

# APLICACIÓN DEL ÍNDICE TAU-U EN DISEÑOS DE EVALUACIÓN DE PROGRAMAS EDUCATIVOS

**MATAS TERRÓN, Antonio**

Universidad de Málaga

Málaga (España)

amatas@uma.es

## **Resumen**

El objetivo de esta comunicación es presentar el índice TAU-U de Parker, Vannest, Davis y Sauber (2011) como complemento a técnicas visuales para confirmar cambios en diseños de caso único AB. Para ello, se muestra un ejemplo de su utilización a partir de la ratio de escolarización entre mujeres y hombres en España entre 1981 y 2010, según datos oficiales del Banco Mundial. Finalmente, se comentan las implicaciones metodológicas de este índice, así como su potencial para la evaluación de programas educativos.

## **Abstract**

The aim of this paper is to show how the TAU-U index by Parker, Vannest, Davis and Sauber (2011) can be used complementing visual techniques in single-case AB designs. Odds of schooling women vs men in Spain, from 1981 to 2010 were analyzed to develop the TAU-U index. Data were from the World Bank database. Eventually, methodological implications, as well as its potential in educational programs evaluation, are commented.<sup>1</sup>

## **Palabras clave**

Evaluación, Investigación social, TAU-U

## **Keywords**

Evaluation, Social Research, TAU-U.

## INTRODUCCIÓN

En investigación evaluativa a veces es necesario recurrir al diseño de caso único, tipo AB o ABA (Léon y Montero, 2003) que suelen clasificarse entre los diseños pre-experimentales (Salkind, 1998).

Tradicionalmente, cuando el diseño de evaluación es de este tipo, el análisis se ha realizado de forma visual a partir de las gráficas generadas con los datos (Bono y Arnau, 2014). No obstante, la literatura al respecto suele advertir que los errores en este tipo de estrategia son elevados (Campbell y Herzinger, 2010) entre otras razones por no contar con medidas suficientes, o por un error de juicio al identificar las tendencias en cada fase.

Por tal motivo, se han propuesto otras técnicas alternativas. Entre estas, los índices de estimación del tamaño del efecto de medidas de no solapamiento. Estos índices presentan algunas ventajas tales como: no requieren supuestos paramétricos, no necesitan gran cantidad de observaciones, aplicables en gran variedad de diseños (Sanz y García-Vera, 2015) y su cálculo es sencillo (Parker, Vannest y Davis, 2010).

Estos índices generan una medición de la magnitud del cambio generado por la intervención, y no sólo la significación estadística. En todo caso, no se debe confundir la existencia de cambio con la utilidad de dicho cambio. Es decir, los cambios registrados por estos índices (como por otras técnicas) pueden o no implicar que el cambio sea suficiente a nivel práctico.

Entre los índices de no solapamiento se encuentra el TAU-U (Parker, Vannest, Davis, & Sauber, 2011). Este índice se desarrolló como un índice que se ajusta a partir de la tendencia de las medidas que conforman la línea base en un diseño AB. Es una evolución del índice NAP (Nonoverlap of All Pairs) de Parker y Vannest (2009).

Para calcular el índice TAU-U supóngase que se tienen datos de las fases A y B de un diseño AB de caso único, donde la fase A tiene  $m$  observaciones y la fase B tiene  $n$  observaciones. Sean entonces  $y_1^A, \dots, y_m^A$  los registros de la fase A, y por otro lado  $y_1^B, \dots, y_n^B$  los de la fase B, TAU-U se calcula como sigue (expresión Expresión):

$$TAU - U = \frac{S_p - S_B}{mn} \text{Expresión 1}$$

Dónde:

$S_p$  es el estadístico S de Kendall calculado por la comparación entre fases para cada par de observaciones (expresión Expresión).

$$S_p = \sum_{i=1}^m \sum_{j=1}^n [I(y_j^B > y_i^A) - I(y_j^B < y_i^A)] \text{Expresión 2}$$

$S_B$  es el estadístico S de Kendall calculado por la comparación dentro de la fase A (línea base) para cada par de observaciones (expresión Expresión).

$$S_B = \sum_{i=1}^{m-1} \sum_{j=i+1}^m [I(y_j^A > y_i^A) - I(y_j^A < y_i^A)] \text{Expresión 3}$$

Para ilustrar el procedimiento, a continuación se expone un ejemplo basado en Brossart,



Esta versatilidad de TAU y TAU-U los hacen índices muy útiles, aunque es necesario tener claro qué se pretende (Brossart, Vannest, Davis y Patience, 2014).

