

Meta-análisis de generalización de la fiabilidad del cuestionario FIT-Choice (Factores que influyen en la elección de la enseñanza como carrera)¹

Meta-analysis of reliability generalisation of the FIT-Choice questionnaire (Factors Influencing Teaching Choice)

DOI: 10.4438/1988-592X-RE-2021-393-492

Enrique Navarro-Asencio

Universidad Complutense de Madrid

Esther López-Martín

Universidad Nacional de Educación a Distancia

Inmaculada Asensio-Muñoz

Universidad Complutense de Madrid

Eva Expósito-Casas

Universidad Nacional de Educación a Distancia

Elvira Carpintero-Molina

Covadonga Ruiz De Miguel

Universidad Complutense de Madrid

Resumen

La eficacia docente depende de factores académicos y no académicos. Entre los segundos, la motivación es una característica muy estudiada para determinar quién accede a los estudios de maestro. En este contexto, el instrumento FIT-Choice (*Factors Influencing Teaching Choice*) es de los más utilizados a

⁽¹⁾ Financiación: Proyecto nº RTI2018-099365-B-100 financiado por el Ministerio de Ciencia e Innovación y la Unión Europea en el Programa Estatal de I+D+i Orientada a los Retos de la Sociedad en la convocatoria de 2018.

nivel mundial y cuenta con diversas adaptaciones de idioma y de estructura dimensional. El objetivo de este trabajo es analizar la fiabilidad de las distintas escalas empleadas y su variabilidad entre los trabajos que la usan, mediante la técnica de meta-análisis de generalización de la fiabilidad. Concretamente se propone un modelo de efectos aleatorios para estimar la fiabilidad media y se estudia la heterogeneidad de los coeficientes con los estadísticos Q de Cochran y I^2 . También se emplea la meta-regresión para conocer el efecto de variables moderadoras relacionadas con el instrumento (tipo de FIT-Choice, idioma de la prueba) y de la muestra (tipo de docente, % de mujeres y edad). Los resultados muestran una buena fiabilidad promedio en la mayor parte de las subescalas, excepto en tres, movilidad en el trabajo, elección de la carrera como opción alternativa y dificultad percibida, con valores por debajo de 0.7. Destaca la gran variabilidad de la fiabilidad entre estudios, siendo la escala de satisfacción la que presenta mayor heterogeneidad. Los hallazgos muestran cómo la variación de la fiabilidad de algunas escalas está vinculada a las características del FIT-Choice utilizado y al tipo de docente.

Palabras clave: motivaciones docentes, cuestionario FIT-Choice, meta-análisis, fiabilidad

Abstract

Teacher effectiveness depends on academic and non-academic factors. Teaching motivation is an element of the latter group of factors that has been widely studied to determine who will enter the teaching profession. In this context, FIT-Choice (Factors Influencing Teaching Choice) is one of the most widely used instruments internationally and has undergone both language and dimensional structure adaptations. The aim of this study is to analyse the reliability of the different scales and how they vary in works that use this instrument, through the reliability generalisation meta-analysis technique using a random effects model. Specifically, we use a random effects model to estimate mean reliability and we study the heterogeneity of coefficients using Cochran's Q and I^2 tests. Meta-regression is also used to study the effect of moderating variables that relate to the instrument (type of FIT-Choice, test language) and the sample (type of teacher, % of women and age). The results show good average reliability for most of the scales, apart from job transferability, fallback career and high demand, which have values below 0.7. High reliability variance between studies stands out, with the satisfaction scale displaying the greatest heterogeneity. The findings show that the reliability of subscales is affected by the type of FIT-Choice and the type of teacher.

Key words: teacher motivations, FIT-Choice instrument, meta-analysis, reliability

