



Revista de Investigación Educativa

ISSN: 0212-4068

rie@um.es

Asociación Interuniversitaria de Investigación
Pedagógica
España

Díaz Costa, Elisabet; Fernández-Cano, Antonio; Faouzi, Tarik; Henríquez, Carlos Felipe
Validación del constructo subyacente en una escala de evaluación del impacto de la investigación
educativa sobre la práctica docente mediante análisis factorial confirmatorio
Revista de Investigación Educativa, vol. 33, núm. 1, enero-julio, 2015, pp. 47-63
Asociación Interuniversitaria de Investigación Pedagógica
Murcia, España

Disponible en: <http://www.redalyc.org/articulo.oa?id=283332966011>

- ▶ Cómo citar el artículo
- ▶ Número completo
- ▶ Más información del artículo
- ▶ Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica

Red de Revistas Científicas de América Latina, el Caribe, España y Portugal
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

Díaz Costa, Elisabet; Fernández-Cano, Antonio; Faouzi, Tarik; Henríquez, Carlos Felipe (2015). Validación del constructo subyacente en una escala de evaluación del impacto de la investigación educativa sobre la práctica docente mediante análisis factorial confirmatorio. *Revista de Investigación Educativa*, 33(1), 47-63.
DOI: <http://dx.doi.org/10.6018/rie.33.1.193521>

Validación del constructo subyacente en una escala de evaluación del impacto de la investigación educativa sobre la práctica docente mediante análisis factorial confirmatorio

Validation of the underlying construct in a rating scale to assess the impact of educational research on teaching practice using confirmatory factor analysis

Elisabet Díaz Costa*

Antonio Fernández-Cano**

Tarik Faouzi***

Carlos Felipe Henríquez****

*Universidad Católica de la Santísima Concepción, **Universidad de Granada,

Universidad del Bío-Bío, *Universidad de Valparaíso

Resumen

Este artículo muestra las etapas de validación de una escala de evaluación del impacto de la investigación educativa en la práctica docente, enfatizando los procedimientos de validez del constructo subyacente. Para ello, se tuvieron en consideración el problema a investigar y los criterios de rigor científico propios de la construcción de instrumentos de encuesta, concretamente la escala de evaluación. Operando con datos procedentes de tal escala aplicada a una muestra de 62 docentes universitarios y 117 docentes no universitarios de Granada (España), mediante un análisis secundario de datos, se desarrollan las diferentes etapas de validación de constructo; obteniéndose tres constructos representados por 16 sentencias (ítems) de los 43 inicialmente disponibles.

Correspondencia: Antonio Fernández-Cano, Dpto. MIDE. Facultad, CC. Educación, Campus de Cartuja, 18071 Granada, Tel./Fax: 958244253. E-mail: afcano@ugr.es

El estudio muestra la potencialidad del análisis factorial confirmatorio frente a la debilidad del exploratorio en la indagación de la validez de constructo de un instrumento de medida con variables ordinales, cual es la escala de valoración, y, al par, en la reducción de ítems de la misma.

Palabras clave: validación de escala, impacto de la investigación educativa, análisis factorial exploratorio, análisis factorial confirmatorio, reducción de ítems.

Abstract

This article examines the validation stages of a rating scale to assess the impact of educational research on teaching practice, paying particular attention to the procedures to validate the underlying construct. To achieve this aim, we considered the research problem, sample availability, and criteria of scientific rigor for developing survey instruments, specifically a rating scale. Data drawn from a rating scale are applied to a sample of 62 teachers at the University of Granada (Spain) and 117 non-university teachers. Through a secondary analysis of the data, the different stages to validate the construct were developed, obtaining three constructs comprising 16 items of the 43 items available. The study shows that confirmatory factor analysis is more powerful than exploratory factor analysis in validating the underlying construct of a measurement instrument (rating scale) and in reducing the scale items.

Keywords: rating scale validation, impact of educational research, exploratory factor analysis, confirmatory factor analysis, item reduction.

Introducción

Es bastante conocido el hecho que la educación no ha logrado articularse, frente a otras disciplinas y campos científicos, como un campo basado en evidencias dado sobre todo por el cuestionable impacto de la investigación sobre la práctica docente (Burkhardt y Schoenfeld, 2003; Escudero, 2006; Fernández Cano, 2001; Kaestle, 1993; Schneider y Keesler, 2007; Vanderlinde y van Braak, 2010). Por ello, este artículo tiene como propósito indagar el constructo latente bajo la noción de impacto de la investigación educativa en la práctica docente; para lo cual se describen las decisiones asociadas a las oportunidades y dificultades halladas en el proceso de validación de un instrumento de recogida de información, como es la escala de evaluación. Dado que la calidad de un instrumento de tipo encuesta está dada por la validez de constructo, la primera tarea del proceso de validación del instrumento (escala) es delimitar las dimensiones asociadas a tal objeto de investigación y con ello, el grado en que el instrumento refleja las teorías relevantes del fenómeno que pretende medir (Martín, 2004; Outón y Suárez, 2010).

Indagar el fenómeno del bajo impacto de la investigación educativa, percibido en la literatura afín, pasa por desarrollar un constructo que represente las conclusiones de los principales estudios desarrollados hasta entonces y arriba ya citados. De esta revisión, se estimarán tres dimensiones teóricas que abarcan el posicionamiento de los científicos a este respecto y que serían la base de los futuros ítems de la escala en curso, que hemos venido en denominar: Escala para la Evaluación del Impacto de la Investigación Educativa sobre la Práctica Docente (desde ahora abreviadamente EI/IE-PD).

En la validación de la escala EI/IE-PD se puede considerar un análisis factorial exploratorio y un análisis factorial confirmatorio. La solución exploratoria y sus índices unifactoriales asociados para la reducción de ítems ya fueron aplicados para analizar las características métricas de este instrumento en el marco de una investigación anterior (Díaz, 2010; Fernández-Cano et al., 2008); en concreto, los siguientes índices para cada ítem: desviación típica, correlación biserial puntual ítem-total, alfa de Cronbach y el tamaño del efecto, entre grupos previamente formados (investigadores/universitarios *versus* docentes/no universitarios: dos micro-culturas desconectadas, que actuarán como grupos divergentes), con su nivel de significación *p* asociado.