Por último, la interpretación del efecto del tamaño sigue el razonamiento de que bajas puntuaciones representan una baja efectividad del programa y viceversa. Para un criterio más operativo, Rakap (2015) proponen los siguientes valores:

- Iguales o inferiores a 0.65: no efectivo o cuestionable.
- Entre 0.66 y 0.92: programa efectivo.
- Superior a 0.93: programa muy efectivo.

El objetivo del presente estudio es mostrar la utilidad del indicador TAU-U para valorar los cambios en series de datos. Se recurre a un ejemplo con datos reales.

## MÉTODO

Para el presente estudio demostrativo se lleva a cabo un estudio de una serie temporal que se analiza con el algoritmo TAU-U. Los datos proceden de la base de datos del Banco Mundial (<http://data.worldbank.org/indicador/SE.ENR.PRSC.FM.ZS>).

De todos los índices disponibles en la base de datos, se seleccionó el porcentaje de mujeres escolarizadas en primaria y secundaria frente al de hombres. El indicador consiste en dividir la tasa de matrícula de mujeres escolarizadas entre la tasa de hombres. Un porcentaje de 100 indica igual proporción entre mujeres y hombres. Si el porcentaje es menor a 100 indica que hay más proporción alumnado masculino, mientras que si el porcentaje es mayor a 100 indica que hay mayor proporción de alumnado femenino escolarizado.

El proceso se inició recogiendo los datos de la base de datos del Banco Mundial y seleccionando solamente la serie histórica de España que tiene un rango de 30 años (desde 1971 a 2011). Posteriormente se elaboró la gráfica de la serie y se seleccionó un momento de entre todos aquellos que visualmente parecían mostrar un cambio de tendencia (e.g. 1974-1975, 1979-1980, 1982-1983, y 1996-1997). Se eligió el cambio entre 1996 y 1997 por ser el más reciente. De esta forma, los datos anteriores a 1997 se utilizaron para elaborar la línea base. En total, eran 25 observaciones.

Puesto que parte de ellas coincidían con los años de la “transición” política en España (1974-1978 aproximadamente), y puesto que se disponía de un número de observaciones amplio, se optó por usar sólo la serie histórica que va desde 1980 a 1996 para la línea base.

Por su parte, los datos de 1997 y posteriores se asignaron a la fase B típica de los diseños AB donde A es la línea base y B es la serie donde se interviene.

Para el análisis se ha usado R (R Core Team, 2015), el plugin SCDA de R-Commander y el script de Tarlow (2013) disponible en [https://dl.dropboxusercontent.com/u/2842869/Tau\\_U.R](https://dl.dropboxusercontent.com/u/2842869/Tau_U.R).

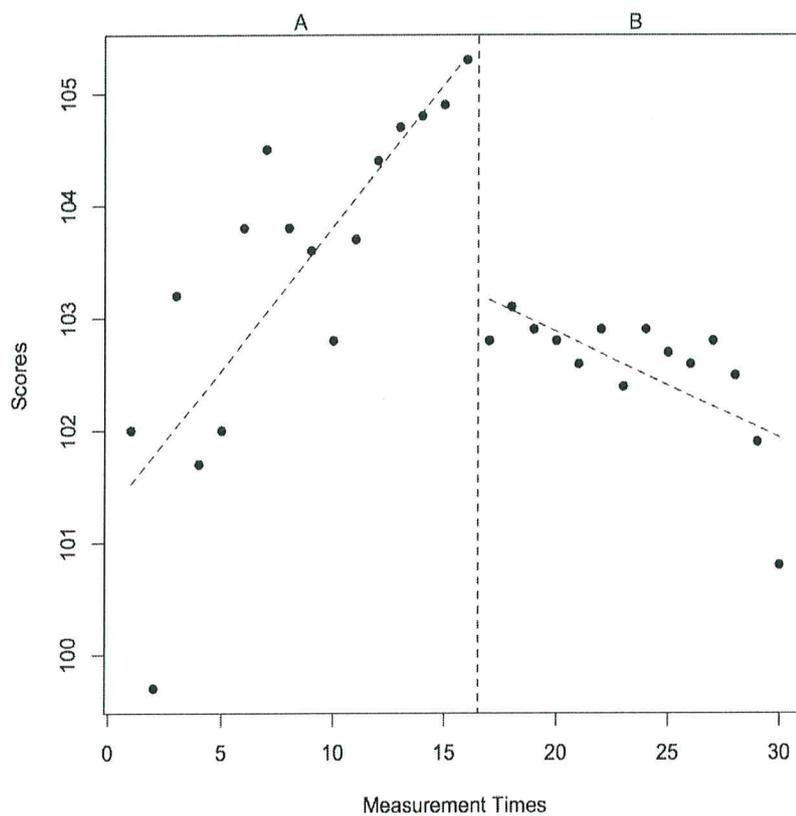
## RESULTADOS

Inicialmente se realizó un análisis de regresión de las dos series de datos: línea base (fase A) y intervención (fase B).