Introducción

La investigación ha demostrado que lo mejor que las escuelas pueden hacer por sus estudiantes es darles buenos maestros, ya que recibir docencia de buenos docentes puede situar incluso a los estudiantes más desfavorecidos en el camino a la universidad y, al revés, años con maestros ineficaces pueden provocar en los estudiantes un fracaso académico difícil de recuperar (Jordan et al., 1997; Schleicher, 2018). El papel fundamental de los maestros en la calidad de la educación que reciben sus alumnos ha sido evidenciado en gran cantidad de estudios empíricos. Hattie (2008 y 2011), que meta-analiza más de 65.000 trabajos de investigación sobre los efectos de cientos de intervenciones en el aprendizaje de 250 millones de alumnos, concluye que aquellos factores a los que se suele dar importancia, como el tamaño de las clases o el nivel de inversión, producen en realidad pocos efectos sobre el aprendizaje de los alumnos. Sin embargo, la calidad del maestro surge como el factor más determinante, una vez que se han controlado otros efectos contextuales, como la procedencia de los alumnos, y un incremento en dicha calidad tiene un impacto más efectivo que cualquier otra inversión educativa, incluso que las reducciones en el tamaño de la clase (Goldhaber, 2010). El conocido *informe McKinsey* (Barber y Mourshed, 2007) también identifica a los docentes como factor clave en los sistemas educativos con mejores resultados en las evaluaciones internacionales, hasta el punto de que un sistema educativo no puede nunca ser mejor que la calidad de sus maestros.

En este contexto, se justifica el interés de la investigación por conocer qué características de los maestros, o aspirantes a serlo, están más relacionadas con su eficacia. Una muestra de este interés se concreta en el *Teaching and Learning International Survey* (TALIS) (INEE, 2020) o el *TALIS Initial Teacher Preparation study* (OECD, 2017), que pone el foco en la importancia de la selección de los candidatos más adecuados para los programas de iniciación docente (*Initial Teacher Education* -ITE-). En esta línea, destacan las políticas de países como Inglaterra y Gales, donde Klassen y Dolan (2015), tras revisar 74 programas ITE, confirman que todos utilizan algún procedimiento de selección de candidatos, basándose en características cognitivas y no cognitivas de los mismos, o como Australia, donde su Instituto de Enseñanza y Liderazgo Escolar (AITSL) establece para los aspirantes, entre otros estándares, un nivel de

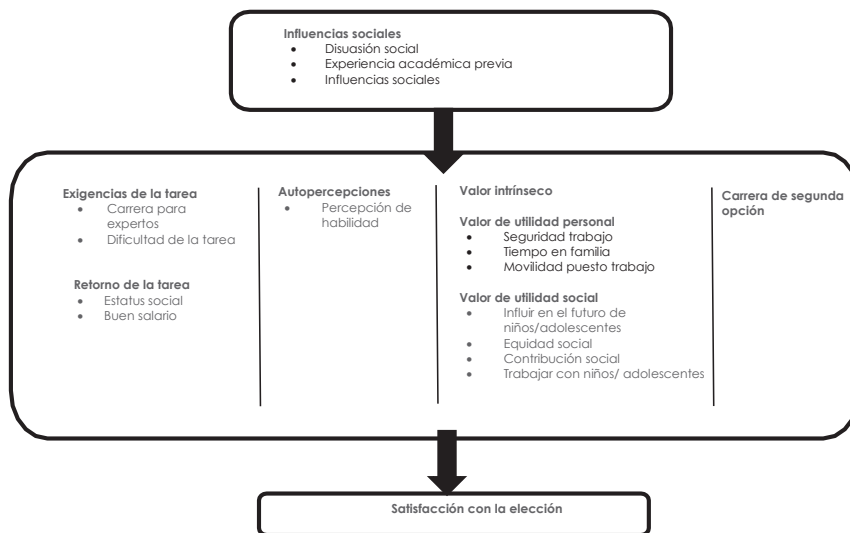
dominio de la lengua y las matemáticas equivalente al 30% superior de la población (AITSL, 2011).

La acumulación de investigación primaria sobre el tema ha llevado a realizar investigación secundaria relevante, como el meta-análisis de Klassen y Kim (2017), que identifica un efecto pequeño, pero significativo, de la eficacia docente referida a un conjunto de atributos académicos de la persona (conocimiento disciplinar, habilidades cognitivas y conocimiento pedagógico) y no académicos (personalidad, motivación, creencias y disposiciones) que, a su vez, interactúan con factores contextuales (culturales, sociales y de experiencia académica previa). También es destacable la revisión sistemática de Coe et al. (2014), que analiza 200 trabajos y llega a la identificación de seis componentes de lo que denominan “docencia magnífica” (*great teaching*), que son el conocimiento pedagógico y la calidad de la formación, principalmente, junto a la gestión del aula y su clima, las creencias sobre la enseñanza y el aprendizaje y elementos profesionales más generales, como la formación permanente o las relaciones con compañeros. En cuanto a la importancia de los factores no cognitivos, hay que citar el meta-análisis de Brookhart y Freeman (1992), que incluyó trabajos en Estados Unidos entre 1960 y 1990, cuya síntesis sugiere que el altruismo y motivaciones intrínsecas, junto con la orientación al servicio eran las principales razones por las que se elige la carrera docente. El trabajo de Heinz (2015) continúa con el análisis de las investigaciones sobre factores no cognitivos, revisando en profundidad trabajos desde 1990 hasta 2014. Esta autora realiza una revisión sistemática de 41 estudios, de los que en 23 se concluye que las razones que llevan a la elección de la docencia como profesión se agrupan en los mismos factores: motivaciones intrínsecas, extrínsecas y altruistas. Por último, en la revisión realizada por Fray y Gore (2018) sobre estudios empíricos publicados entre 2007 y 2016, se concluye que en la mayoría de ellos (63 de 70) las motivaciones docentes intrínsecas y altruistas tienen más peso que las extrínsecas.