En este estudio se realizará para una validación constructual completa con ambos análisis (exploratorio y el confirmatorio) y como consecuencia de tal validación, la plausible reducción de ítems de la escala. Un objetivo adicional de este estudio será desarrollar entonces el modelo más parsimonioso, reduciendo el número de ítems de la escala sin alterar el constructo representado por el mejor modelo; entendiendo que la búsqueda de parsimonia métrica, como principio teórico básico de la medida y como norma en investigación aplicada, debiera ser un logro capital, pese a que la investigación al respecto no ha dado aún respuestas totalmente satisfactorias según se constata en Larwin y Harvey (2012).

Método

Diseño

El diseño de este estudio se caracteriza por ser descriptivo, cuantitativo, tipo encuesta (Buendía, Colás y Hernández-Pina, 1998); más en concreto, se trata de un estudio muestral prométrico que coadyuva al establecimiento de inferencias evaluativas. También es un estudio explicativo de la estructura subyacente (constructo) en un instrumento de medida, escala de valoración.

Instrumento

Las dimensiones teóricas, posibles futuros factores, se conformaron inductivamente a partir de categorías conceptuales obtenidas tras realizar una acuciosa revisión de los antecedentes científicos (ver literatura citada en *Introducción*); para posteriormente redactar 43 ítems que constituyen los elementos iniciales de tal escala, tal como se expone en la Tabla 1.

Se definieron tres dimensiones teóricas del impacto de la investigación educativa en la práctica docente: a) Diagnóstico del impacto de la investigación educativa sobre la práctica docente, en adelante DIIE, b) Acciones para mejorar el impacto de la investigación educativa sobre la práctica docente, en adelante AMIIE y c) Oportunidades que ofrece la mejora del impacto de la investigación educativa sobre la práctica docente, OMIIIE. Las sentencias, según dimensiones, se distribuyeron de la siguiente manera: el primer constructo, DIIE, con 20 ítems, mientras que el segundo constructo, AMIIE, estuvo compuesto de 13 ítems y el último, OMIIIE, con 10 ítems (Tabla 1).

Tabla 1

Dimensiones teóricas y categorías conceptuales propias de la escala ei/ie-pd, validada por expertos (Díaz, 2010)

| Dimensión teórica | Categorías conceptuales generadoras de ítems | Ítems derivados (#) |
|--|---|---|
| Diagnóstico del impacto de la investigación educativa sobre la práctica docente | <ul style="list-style-type: none"> - Lejanía entre profesores y conocimiento científico. - Dificultades del conocimiento científico para impactar sobre la práctica docente. - Tradición de investigación cuantitativa con baja incidencia social. - Actividad educativa como acto previo a la investigación educativa - Mala calidad de la información disponible sobre el alumnado y sus maneras de aprender. - Escasa importancia de la investigación educativa como consecuencia del bajo impacto que tiene sobre la práctica. - Falta de rigurosidad en la reflexión sobre la práctica docente. - Escasez de financiación para la investigación educativa. | 1-2-3-5-7-8-9-11-16-17-22-26-27-28-30-36-39-43 |
| Acciones para mejorar el impacto de la investigación educativa sobre la práctica docente | <ul style="list-style-type: none"> - Fomentar desde el aula la reflexión crítica, la práctica juiciosa, la colaboración y la investigación acción. - Plantear problemas de investigación desde contextos de aplicación. - Favorecer las aportaciones transdisciplinares. - Ampliar situaciones en las que se podría generar conocimiento. - Promover la responsabilidad social en la producción de conocimiento orientándolo a la solución de problemas. - Establecer nuevos criterios para evaluar la calidad de la producción científica. - Emprender investigaciones con incidencia directa (intervención o mejora) e indirecta (conocimiento acumulado). - Contrastar y confirmar la teoría educativa en la práctica. - Potenciar el desarrollo profesional a partir de la figura del “profesor investigador”. | 12-13-14-15-18-20-21-23-25-31-32-33-34-35-40-41 |
| Oportunidades para la mejora del impacto | <ul style="list-style-type: none"> - Agendas de investigación cargadas de significación para los actores. - Posibilidad para los docentes de reconstruir el conocimiento pedagógico a través de la reflexión y la utilización del mismo. - Enriquecimiento profesional de todos los participantes del proceso de investigación educativa. - Consolidación de políticas educativas trascendidas de manera democrática. - Aportaciones de otras áreas del conocimiento, con el consiguiente prestigio para ellas. - Cambio de percepción del impacto de la investigación educativa sobre la práctica docente. | 4-6-10-19-24-29-37-38-42 |

Como primera versión se diseñó escala criterial de cinco puntos, que constaba de 43 ítems a valorar de 1 a 5 (1= Muy en desacuerdo, 2= En desacuerdo, 3= Indiferente, 4= De acuerdo y 5= Muy de acuerdo) y fue administrada individualmente como un cuestionario a docentes universitarios y no universitarios (Díaz, 2010).

Muestra disponible

No es fácil acceder a una muestra representativa y de tamaño suficiente de docentes para indagar este problema. Para el estudio, se recurrió a datos de la muestra utilizada

por Díaz (2010) realizando un análisis secundario de los mismos. La accesibilidad a sujetos respondientes y su participación altruista fue ardua y muchas veces resultó difícil obtener sus respuestas; a través de un muestreo no probabilístico de tipo disponible se obtuvo una muestra compuesta por 179 respondientes cuantificada por 62 docentes universitarios y 117 no universitarios.

Técnicas de análisis de datos

A partir del análisis de datos generados por la muestra disponible a la que se le aplicó la escala EI/IE-PD, se continuó con la validación de la misma insistiendo en la estructura del constructo subyacente. Para ello, se aplicó el análisis confirmatorio sobre dos modelos mencionados en la tesis de Díaz (2010), mientras que se usó el análisis confirmatorio para hallar un modelo que se ajustara más a los datos. Finalmente, se compararon los tres modelos para encontrar el más idóneo. En este análisis se utilizó el software libre del programa R-Project, versión 3.0.2.

La diferencia entre un AFE y un AFC radica en que el análisis exploratorio trata de encontrar el modelo factorial subyacente que se ajuste mejor a los datos, mientras que el AFC permite evaluar el ajuste de los datos a determinada estructura factorial previamente hipotetizada (Lévy-Mangin y Varela, 2006; Thompson, 2004). En primer lugar, hubo que comprobar la unidimensionalidad, fiabilidad y validez de los constructos adoptados para la medición de las diferentes variables latentes (Farrell, 2010).