En el caso de la fase A (figura Illustration) se puede observar una tendencia ascendente (expresión Expression) que se inicia en 103.8 puntos y concluye en 105.3 puntos con una mediana de 103.7 (media= 103.4; d.t.=1.48).

$$Y = 0.254X + 101.2 \text{ Expresión 5}$$

En la expresión Expression se muestra la ecuación de la recta procedente de la regresión lineal.



**Figura 2.** Líneas de tendencias

Con relación a la fase B, la figura Figura muestra la distribución de puntos para la observaciones desde 1997 a 2010. En el mismo se puede observar una clara tendencia descendente que queda también de manifiesto en la regresión lineal (expresión Expression).

$$Y = -0.094X + 103.27 \text{ Expresión 6}$$

La fase B se inicia en 102.8 puntos y concluye en los 100.8 puntos, con una mediana de 102.7 (media=102.5; d.t.=0.58).

La dispersión de los datos parece también ser distinta para las dos series. Mientras que la fase A presenta un Coeficiente de Variación de 0.014, en la fase B es de 0.005. En la figura se muestra la variabilidad de las observaciones en ambas fases.

Por su parte, los resultados de la prueba TAU-U se han llevado a cabo con el script de Tarlow (opus cit.). En primer lugar, los resultados ofrecen la proporción de comparaciones en la cual las observaciones de la fase B son más grandes que en la fase A (224 pares, siendo 55 positivos y 166 negativos). Esta comparación entre la fase A y la fase B obtienen un valor Tau de -0.49 (d.t.=48.01; p=0.022). En segundo lugar, en la fase A se contabilizan 120 pares (101 positivos y 17 negativos). Este resultado indica la tendencia de la fase A.

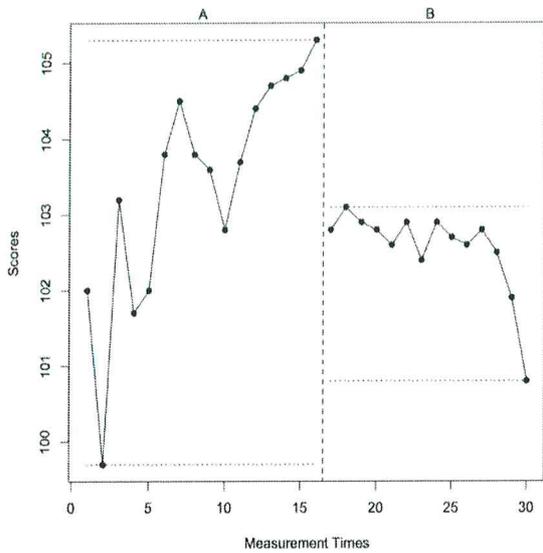


Figura 3. Variabilidad de las obseraciones

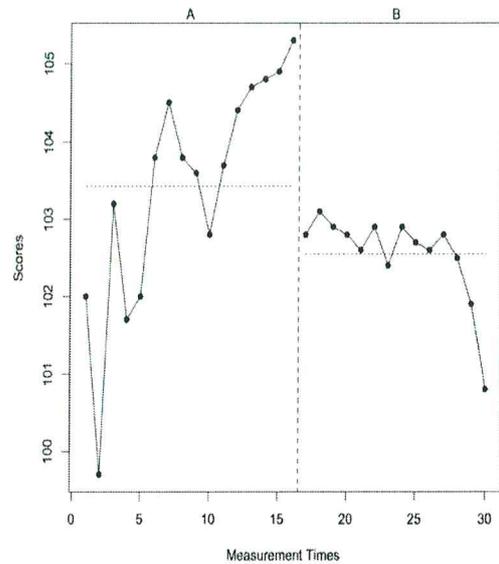


Figura 4. Promedio de las fases

Por último, la comparación de la fase A con la fase B, menos la tendencia de la fase A, indica la cantidad de observaciones no solapadas controlando el efecto de la fase A. En este caso, el valor de Tau es de -0.56 (d.t.=52.88; p=0.0002). En la tabla se muestran estos resultados junto con las diferentes comparaciones que genera el script.