Un modelo de medida de los factores implicados en la elección de la profesión docente que incluye todos estos componentes es el propuesto por Helen Watt y Paul Richardson, que vincula la decisión de elegir la carrera docente con la teoría del valor de las expectativas (Eccles, 2005). Así, en la elección intervienen, por un lado, el valor que se da a la profesión, que depende del disfrute o de la utilidad percibida, y, por otro, las expectativas de éxito, que dependen de las creencias sobre

la profesión y de la percepción de autoeficacia en la realización de la tarea, ambas relacionadas a su vez con las experiencias previas y su interpretación (Watt et al., 2012). En la figura I se representa el modelo, con una dimensión de antecedentes académicos e influencias sociales que pueden determinar las creencias sobre la profesión, la percepción de autoeficacia, las motivaciones y la posibilidad de haber elegido una profesión alternativa. El modelo distingue entre motivaciones intrínsecas, de utilidad personal y de valor social. También incluye un factor de percepciones sobre la profesión relativas a la dificultad y exigencia profesional así como a los posibles beneficios (salario, estatus, etc.). Finalmente, contempla un factor sobre la habilidad docente autopercebida. Este modelo es el sustento teórico del instrumento de medida de la motivación denominado FIT-Choice (*Factors Influencing Teaching Choice*).

FIGURA I. Modelo teórico FIT-Choice.



Fuente: Adaptado de Watt y Richardson (2007)

El FIT-Choice presenta un total de 18 subescalas agrupadas en cuatro factores de segundo orden: valor de utilidad personal, valor de utilidad

social, exigencia de la profesión y beneficios de retorno de la tarea, referidos los dos primeros a aspectos motivacionales y los segundos a las creencias sobre la profesión. A su vez, las escalas de antecedentes también hacen referencia a las motivaciones (experiencias previas de enseñanza-aprendizaje e influencias sociales) y a las creencias (disuasión social y satisfacción con la elección).

Las revisiones sistemáticas realizadas hasta el momento justifican el interés por el estudio de este instrumento, destacando la extensión de su uso a nivel internacional: el FIT-Choice se emplea en 10 de los 41 artículos incluidos en la revisión de Heinz (2015) y en 17 de los 70 trabajos revisados por Fray y Gore (2018). Entre los países en que se ha utilizado se encuentra España, donde ha sido validado por Gratacós y López-Jurado (2016). La aplicación del cuestionario a poblaciones muy diversas ha provocado no solo adaptaciones asociadas a la traducción a otros idiomas sino también ajustes en la estructura, ítems y dimensiones originales, lo que demanda un estudio global de sus características psicométricas. En esta línea, el presente trabajo se centra en la fiabilidad general del FIT-Choice y persigue sintetizar de forma cuantitativa los resultados de fiabilidad obtenidos en investigaciones educativas que han aplicado este instrumento además de analizar los factores asociados con ella. Como objetivos específicos se pretende, por tanto:

- Describir las características del instrumento y la muestra que han utilizado los diferentes estudios.
- Estimar la fiabilidad media de las escalas del instrumento entre los distintos estudios
- Identificar el efecto de posibles factores asociados con la variabilidad de la fiabilidad entre investigaciones (tipo de adaptación del instrumento, tipo de docente evaluado, edad promedio o porcentaje de mujeres en la muestra)

Método

Para lograr los objetivos propuestos se emplea la técnica de meta-análisis de generalización de la fiabilidad (GF en adelante) (Vacha-Haase, 1998). En términos generales, la fiabilidad en la teoría clásica de los test (TCT) se define como la cantidad de varianza de las respuestas que no se debe

a errores de medida, siendo el coeficiente alfa de Cronbach uno de los procedimientos más utilizados para su estimación.

