encontrado significancia estadística refuerza la idea de que el “Efecto Pigmalión”, de existir, es un fenómeno de magnitud muy pequeña o altamente dependiente del contexto.

8. Discusión y Alcance de los Resultados

Este trabajo ha sintetizado la evidencia de 19 experimentos sobre la inducción de expectativas docentes. Los resultados desafían la narrativa convencional y simplista sobre la potencia de las expectativas del profesor.

1. **Efecto Global Nulo:** Al contrario de lo que sugiere la literatura clásica (Rosenthal & Jacobson, 1968), nuestro metaanálisis de efectos aleatorios no encontró evidencia significativa de un efecto global ($p = 0.10$). Tras ajustar por el sesgo de publicación detectado (mediante *Trim-and-Fill*), el efecto se reduce prácticamente a cero ($SMD_{ajustado} = 0.02$).
2. **El Rol Crítico del Tiempo (Metaregresión):** El hallazgo más robusto y revelador de este estudio proviene del análisis de metaregresión. La variable **weeks** mostró una relación inversa significativa ($p < 0.01$).
 - Los estudios de contacto muy breve (0-2 semanas) mostraron efectos grandes.
 - Los estudios de mayor duración mostraron efectos nulos.
 - **Alcance:** Esto sugiere que las expectativas inducidas artificialmente son efímeras. Los profesores pueden alterar su comportamiento inicialmente basándose en la información falsa recibida, pero conforme pasa el tiempo (semanas), la interacción real con el alumno y su desempeño objetivo prevalecen sobre la expectativa inducida, haciendo desaparecer el efecto.
3. **Heterogeneidad Explicada:** El modelo logró explicar una gran parte de la varianza ($R^2 \approx 40\%$) a través de la duración del contacto, lo que indica que la inconsistencia en la literatura previa no era aleatoria, sino sistemática y dependiente del diseño temporal.

9. Conclusiones y Futuros Pasos

En base al análisis estadístico realizado, se concluye que:

1. **No existe evidencia robusta** para afirmar que las expectativas docentes aumenten el CI de los alumnos de manera generalizada.
2. **La duración modera el efecto:** El fenómeno parece ser un artefacto de corto plazo que se diluye rápidamente con la experiencia directa del docente.
3. **Sesgo de Publicación:** Existe evidencia de que los estudios pequeños con resultados negativos han sido sistemáticamente menos publicados, inflando la percepción histórica del efecto.

Futuros Pasos: Se recomienda que futuras investigaciones abandonen el uso del CI global como variable dependiente, ya que es una medida estable y difícil de modificar, y se enfoquen en variables más sensibles como el rendimiento académico específico, la autoeficacia o la motivación del estudiante, controlando rigurosamente el tiempo de exposición.

10. Referencias Bibliográficas

- Harrer, M., Cuijpers, P., Furukawa, T.A., & Ebert, D.D. (2021). *Doing Meta-Analysis with R: A Hands-On Guide*. Chapman & Hall/CRC Press.

- Raudenbush, S. W. (1984). Magnitude of teacher expectancy effects on pupil IQ as a function of the credibility of expectancy induction: A synthesis of findings from 18 experiments. *Journal of Educational Psychology*, 76(1), 85–97.
- Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom: Teacher expectation and pupils' intellectual development. *Holt, Rinehart & Winston*.
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1-48.

Citas de paquetes de R utilizados:

Apéndice: Sentencias de R Utilizadas

Para garantizar la reproducibilidad de este estudio, se adjunta la totalidad del código utilizado, el cual se ha presentado y ejecutado a lo largo de las secciones previas de este documento dinámico.