Tabla 1. Resultados del script de R para el cálculo de TAU-U

|        | A vs B  | trendA  | Trend B | A vs B – trend<br>A | A vs B +<br>trendB | A vs B + trendB - trendA |
|--------|---------|---------|---------|---------------------|--------------------|--------------------------|
| #pairs | 224     | 120     | 91      | 344                 | 315                | 435                      |
| #pos   | 55      | 101     | 17      | NA                  | NA                 | NA                       |
| #neg   | 166     | 17      | 67      | NA                  | NA                 | NA                       |
| S      | -111    | 84      | -50     | -195                | -161               | -245                     |
| Tau    | -0.4955 | 0.7     | -0.5495 | -0.5669             | -0.5111            | -0.5632                  |
| SD(S)  | 48.0199 | 22.1660 | 18.0370 | 52.8890             | 51.3495            | 55.9136                  |
| p      | 0.0220  | 0.0002  | 0.0066  | 0.0002              | 0.0018             | 0.0000                   |

## DISCUSIÓN Y CONCLUSIONES

La discusión y las conclusiones derivadas se organizan en dos partes. La primera sobre los resultados del ejemplo, y la segunda sobre las implicaciones del índice TAU-U como recurso para la evaluación de programas.

### Sobre el ejemplo

En primer lugar, hay que advertir que este estudio se ha recurrido a un artificio conceptual. En el diseño AB, donde la fase B es de intervención, no se ha producido ninguna intervención. Sin embargo, se ha partido de la hipótesis de trabajo implícita de que en el sistema educativo se produjo un cambio en 1997 que provocó un cambio en los registros documentales.

Bajo estas hipótesis de trabajo, el análisis visual mostró un cambio en la pendiente de la tendencia entre las fases. Mientras que la fase A la pendiente era positiva, en la fase B dicha pendiente se invirtió claramente. Por otro lado, se registró una variabilidad menor en la serie de la fase B. Sin embargo, el solapamiento entre las fases fue evidente. A pesar de ello, los gráficos sugieren una diferencia de nivel entre fases en función del promedio.

Los mismos resultados anteriores eran coherentes con el coeficiente de variación. En definitiva, los resultados a través del análisis visual y a través de un análisis estadístico descriptivo clásico sugieren diferencias en la variabilidad de las observaciones, así como en la tendencia.

Sin embargo, también queda patente que la nube de puntos de la fase A presentaba un claro solapamiento al con distribución de la nube de puntos de la fase B. Esta situación podría sugerir que, aunque existan cambios entre fases, estos no sean los suficientemente destacados como para tenerlos en cuenta. Dicho de otra forma, queda la sospecha de que la serie de la fase B, pueda considerarse o no de la misma población de los datos de la serie A.

Por su parte, el índice TAU da una medida del cambio producido (tamaño del efecto). En el ejemplo, el gráfico muestra un cambio claro de tendencia, sin embargo, los valores TAU indican que dicho cambio es moderado. No obstante, el tamaño del efecto pasa a ser algo mayor cuando se controla el efecto de la tendencia de A (TAU-U). En cualquier caso, el cambio es muy bajo y cuestionable, no pudiendo aportar información que sustente la idea de que realmente se asistió a un cambio de tendencia en la serie histórica, aunque los índices sean estadísticamente significativos.

### Sobre el índice TAU-U

El principal interés del índice TAU-U en investigación en Educación, radica en su potencial en la evaluación de programas educativos.

En primer lugar, es un índice cuyo algoritmo es relativamente fácil de calcular. Al mismo tiempo, presenta pocas restricciones para su aplicación. No obstante, algunos investigadores han hecho notar que el índice TAU-U es sensible al número de observaciones en cada fase (Pustejovsky, 2016) lo que debe ser tenido en cuenta.

Por otro lado, ofrece una aproximación al tamaño del efecto del cambio producido, y no

solamente a su significación estadística. En este sentido, es una estrategia analítica que puede ayudara superar la dicotomía entre resultados útiles para la práctica de la evaluación frente a los resultados estadísticos de la propia evaluación. No obstante, es evidente que este debate está presente y que debe ser abordado desde el rigor metodológico y desde el sentido común, tal como queda patente en la revisión literaria realizada.

En otro orden de cosas, el índice TAU y TAU-U se diseñaron para estudios de caso único tipo AB, ABA y similares. Por tanto, en evaluación de programas, la fase B coincide con la puesta en marcha del programa educativo.