El meta-análisis de GF está creciendo en importancia y es una forma de revisión sistemática que supera los problemas de medición presentes en los estudios primarios (Greco et al., 2018). Graham et al. (2011) señalan tres tipos de aplicaciones de este tipo de meta-análisis: 1) estimar el promedio en la puntuación de fiabilidad de las medidas utilizadas entre diferentes estudios, 2) estudiar los factores que afectan a la variabilidad de la fiabilidad entre los distintos estudios, y 3) dar respuesta a la pregunta ¿cómo son de fiables las puntuaciones obtenidas en estudios con diferentes muestras y características?

En esta técnica, se parte de los datos de fiabilidad de los diferentes estudios para calcular la tendencia central y la variabilidad de los índices, que se analizan con una metodología específica, ya que, como apuntan Rodríguez y Maeda (2006), la distribución de puntuaciones de los coeficientes de fiabilidad no tiene las mismas propiedades que la de los tamaños del efecto empleados en los meta-análisis tradicionales. Aunque existen diferentes estrategias de análisis, en este trabajo se opta por un modelo de efectos aleatorios con una transformación ponderada de los coeficientes alfa, incluyendo el dato de variabilidad para considerar el posible efecto de la heterogeneidad de la muestra en la estimación de la fiabilidad. Este modelo de efectos aleatorios es recomendado por Botella et al. (2010) y Sánchez-Meca et al. (2013) por su mayor capacidad para generalizar las conclusiones y porque es la forma más correcta de hacer inferencias sobre la fiabilidad, al permitir estimar los efectos de diferentes fuentes de variación.

Estrategia de búsqueda y selección de estudios

La búsqueda de los artículos se llevó a cabo el 6 de julio de 2020 en las siguientes bases de datos internacionales: colección principal de la *Web of Science* (WoS), base de datos del *Education Resource Information Center* (ERIC), *Scopus*, *Google Scholar* y *Dialnet*. Los términos de búsqueda utilizados fueron FIT-Choice y *Factors Influencing Teaching Choice* y no se estableció límite temporal. La búsqueda se efectuó en el campo “por defecto” de ERIC y Dialnet, en el campo “tema” de la WoS, en el campo “*title, abstract y keywords*” de *Scopus* y en el campo “*title*” de Google

Académico. Esta última base de datos solo permite acotar la búsqueda al campo título, pero no al campo título, resumen y palabras clave o campos por defecto. La ecuación de búsqueda utilizada fue la siguiente²:

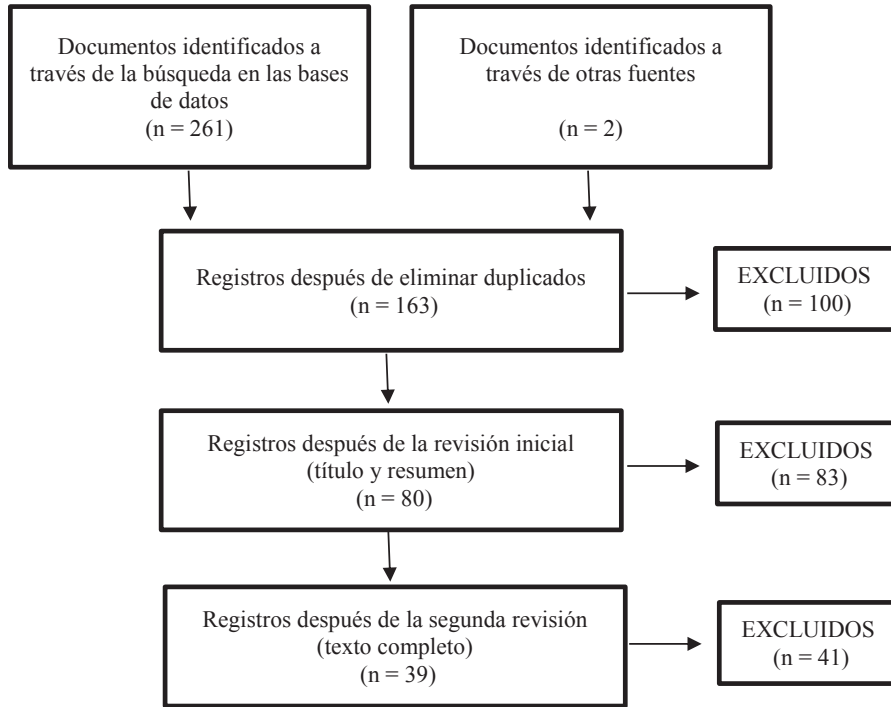
*(“fit choice” OR “fit*choice” OR “factors influencing teaching choice”)*

A su vez, se realizaron consultas en los principales repositorios internacionales de tesis doctorales: DART-Europe y *Open Access Theses and Dissertations*.