Procedimientos analíticos de la validez de constructo de la escala EI/IE-PD

La validez de constructo se llevó a cabo a través del método de análisis factorial confirmatorio (Lévy-Mangin y Varela, 2006) tras analizar el ajuste a los siguientes criterios dados en la Tabla 2:

Tabla 2

Criterios de validez del modelo de análisis factorial confirmatorio (AFC)

| | |
|---|--|
| <ul style="list-style-type: none"> - Los índices de ajuste de los datos de un modelo: CFI (<i>Comparative Fit Index</i>/Índice de Ajuste Comparativo); TLI (<i>Tucker Lewis Index</i>/Índice de Tucker-Lewis) y Chi cuadrado normalizada (χ^2/g_l). - El índice de bondad de ajuste más robusto: Error Cuadrático Medio Cuadrático de Aproximación (RMSEA). - La contribución factorial de cada ítem. Ésta debe ser mayor que 0.6 y significativa según el test W de Wald (ver figuras). - Los índices de modificación calculados por el test W de Wald. Las modificaciones sugeridas serán aceptadas sólo si pueden ser justificadas por el marco teórico. | <ul style="list-style-type: none"> - La fiabilidad de cada constructo verificada con rhô de Jöreskog ($\rho_{Jor} > 0.7$). - La validez convergente verificada con rhô de la varianza extraída ($\rho_{vc} > 0.5$). - El índice estadístico del coeficiente de determinación, R^2, está verificada si es mayor que 0.5 para cada ítem. - La validez discriminante estará verificada cuando la varianza compartida entre los constructos (variables latentes) es menor que la varianza compartida entre cada constructo y sus ítems. - La unidimensionalidad del conjunto de los constructos dada por la validez discriminante. |
|---|--|

Resultados

En primer lugar, expondremos los resultados del análisis factorial exploratorio (desde ahora AFE), por componentes principales con rotación varimax, que ya fue realizado por Díaz (2010). En segundo lugar, se hará revisión de lo obtenido en el análisis factorial confirmatorio (desde ahora AFC) para el modelo propuesto por los expertos en sus producciones escritas, luego del análisis de contenido previo, y para el modelo derivado del AFE.

Aplicando análisis factorial exploratorio (AFE)

El trabajo de Díaz (2010) consistió en un estudio factorial basado en la versión española de la EI/IE-PD para lo que fue necesario llevar a cabo un análisis factorial de tipo exploratorio; el cual se efectuó mediante el método de extracción de componentes principales y con normalización varimax como método de rotación. La solución obtenida fue satisfactoria, pues la medida de adecuación muestral de Kaiser-Meyer-Olkin alcanzó un valor de 0.84 y la prueba de esfericidad de Bartlett es altamente significativa ($p<.001$).

En la tabla 3, se muestran las componentes cuyo valor o propio *eigen value* (λ) fue igual o superior a 1 (Boza y Méndez, 2012). Las once primeras componentes cumplieron con esta condición y representaron un 64.9% de la varianza total explicada. A través del análisis de las propiedades métricas del instrumento, se pudo identificar entre estas once componentes, cuatro que contenían el total de los ítems mejores evaluados según sus características métricas (Díaz, 2010); o sea, eran los ítems con valores métricos unifactoriales más altos: mayor correlación biserial puntual ítem-total alta y significativa, alfa de Cronbach alta y significativa y el alto tamaño del efecto entre grupos previamente formados ($d>.05$) con su nivel probabilidad p de la diferencia significativa ($\alpha>.05$).

Tabla 3

Varianza total explicada y valores propios mayores que 1 en el AFE (Díaz, 2010)

| Componente | Sumas de las saturaciones al cuadrado | | | | | |
|------------|---------------------------------------|---------------------------|----------------------|----------------|---------------------------|----------------------|
| | de la extracción | | | de la rotación | | |
| | Total | Porcentaje de la varianza | Porcentaje acumulada | Total | Porcentaje de la varianza | Porcentaje acumulada |
| 1 | 9.53 | 22.15 | 22.15 | 7.15 | 16.63 | 16.63 |
| 2 | 4.96 | 11.53 | 33.68 | 3.53 | 8.21 | 24.85 |
| 3 | 2.41 | 5.61 | 39.30 | 2.86 | 6.67 | 31.52 |
| 4 | 1.87 | 4.35 | 43.65 | 2.20 | 5.12 | 36.64 |
| 5 | 1.63 | 3.80 | 47.45 | 2.19 | 5.11 | 41.75 |
| 6 | 1.53 | 3.57 | 51.01 | 1.88 | 4.39 | 46.15 |
| 7 | 1.38 | 3.21 | 54.22 | 1.83 | 4.27 | 50.42 |
| 8 | 1.29 | 2.99 | 57.22 | 1.61 | 3.75 | 54.18 |
| 9 | 1.22 | 2.83 | 60.05 | 1.57 | 3.66 | 57.84 |
| 10 | 1.05 | 2.45 | 62.50 | 1.54 | 3.58 | 61.43 |
| 11 | 1.03 | 2.40 | 64.89 | 1.48 | 3.46 | 64.89 |

*Valores truncados al segundo decimal.

Estas componentes son: la primer componente con un total de varianza explicada de 22.15%, la segunda con una varianza explicada de 11.53%, la tercer componente con una varianza explicada menor que sus antecesores pero que se retuvo pues su valor propio es mayor que 1 para el total, 5.61%, y por último, se retuvo una cuarta componente que explica un 4.35% de la varianza total explicada. Además, todos los ítems contribuían al menos a una de las cuatro componentes consideradas.

Las cuatro primeras componentes alcanzaron valores de λ suficientes, mayores que 1, como para ser retenidos; además estas cuatro componentes eran manifiestamente interpretables y los ítems que las saturaban alcanzaban índices unifactoriales asociados para la reducción de ítems aceptables (Díaz, 2010). Su denominación e interpretación es como sigue:

La componente I se denominó *Diagnóstico del bajo impacto de la investigación educativa en la práctica docente*. Esta componente incorporó 18 ítems con cargas significativas (al usual $a > 10.31$) y, según los resultados de la validación, es el que mejor explica la varianza del instrumento, por lo que es el componente más significativo en el constructo que se ha propuesto indagar.