Sin embargo, en evaluación educativa también se puede utilizar el índice cuando realmente no hay programa, sino cuando se ha producido un hecho social de impacto. Es el caso del ejemplo que se ha desarrollado en estas páginas. En este caso, la fase B es realmente una fase de post-impacto. En estos casos, la validez del estudio queda en entredicho puesto que el hecho impactante no suele estar controlado por el investigador. Para este tipo de estudios, es recomendable recurrir a diseños de investigación evaluativa basados en la “Teoría del Cambio” y entre ellos se recomienda el Análisis de la Contribución (Mayne, 2012).

Por último, tal como advierten diversos autores con relación a los diseños experimentales de caso único (DECU o SCED de su denominación en inglés) es siempre aconsejable utilizar varias estrategias para contrastar los resultados (Manolov, Moeyaert & Evans, 2015).

Finalmente, se sintetizan los pasos seguidos en este estudio. Con ello se pretende ofrece un esquema para la utilización de estos índices de forma coherente en los casos donde no existe realmente una intervención educativa programada:

- Identificar las variables de interés y acceder a las series históricas.
- Identificar momentos dentro de la serie donde se observe un aparente cambio de tendencia o una diferencia destacada.
- Tomar los datos previos como fase A y los posteriores como fase B.
- Analizar visualmente la secuencia AB.
- Analizar con el índice TAU y TAU-U la serie A, secuencia AB (índice TAU) y secuencia AB corrigiendo el efecto de la serie A (índice TAU-U).
- Extraer conclusiones sobre la significatividad del cambio.
- Extraer conclusiones sobre la magnitud del cambio.

## REFERENCIAS BIBLIOGRÁFICAS

- Bono, R., & Arnau, J. (2014). Diseños de caso único en Ciencias Sociales y de la Salud. Madrid: Síntesis.
- Brossart, D.F., Vannest, K.J., Davis, J.L., & Patience, M.A. (2014). Incorporating nonoverlap indices with visual analysis for quantifying intervention effectiveness in single-case experimental designs. *Neuropsychological Rehabilitation: an International Journal*, 24(3-

4), 464-491.

- Campbell, J.M., & Herzinger, C.V. (2010). Statistics and single subject research methodology. En D.L. Gast (Ed.), *Single subject research methodology in behavioral sciences* (pp. 417-453). Nueva York: Routledge.
- León, O. & Montero, I. (2003). *Métodos de investigación en Psicología y Educación*. Madrid: Mc Graw-Hill.
- Manolov, R., Moeyaert, M. & Evas, J.J. (2015). *Resources and guidelines for analysing SCED data*. Disponible en: [http://www.ub.edu/gcai/soft/Software\\_resources\\_for\\_applied\\_researchers\\_April2015.pdf](http://www.ub.edu/gcai/soft/Software_resources_for_applied_researchers_April2015.pdf) [Consultado 29 noviembre 2016].
- Mayne J. (2012). Contribution analysis: Coming of age? *Evaluation* 18(3), 270–280.
- Parker, R. I., & Vannest, K. J. (2009). An improved effect size for single case research: Non-overlap of all pairs (NAP). *Behavior Therapy*, 40(4), 357-367.
- Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011). Combining non-overlap and trend for single-case research: Tau-U. *Behavior Therapy*, 35(4), 202-322.
- Parker, R.I., Vannest, K.J., Davis, J.L., & Sauber, S.B. (2010). Combining non-overlap and trend for single case research: Tau-U. *Behavior Therapy*, 42, 284–299.
- Pustejovsky, J.E. (2016). TAU-U. Disponible en : <http://jepusto.github.io/Tau-U> [Consultado 29 Enero 2017].
- R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Rakap, S. (2015). Effect sizes as result interpretation aids in single- subject experimental research: description and application of four nonoverlap methods. *British Journal of Special Education*, 42(1), 11-33.
- Salkind, N. (1998). *Métodos de Investigación*. Pearson: México D.F.
- Sanz, J. & García-Vera, M.P. (2015). Técnicas para el análisis de diseños de caso único en la práctica clínica: ejemplos de aplicación en el tratamiento de víctimas de atentados terroristas. *Clínica y Salud*, 26, 167-180.
- Tarlow, K. (2013). Tau\_R. Disponible en: [https://dl.dropboxusercontent.com/u/2842869/Tau\\_U.R](https://dl.dropboxusercontent.com/u/2842869/Tau_U.R). [Consultado 6 febrero 2017].