La búsqueda de la literatura permitió recuperar 261 documentos: 62 registros de la WoS, 50 registros de ERIC, 79 registros de *Scopus*, 5 registros de Dialnet, 53 registros de *Google Scholar*, 1 registro de DART-Europe y 11 registros de *Open Acces Theses and Dissertation*. También se recuperaron dos artículos de otras fuentes. Tras eliminar 100 documentos duplicados, se llevó a cabo una revisión inicial de los 163 trabajos restantes basada en el título y el resumen. Esta revisión permitió excluir 83 documentos que no analizaban las propiedades psicométricas del FIT-Choice (temática diferente o estudios teóricos) o estaban publicados en un idioma diferente al inglés o el español. Esta selección inicial fue llevada a cabo por dos investigadores independientes, alcanzándose un índice de acuerdo del 94%. Los 80 trabajos seleccionados en esta primera fase fueron sometidos a una evaluación basada en la lectura del texto completo que permitió descartar 41 documentos, principalmente por no facilitar información sobre la fiabilidad del FIT-Choice o remitir a los valores reportados por otros trabajos que se habían incluido previamente. La Figura II resume el proceso descrito.

⁽²⁾ En la búsqueda en Google Scholar se eliminó el truncamiento para evitar recuperar documentos irrelevantes.

FIGURA II. Diagrama de flujo del proceso de selección de los estudio



Atendiendo a lo anterior, en total se han analizado 39 estudios que, en su conjunto, proporcionan información sobre las propiedades psicométricas del FIT-Choice tras su aplicación a 48 muestras de sujetos diferentes. Los criterios utilizados en la codificación de estos estudios se recogen en la Tabla I.

Codificación

Además de la información descriptiva de los estudios y la muestra, la caracterización del instrumento FIT-Choice se lleva a cabo considerando, en primer lugar, el tipo de escala utilizada, diferenciándose entre la aplicación del instrumento original validado por Watt y Richardson (2007) o una modificación del original, ya sea por reducción de la escala,

por cambios en el constructo o por la adaptación a otro idioma. A su vez, se especifica si el estudio valida la escala como propósito principal del trabajo o como estudio previo para garantizar su uso, registrando si no lo hacen, lo hacen solo mediante un estudio de la fiabilidad o aplican además técnicas factoriales exploratorias y confirmatorias. En el caso de los estudios que emplean traducciones del instrumento, también se indica de qué idioma se trata. Finalmente, se incluye la información necesaria para poder realizar el resumen cuantitativo de la fiabilidad: a) alfa de Cronbach (en dos estudios eran datos de la fiabilidad compuesta obtenida con las puntuaciones factoriales en un estudio confirmatorio), b) el número de ítem de cada escala y c) la desviación típica de la muestra.

En los trabajos que adaptan el modelo, algunos proponen una nueva estructura organizativa de dimensiones. En estos casos, se añade el nombre del nuevo factor junto a la información anterior.

TABLA I. Sistema de codificación de los artículos

	Variable	Etiquetas	Tipo
1	Identificador		Cadena
2	Autor/es		Cadena
3	Año de publicación		Númérica
4	Título del documento		Cadena
5	Tipo de publicación	1 - Artículo 2 - Libro 3 - Tesis doctoral 4 - Informe de investigación 5 - Otros	Númérica
6	Título de publicación		Cadena
7	Idioma de la publicación		Cadena
8	Palabras clave		Cadena
9	FIT-Choice (tipo de escala)	1 - Original (Watt y Richardson, 2007) 2 - Escala que adapta la dimensionalidad del constructo 3 - Escala traducida 13 - Original y traducida 23 - Escala adaptada (traducción + dimensionalidad)	Númérica
10	Estudio de validación	1- No 2- Sí, solo fiabilidad 3- Sí, con técnicas factoriales exploratorias 4- Sí, con técnicas factoriales confirmatorias	Númérica
11	Idioma FIT-Choice		Cadena