La componente II y la componente IV podrían interpretarse agrupadamente bajo la nominación de: *Acciones para mejorar el impacto de la investigación educativa sobre la práctica docente*. Estas componentes fueron explicadas a partir de las cargas significativas de 16 ítems. La racionalidad de esta agregación estriba en que ambas componentes podrían ser interpretadas genéricamente por determinadas acciones; así, la componente II estaba saturada en su mayoría por ítems que representaban acciones positivas a mantener mientras que la componente IV se configuraba en mayor grado por acciones negativas a omitir según los ítems que la saturaban (Díaz, 2010).

La componente III se denominó *Oportunidades que ofrece la mejora del impacto de la investigación educativa sobre la práctica docente*. A esta componente contribuyeron 9 ítems. Obviamente algunos ítems saturaban dos o más componentes.

Al analizar las coincidencias entre los ítems reducidos a partir de las propiedades métricas, y el AFE, se concluye que sólo 25 ítems saturaban en una y sólo una componente con cargas superiores a 10.61, un valor de parsimonia aceptable por conservador y riguroso, aunque sólo 23 de ellos mostraron propiedades métricas unifactoriales aceptables (Tabla 4).

Tras hacer el análisis de fiabilidad de los 25 ítems seleccionados, se comprobó además que el índice alfa de Cronbach de cada ítem disminuye al compararlo con el alfa de la prueba total ($\alpha_{cro} = 0.76$), y también disminuía en función del total de ítems de los cuatro primeros componentes ($\alpha_{cro} = 0.77$). Finalmente, se optó por mantener los 25 ítems de las cuatro primeras componentes, a pesar de que dos ítems (12 y 35) no se mostraron como significativos según el análisis de las características métricas del instrumento (Tabla 5). Esta decisión fue tomada atendiendo al aumento de la fiabilidad de estos 25 ítems sobre los resultados de fiabilidad de los 23 ítems obtenidos del estudio de propiedades métricas. Así, la escala EI/IE-PD quedó compuesta por los ítems que incluían los cuatro primeros componentes del AFE con un total de varianza explicada igual a 43.65% (Díaz, 2010).

Tabla 4

Dimensiones e ítems de la escala ei/ie-pd deducida del AFE (Díaz, 2010)

| Dimensión | Categorías conceptuales generadoras de ítems | Ítems derivados (#) |
|--|--|--|
| Diagnóstico del impacto de la investigación educativa sobre la práctica docente | <ul style="list-style-type: none"> - Investigadores y docentes no tienen intereses y motivaciones comunes. - Los docentes consideran que su práctica no mejora con la ayuda de la investigación educativa. - La investigación educativa está descontextualizada de la realidad del aula. - Los hallazgos de la investigación educativa no son aplicables al trabajo de aula. | 1-2-3-7-8-9- 17-22-24-27- 30-36-43 |
| Acciones para mejorar el impacto de la investigación educativa sobre la práctica docente | <ul style="list-style-type: none"> - La investigación educativa se vería beneficiada con un lenguaje común entre investigadores y docentes. - La investigación educativa debería incentivarse también para y con los docentes. - Necesidad de crear una figura que sirva de nexo entre la investigación educativa y los centros escolares (experto) - Mejorar la formación en investigación de los docentes tendente a mejorar su participación en las mismas. | 12-14-16-18- 21-25-33-35 |
| Oportunidades que ofrece la mejora del impacto de la investigación educativa sobre la práctica docente | <ul style="list-style-type: none"> - Las innovaciones educativas basadas en investigaciones educativas tienen más posibilidades de ser beneficiosas para los alumnos que las que no se basan en sus hallazgos - La investigación educativa podría ser una labor más del profesor, siendo responsables o colaboradores de la misma. - Oportunidad de acercamiento entre la investigación educativa y la práctica del docente. | 10-19-26-38 |

a > |0,60|; λ > 1

Tabla 5

Índices de bondad de ajuste de los datos al modelo de conjunto de medidas (Díaz, 2010)

| Modelo | Total ítems | Ítems retenidos | Criterion* |
|----------------------------|-------------|---|--|
| Modelo deducido por el AFE | 25 | 1, 2, 3, 7, 8, 9, 10, 12, 14, 16, 17, 18, 19, 21, 22, 24, 25, 26, 27, 30, 33, 35, 36, 38 y 43 | Regla de Kaiser, altas cargas y ajuste a propiedades métricas mencionadas (Díaz, 2010) |

*a>|0,60|; λ>1,

Aplicando análisis factorial confirmatorio (AFC)

Este análisis permite probar la unidimensionalidad de las variables latentes o los constructos, DIIE, AMIIE, y OMIIIE con el método WLSMV (*Weighted Least Squares Mean and Variance Adjusted*) –mínimos cuadrados ponderados con estimación robusta de errores

estándar— aplicado a correlaciones policóricas (para datos ordinales, se busca conjuntos de ítems que estén intra-correlacionados, usando tal correlación policórica). Por defecto, WLSMV usa la correlación policórica en el caso que como aquí los ítems estén definidos como ordinales, tal como acontece en las escalas de evaluación sumaria (tipo Likert, 1932) y con las escalas de valoración (*rating scales*). Téngase presente que este método ha sido el más utilizado para los datos categóricos (Muthén, 1993), a pesar de que fue diseñado originalmente para su uso con datos no-normales y continuos (Brown, 2006).

Para investigar la estructura interna de cada constructo: DIIE, AMIIE y OMIIE, se realizó un estudio de la validez de la escala de medida para cada constructo con todos los ítems. En la Tabla 6, se encuentran los valores de los índices de ajuste de los modelos de medidas de las variables latentes DIIE, MIIE y OMIIE para los dos modelos utilizados por Díaz (2010), el validado por expertos (Tabla 1) y el reducido a partir del AFE (Tabla 4). Se trata de verificar si existe un buen ajuste según el AFC, para posteriormente compararlos con el mejor modelo que tiene un buen ajuste, el hallado por la matriz policórica.