12	Tipo de docente	1 - Educación Infantil 2 - Educación Primaria 3 - Educación Secundaria 4 - Profesor Idiomas 12 - Educación Infantil y Primaria 13 - Educación Infantil y Educación Secundaria 23 - Educación Primaria y Secundaria 123 - Educación Infantil, Primaria y Secundaria	Numérica
13	País	En estudios que comparan varios países se incluye cada uno como una muestra	Cadena
14	Tamaño de la muestra	Número de casos totales	Numérica
15	N mujeres	Frecuencia o porcentaje	Numérica
16	Edad	Edad promedio de la muestra	Numérica
17	Actividad docente	1 - Docente en formación (<i>pre-service</i>) 2 - Docente en ejercicio	Numérica

La selección final recoge información de más de 29640 docentes (7.7%) y futuros docentes (92.3%), cuya edad media es de 26.18 años y de los que el 66% son mujeres. Estas investigaciones se han publicado en el periodo temporal 2007-2020, siendo 2012 y 2019 los años que más trabajos concentran (37.5% y 16.7% respectivamente). La mayoría de los documentos son artículos (87.5%), seguidos de tesis doctorales (8.3%) y libros (4.2%). Más del 93% de los trabajos (N=45) están escritos en inglés, frente a un 6% en español (N=3). En su mayor parte, son estudios realizados en EEUU (18.8%), Suiza (12.5%) y Turquía (10.4%).

Plan de análisis de datos

En primer lugar, como no es posible asumir la normalidad de la distribución de los coeficientes alfa entre los distintos estudios, se aplica la transformación propuesta por Bonett (2010):

$$L_i = \ln(1 - \alpha_i) \quad (1)$$

Donde es el coeficiente de fiabilidad de cada estudio i . La varianza de esta transformación se estima con la fórmula 2:

$$Var(L_i) = \frac{2q_i}{(q_i-1)(n_i-2)} \quad (2)$$

Donde q es el número de ítems de la escala en el estudio i y n el tamaño de la muestra empleada. A continuación, los resultados transformados se ponderan por la inversa de la varianza que en un modelo de efectos aleatorios tiene dos términos: $\text{Var}(L_i) \tau^2$ y, esto es, la variabilidad intra y entre estudios. Como la varianza entre estudios es un parámetro desconocido, se estima a partir de la información de fiabilidad proporcionada por las investigaciones analizadas con el método de Dersimonian-Laird (Sánchez-Meca et al., 2013). Finalmente, una vez ponderados, para facilitar su interpretación como alfas, se revierte la transformación.

La homogeneidad se contrasta con el estadístico Q de Cochran (Hedges y Olkin, 1985) y la proporción de heterogeneidad se calcula con I^2 (Higgins y Thompson, 2002), que por encima de 0.75 se considera alta.

El efecto de las variables moderadoras se calcula utilizando meta-regresión. Finalmente, el sesgo de la publicación se estudia con el método de regresión de Egger, donde valores no significativos indican ausencia de sesgo.

Variables moderadoras

Para el estudio del efecto de variables contextuales sobre la variabilidad de los coeficientes de fiabilidad se utilizaron como moderadoras las siguientes variables de la tabla I: 9, 10 y 11 (relativas a la adaptación del instrumento) y 12, 15, 16 y 17 (relativas a la muestra). Con el propósito de comprobar los efectos independientes de los grupos de casos representados en las distintas variables moderadoras, se dicotomizaron de modo que pudieran incluirse como predictores separados en el estudio de regresión. Todos los análisis se realizaron con el software *Jamovi* (2020) y el paquete *metafor* (Viechtbauer, 2010) de R.

Resultados

Para dar respuesta al primer objetivo específico, los estudios seleccionados y sus características se sintetizan en la tabla II