Tabla 6

Índices de bondad de ajuste de modelos total, deducido y reducido

| Modelo | Variable latente | Nº ítems | χ^2/gl | CFDLI | RMSEA | CFI | TLI | VR ² | ρ_{jor} | ρ_{vc} |
|--|------------------|----------|-------------------|-------|-------|-------|-------------|-----------------|--------------|-------------|
| Modelo validado por expertos (con todos los ítems) | DIIE | 18 | 3.54 [-0.04;0.77] | 0.119 | 0.845 | 0.824 | [0.00;0.60] | 0.36 | 0.88 | |
| | AMIIE | 16 | 2.52 [0.00;0.65] | 0.092 | 0.732 | 0.690 | [0.00;0.42] | 0.19 | 0.75 | |
| | OMIIE | 9 | 2.30 [0.30;0.75] | 0.085 | 0.960 | 0.946 | [0.09;0.56] | 0.34 | 0.81 | |
| Modelo generado por AFE (modelo deducido) | DIIE | 13 | 1.85 [-0.45;0.78] | 0.069 | 0.973 | 0.968 | [0.20;0.62] | 0.46 | 0.92 | |
| | AMIIE | 8 | 4.10 [0.32;0.68] | 0.131 | 0.864 | 0.809 | [0.10;0.46] | 0.30 | 0.77 | |
| | OMIIE | 4 | 0.32 [0.55;0.73] | 0.000 | 1.000 | 1.018 | [0.30;0.53] | 0.40 | 0.72 | |
| Modelo generado por el AFC (Modelo reducido) | DIIE | 9 | 1.00 [0.65;0.80] | 0.000 | 1.000 | 1.000 | [0.43;0.64] | 0.51 | 0.90 | |
| | AMIIE | 3 | 1.33 [0.61;0.78] | 0.043 | 0.992 | 0.987 | [0.38;0.61] | 0.50 | 0.74 | |
| | OMIIE | 4 | 1.30 [0.65;0.79] | 0.041 | 0.992 | 0.995 | [0.43;0.62] | 0.49 | 0.79 | |

CFDLI: Intervalo de la contribución factorial de los ítems.

RMSEA: Root Mean Squared Error of Approximation (Error Medio Cuadrático de Aproximación).

VR²: Intervalo de la variación de los coeficientes de determinación R².

Para valores de referencia, véase Tabla 2.

Los valores de los índices de los constructos del primer y segundo modelo no ajustan adecuadamente. Probablemente se debe a la existencia del problema de multicolinealidad entre los ítems. Por ello, frente a la insuficiencia de los resultados obtenidos se realizó un análisis de la contribución factorial de cada ítem con el fin de generar un modelo que se ajuste mejor a los datos. En este modelo, se aceptaron los ítems con correlaciones policóricas mayores que 0.45; a partir de este criterio, para el constructo DIIE, se retuvieron los ítems 1, 2, 7, 8, 9, 17, 22, 27 y 43, mientras que para el constructo

AMIIIE se retuvieron los ítems 14, 21 y 42. La variable latente OMIIIE quedó configurada por los ítems 24, 28, 38, y 42. El conjunto de estos ítems retenidos configuran la base del tercer modelo para el análisis confirmatorio.

El constructo AMIIIE reducido está compuesto ahora por tres ítems, lo que no permite su validación. Con el fin de evitar el problema de identificabilidad, es necesario validarlos conjuntamente con otro constructo previamente validado, por ejemplo el constructo DIIE. En la Tabla 6, se presentan también los valores de los índices de ajuste de los modelos de las variables latentes reducidas DIIE (Figura 1), AMIIIE (Figura 2) y OMIIIE (Figura 3). Estos sugieren que el modelo propuesto es el mejor para definir las variables latentes DIIE, AMIIIE y OMIIIE. Las variables reducidas DIIE y AMIIIE tienen un buen ajuste: $\chi^2/gf < 0.20$; RMSEA < 0.05; CFI > 0.90; TLI > 0.90; $\rho_{vc} > 0.5$ y $\rho_{Jor} > 0.7$; mientras que la variable OMIIIE tiene un ajuste medianamente aceptable.

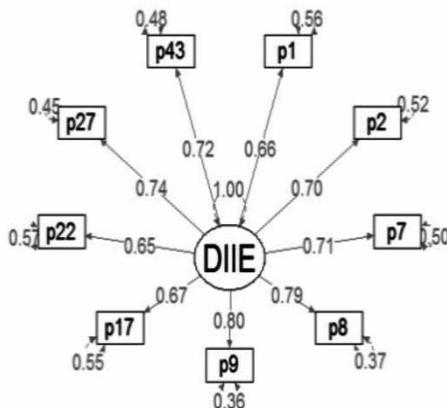


Figura 1. Modelo de medida del constructo reducido DIIE.

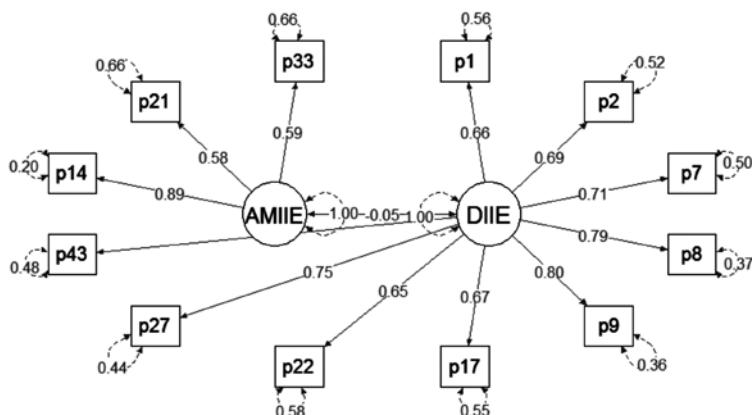


Figura 2. Modelo de medida del constructo reducido AMIIIE.

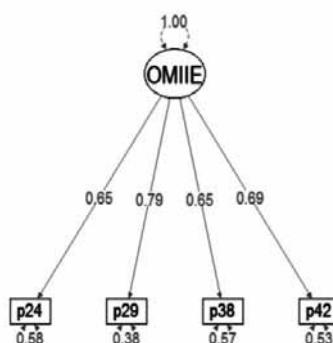


Figura 3. Modelo de medida constructo reducido OMIIE.

Se calculó la fiabilidad de los constructos con dos índices: el coeficiente de determinación (fiabilidad individual) y la fiabilidad compuesta (ρ ho de Jöreskog). Como indica la Tabla 6, la fiabilidad individual de cada uno de los ítems es significativa, dado que superan, en la mayoría de los casos, el valor de 0.50. En cuanto la fiabilidad compuesta, su valor para cada constructo es mayor que 0.70. Ello implica que cada conjunto de ítems explica una parte de variación del concepto con la medida del error. Finalmente, la validez convergente se comprobó calculando la varianza extraída para cada constructo. Para los constructos reducidos, se puede observar que todos los valores son mayores o próximos a 0.50; que resultan relativamente aceptables. En el modelo de conjunto de medidas para los tres constructos (Tabla 6 y Figura 4), a pesar de que el χ^2 es significativo con un valor p menor que 0.001, los índices $\chi^2 / gl = 1.54$; TLI=0.96; CFI=0.96 y RMSEA=0.055 indican un adecuado ajuste a los datos, pues siguiendo a Fornell y Larcker (1981), el índice RMSEA se considera bueno si es menor que 0.05 y aceptable si varía entre 0.05 y 0.08.

Tabla 7

Estadísticos de bondad de ajuste del tercer modelo propuesto

| Modelo | Nº items | χ^2/gl | CFDLI | RMSEA | CFI | TLI | VR^2 | ρ_{jor} | ρ_{vc} |
|------------------|----------|-------------|--------------|-------|-------|-------|-------------|--------------|-------------|
| Modelo propuesto | 16 | 1.54 | [0.61; 0.81] | 0.055 | 0.968 | 0.963 | [0.37;0.66] | 0.48 | 0.94 |

Para ver valores de referencia, véase Tabla 2.

La validez discriminante consiste en demostrar que dos medidas de constructos diferentes no deben estar relacionadas entre sí. Concretamente, se trata de verificar que la varianza de un constructo dado, está mejor explicado por sus ítems que por los ítems de otro constructo (Fornell y Larcker, 1981). Se comprobó la validez discriminante entre las variables que conforman el modelo de medida propuesto (Figura 4). Para

estos efectos, se considera como indicador de validez discriminante, cualquier valor mayor que uno; o sea, la razón, ρ_{vc}/ρ^2 , de la varianza extraída entre el constructo y sus ítems y la correlación o covarianza del mismo constructo con otro debe ser mayor que 1, incluso aunque la matriz no sea simétrica. La Tabla 8, presenta los resultados de la validez discriminante.

Tabla 8

Validez discriminante del modelo propuesto

| | DIIE | AMIIIE | OMIIIE |
|--------|------|--------|--------|
| DIIE | 0.51 | 0.0036 | 0.32 |
| AMIIIE | | 0.50 | 0.23 |
| OMIIIE | | | 0.49 |

La Tabla 8 muestra que la validez discriminante entre las variables latentes está garantizada ya que la varianza para cada variable latente es mayor que las correlaciones entre éstas, y por lo tanto los constructos son diferentes entre sí. La configuración final subyacente del modelo propuesto queda tal como se visualiza en la Figura 4:

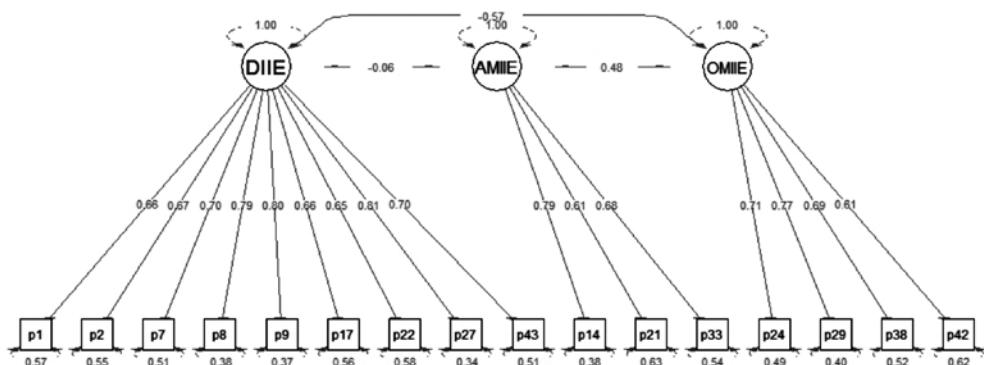


Figura 4. Modelo de medida del modelo propuesto (DIIE, AMIIIE y OMIIIE).

Se puede afirmar que el AFC para el tercer modelo propuesto dio resultados satisfactorios. La evaluación de fiabilidad de la consistencia interna dada por ρ de Jöreskog (ρ_{Jor}) presenta valores considerados muy positivos mayores que 0.70, para todos los constructos. En cuanto a la validez convergente, todas las contribuciones factoriales que unen los ítems de las variables latentes son significativas (W -estadístico > 10.82 con $p=0.001$). Además, los coeficientes ρ de la validez convergente ρ_{vc} son aceptables, valores cercanos a 0.5.

Por otra parte, y debido a los malos resultados del análisis confirmatorio para el modelo deducido del AFE, se intentó comprobar que la causa de estos resultados no

fuerá el modo en que se consideraron las variables para el AFC (variables ordinales). Los índices de bondad de ajuste para el modelo inferido del AFE pero con variables continuas se ofrecen en Tabla 9:

Tabla 9

Índices de bondad de ajuste de un modelo inferido del AFE, considerando variables continuas

| | Variable latente | Nº ítems | χ^2/gl | CFDLI | RMSEA | CFI | TLI | VR^2 | ρ_{vc} | ρ_{jor} |
|-------------------------------|------------------|----------|-------------|--------------|-------|------|------|-------------|-------------|--------------|
| Constructos deducidos del AFE | DIIE | 13 | 1,27 | [-0,45;0,79] | 0,039 | 0,98 | 0,97 | [0,20;0,62] | 0,47 | 0,89 |
| | AMIIIE | 8 | 2,63 | [0,32;0,68] | 0,096 | 0,84 | 0,77 | [0,10;0,46] | 0,31 | 0,75 |
| | OMIIIE | 4 | 0,29 | [0,55;0,73] | 0,000 | 1,00 | 1,04 | [0,30;0,53] | 0,40 | 0,70 |

Para ver valores de referencia, véase Tabla 2.

Los resultados muestran nuevamente, un ajuste poco satisfactorio del modelo arrojado por el AFE al modelo surgido del AFC con variables continuas. En las Figuras 5, 6 y 7 se muestran los modelos de medida deducidos de este análisis.

En efecto, los índices: χ^2/gl , RMSEA, CFI y TLI no superan el umbral de aceptación para el constructo AMIIIE (Figura 6), mientras que estos índices sí expresan un mejor ajuste para los constructos DIIE (Figura 5) y OMIIIE (Figura 7), exceptuando los índices para ρ de la varianza extraída y ρ de Jöreskog. Téngase presente que ρ de Jöreskog se considera como un mejor indicador de la consistencia interna (Gerbing y Anderson, 1988) y también representa una mejora sustancial, ya que no depende del número de ítems que constituyen los factores (Roehrich, 1993).

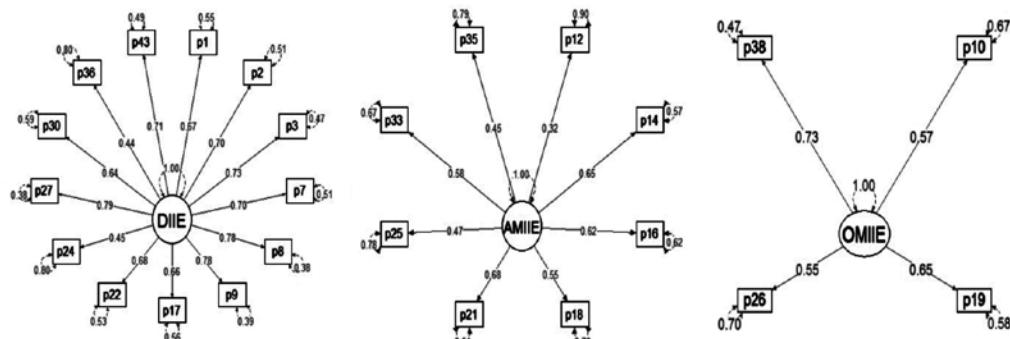


Figura 5. Modelo de medida del constructo DIIE deducido del AFE

Figura 6. Modelo de medida del constructo AMIIIE deducido del AFE

Figura 7. Modelo de medida del constructo OMIIIE deducido del AFE

Comparando el modelo deducido del análisis confirmatorio (modelo propuesto, con 16 ítems) con el modelo factorial exploratorio (modelo deducido del AFE, con 25 ítems), se constata que los tres constructos de este último, no se ajustan bien a los datos. Lo anterior se confirma al contrastar los resultados de la aplicación del AFC a los constructos deducidos del AFE. El análisis confirmatorio, además de las contribuciones factoriales que también son una salida del exploratorio, provee los valores de la estimación de los errores de medida y la bondad de ajuste del modelo. No considerar los errores de medida es en cierto grado, una limitación del análisis factorial exploratorio que lo hace no aplicable en algunas situaciones, como por ejemplo, cuando se usan escalas tipo Likert (Likert, 1932) para medir un fenómeno.

Discusión de los resultados

El objetivo principal de este trabajo fue realizar un estudio sobre la validez de constructo de la escala EI/IEPD. Se puede observar que el conjunto de sentencias y, por ende, el conjunto de dimensiones, muestran una buena fiabilidad entre 0.71 y 0.89; tomando como referencia a Ortega et al. (2008) que plantean que, aquellos valores alfa que se encuentran entre 0.61 y 0.80, son considerados aceptables. Para el caso de la escala EI/IE-PD, modelo deducido del AFE, el alfa de Cronbach fue de 0.77, superando a la versión extendida que fue validada por Díaz (2010). Ahora bien, al analizar este modelo con el análisis AFC, fue posible evidenciar que no muestra robustez con el método *Weighted Least Squares Mean and Variance Adjusted* (WLSMV). Por lo tanto, es preciso señalar que el alfa de Cronbach, por sí mismo, no es un indicador que permita concluir que se está frente a un buen instrumento; ya que ésta además es sensible al número de los ítems en cada constructo.

Otro hallazgo positivo conveniente de denotar es el bajo número de ítems que terminan siendo seleccionados para constituir el modelo final propuesto y aceptado (16 ítems); tal modelo es el que obtiene mejores ítems para el AFC, y representa a su vez un número muy reducido de ítems iniciales valorados por los expertos. Es evidente que, se debe ponderar que si un instrumento posee pocos ítems y goza de un modelo robusto, tendrá una más fácil aplicación.

Respecto a la validez discriminante del modelo propuesto, se puede afirmar que los constructos OMIIE y AMIIE presentan cierta dependencia según el análisis mostrado en la Tabla 6. Esto obliga a reconsiderar estos dos constructos como dimensiones no totalmente diferentes del impacto de la investigación educativa en la práctica docente. Esta dependencia podría tener relación con que las oportunidades, que mejoran el impacto de la investigación educativa, estarán siempre en correspondencia con las acciones para hacer efectivas esas oportunidades, lo que podría ayudar a interpretar los resultados de la dependencia observada en el análisis discriminante. Tras la realización del AFC al modelo propuesto, se pudo observar que esta división no se ajusta al modelo que mejor representa el constructo a medir, por lo que los ítems de AMIIE y OMIIE fueron reorganizados en la propuesta final de instrumento que se presente en este artículo (ver Anexo final).

Conclusiones

Se hace visible la dificultad de disponer de instrumentos validados que permitan conocer la evaluación del impacto de la investigación educativa en la práctica docente, pues la aproximación clásica de conjugar AFE y propiedades métricas no sería suficiente para definir un constructo consistente; se hace preciso AFC. Además, el impacto de la investigación educativa en la práctica docente se presenta como un complejo fenómeno multidimensional. Más investigación al respecto sería conveniente mejorando la muestra aquí disponible.

Se destaca la funcionalidad del AFC, ya puesta de manifiesto por Lavega, March y Filella (2013), para la validación de constructo de instrumentos de investigación cuantitativa, al mostrar aristas que un AFE no hace explícitas (análisis discriminante y convergente de los ítems), cuestión que se presenta en las diferencias de los índices de varianza extraída, R^2 , obteniendo valores más modestos de dichos índices para el modelo de AFE que para el modelo AFC, en el primer y tercer componente.

A partir del análisis métrico con AFE y AFC, mostrados anteriormente, se puede concluir que la escala EI/IE-PD (modelo reducido propuesto) se presenta como un instrumento aceptablemente válido para evaluar el impacto de la investigación educativa en la práctica docente.

Referencias

- Boza, A. y Méndez, J. (2013). Aprendizaje motivado en alumnos universitarios: Validación y resultados generales de una escala. *Revista de Investigación Educativa*, 31(2), 331-345.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. Nueva York: The Guilford Press.
- Buendía, L., Colás, P. y Hernández-Pina, P. (1998). *Métodos de investigación en psicopedagogía*. Madrid: McGraw-Hill.
- Burkhardt, H. y Schoenfeld, A. H. (2003). Improving educational research: Toward a more useful, more influential, and better-funded enterprise. *Educational Researcher*, 32(9), 3-14.
- Díaz, E. (2010). *Validación de la escala de Evaluación del Impacto de la Investigación Educativa en la Práctica Docente (EI/IE-PD)* (Tesis doctoral). Granada: Universidad de Granada. Disponible en <http://digibug.ugr.es/bitstream/10481/4873/1/18656158.pdf>.
- Escudero, J. (2006). Compartir propósitos y responsabilidades para una mejora democrática de la educación. *Revista de Educación (Madrid)*, 339, 19-41.
- Farrell, A. M. (2010). Insufficient discriminant validity: A comment on Bove, Oervan, Beatty and Shiu (2009). *Journal of Business Research*, 63(3), 324-327.
- Fernández-Cano, A. (2001). Valoración del impacto de la investigación educativa sobre la práctica docente. *Revista de Educación (Madrid)*, 324, 155-170.
- Fernández-Cano, A., Díaz, E., Torralbo, M. y Ruiz, M. (2008). Inquiring into the impact of educational research on teaching practices. A rating scale for evaluating it. Presentada a *European Conference on Educational Research (ECER-2008). From Teaching to Learning?* (p. 49). Göteborg: EERA.

- Fornell, C. y Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error: Algebra and statistics. *Journal of Marketing Research*, 18, 39-50.
- Gerbing D.W. y Anderson J.C. (1988). An update paradigm for scale development incorporating unidimensionality and its assessment. *Journal of Marketing Research*, 25(2), 186-1992.
- Kaestle, C.F. (1993). The awful reputation of education research. *Educational Researcher*, 22(1), 23+26-31.
- Larwin, K. y Harvey, M. (2012). A demonstration of a systematic item-reduction approach using structural equation modeling. *Practical Assessment, Research & Evaluation*, 17(8). Disponible en: <http://pareonline.net/pdf/v17n8.pdf>
- Lavega, P., March, J. y Filella, G. (2013). Juegos deportivos y emociones. Propiedades psicométricas de la escala GES para ser aplicada en la educación física y el deporte. *Revista de Investigación Educativa*, 31(1), 151-165.
- Lévy-Mangin, J. P. y Varela, J. (2006). *Modelización con estructuras de covarianzas en ciencias sociales*. La Coruña: Netbiblo.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 140, 1-55.
- Martín, M. C. (2004). Diseño y validación de cuestionarios. *Matronas Profesión*, 5(17), 23-29.
- Muthén, B. (1993). Goodness of fit with categorical and other non-normal variables. En K. A. Bollen, y J. S. Long (Eds.), *Testing structural equation models* (pp. 205-243). Newbury Park, CA: Sage.
- Ortega, E., Calderón, A., Palao, J. M. y Puigcerver, C. (2008). Diseño y validación de un cuestionario para evaluar la actitud percibida del profesor en clase y de un cuestionario para evaluar los contenidos actitudinales de los alumnos durante las clases de educación física en secundaria. *Retos*, 14, 22-29.
- Outón, P. y Suárez, A. (2010). Adaptación y validación del test de dislexia Bangor. *Revista Investigación Educativa*, 28(2), 445-457.
- Roehrich, G. (1993). Validité convergente et validité discriminante: L'apport des modèles d'équations structurelles. *ESA Grenoble, Cahiers de Recherches*: CERAG, 93-23.
- Schneider, B. y Keesler, V. (2007). School reform 2007: Transforming education into a scientific enterprise. *Annual Review of Sociology*, 33, 197-217.
- Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. Washington: American Psychological Association.
- Vanderlinde, R. y van Braak, J. (2010). The gap between educational research and practice: Views of teachers, school leaders, intermediaries and researchers. *British Educational Research Journal*, 36(2), 299-316.

Anexo
MODELO VALIDADO Y REDUCIDO DE ESCALA EI/IE-PD

| Ítems | MD | D | IND | A | MA |
|---|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| 1. Investigadores y docentes tienen intereses y motivaciones profesionales diferentes | <input type="checkbox"/> |
| 2. Las investigaciones son realizadas por aquellos que desconocen la práctica cotidiana en el aula | <input type="checkbox"/> |
| 7. Los/as profesores/as consideran que la Investigación Educativa no resuelve sus problemas en el aula | <input type="checkbox"/> |
| 8. La Investigación Educativa propone soluciones poco aplicables para la práctica educativa | <input type="checkbox"/> |
| 9. Los temas de la Investigación Educativa son ajenos a los problemas del profesor/a en el aula | <input type="checkbox"/> |
| 14. El lenguaje de los informes de Investigación Educativa es demasiado técnico para la comprensión de los docentes | <input type="checkbox"/> |
| 17. El desprecio de la Investigación Educativa es cada vez más generalizado entre los docentes | <input type="checkbox"/> |
| 21. La utilización de un lenguaje preciso y riguroso es un obstáculo para la comprensión de informes de Investigación Educativa | <input type="checkbox"/> |
| 22. Hay una pérdida de confianza de los/as profesores/as en la utilidad de la Investigación Educativa | <input type="checkbox"/> |
| 24. El/la profesor/a podría mejorar su práctica a partir de la información contenida en las Investigaciones Educativas | <input type="checkbox"/> |
| 27. La contribución de la Investigación Educativa a la mejora de la práctica educativa ha sido nula | <input type="checkbox"/> |
| 29. A largo plazo, la Investigación Educativa puede mejorar la práctica del profesor/a | <input type="checkbox"/> |
| 33. Los docentes no poseen formación en Investigación Educativa que les permita utilizar en la práctica los hallazgos de la misma | <input type="checkbox"/> |
| 38. La investigación puede ser una tarea más del profesor/a | <input type="checkbox"/> |
| 42. La participación del profesor/a en equipos de investigación mejora la práctica docente | <input type="checkbox"/> |
| 43. CONCLUSIÓN: la Investigación Educativa no tiene impacto sobre la mejora de la docencia diaria del profesor/a | <input type="checkbox"/> |

*MD= Muy Desacuerdo, D= Desacuerdo, IND= Indiferente, A= Acuerdo y MA= Muy Acuerdo.

Se respeta el orden de presentación de la escala amplia original.

Fecha de recepción: 17 de septiembre de 2013.

Fecha de revisión: 19 de septiembre de 2013.

Fecha de aceptación: 6 de mayo de 2014.