TABLA II. Estudios finalmente incluidos tras el proceso de revisión sistemática

ID	Estudio	n° de muestras	Tipo de escala	Estudio de validación	Tipo de docente	Actividad docente	Idioma Fit-Choice
43	Alpaslan et al. (2018)	1	2	1	3	1	Inglés
61	Akpochafo (2020)	1	23	1	23	1	Turco
21	Berger y D'Ascoli (2012a)	2	3	4	3	2	Alemán y francés
105	Berger y D'Ascoli (2012b)	2	23	1	3	1	Alemán y francés
146	Berger y Girardet (2020)	1	3	4	3	2	Inglés
81	Choi S., y Song J. (2015)	1	23	1	23	1	Inglés
125	de Zordo et al. (2019)	1	3	1	12	1	Inglés
59	Eren y Tezel (2010)	1	3	4	4	1	Turco
13	Fokkens-Bruinsma y Canrinus (2012a).	1	23	4	2	1	Holandés
133	Fokkens-Bruinsma y Canrinus (2012b).	1	3	2	3	1	Holandés
86	Fokkens-Bruinsma y Canrinus (2014)	2	3	1	23	1	Holandés
63	Goller et al. (2019)	2	3	4	12	1	Alemán y finlandés
143	Gratacós y López-Jurado (2016)	1	3	4	12	1	Español
23	Heinz et al. (2017)	1	1	1	3	1	Inglés
2	Hennessy y Lynch (2017)	1	2	3	3	1	Inglés
46	Jablanovic y Vracar (2019)	1	1	1	123	1	Inglés
85	Jugovic et al. (2012)	1	3	4	2	1	Croata

58	Kilinc et al. (2012)	1	3	4	23	1	Turco
102	Konig y Rothland (2012)	1	3	4	23	1	Alemán
51	Lawver (2009).	1	1	1	3	1	Inglés
15	Lawver (2012)	1	1	1	3	1	Inglés
76	Lin et al. (2012)	2	13	4	23	1	Inglés y chino
26	MacKenzie (2013)	1	1	1	23	1	Inglés
91	Nesje et al. (2018)	1	3	4	23	1	Noruego
164	Ozturk-Akar (2012)	1	3	4	3	1	Turco
14	Ozturk-Akar (2019)	1	23	1	3	1	Turco
96	Ponnock (2018)	1	2	4	123	1	Inglés
80	Ridgewell (2018)	1	1	4	2	2	Inglés
48	Said-Hung et al. (2018)	1	3	1	23	1	Español
3	Silvestre et al. (2020)	1	3	4	123	1	Español
126	Suryani et al. (2016)	1	23	4	123	1	Indonesio
128	Taimalu et al. (2017)	1	23	4	123	1	Estonio
165	Topkaya y Uz-tosun (2012)	1	2	4	4	1	Inglés
138	Torsney et al. (2019)	1	2	4	2	1	Inglés
94	Van Overschelde y Garza (2019)	1	2	4	3	1	Inglés
75	Wang (2019)	1	1	4	23	1	Inglés
97	Watt y Richardson (2007)	2	13	4	23	1	Inglés y alemán
101	Watt et al. (2012)	4	1	4	123	1	Inglés
154	Zhang et al. (2020)	1	23	4	4	1	Chino

La mayor parte de las investigaciones realizan el análisis en una única muestra (82%), el resto comparan los resultados de 2 grupos (15.4%) o,

en un estudio, comparan cuatro muestras distintas. En relación con las etapas educativas, en el 30.8% de los estudios participan docentes de Educación Secundaria, en el 28.2% docentes de Educación Primaria y Secundaria, en el 15.4% profesores de las tres etapas (Infantil, Primaria y Secundaria) y en el 10.3% docentes de Educación Primaria. Unido a ello en torno al 7.7% hacen referencia a muestras mixtas de Infantil y Primaria y un 7.7% a profesores de idiomas.

Los textos utilizan muestras compuestas de forma mayoritaria por docentes en formación (92.3%), solo tres artículos (7.7%) incluyen al profesorado en activo.

En el 20.5% de las investigaciones se emplea la escala Fit-Choice en su versión original (Watt y Richardson, 2007) y otro 5% utiliza la versión original junto con una traducción, frente a un 15.4% en los que se realiza una adaptación de la dimensionalidad del constructo, un 38.4% en los que se adapta el idioma o un 20.5% en los que se adaptan tanto dimensionalidad como idioma. El 66.6% realiza algún tipo de estudio psicométrico, la mayor parte (61.5%) aplica técnicas factoriales confirmatorias. Las investigaciones seleccionadas recogen el uso de la escala en 12 idiomas diferentes, siendo el inglés (41.7%), el alemán (10.4%), el turco (10.4%) y el estonio (10.4%) los más frecuentes.

En lo que se refiere al **segundo objetivo** específico, en la tabla III se presentan los resultados del meta-análisis para las 18 subdimensiones, los cuatro factores de segundo orden y los constructos de motivación y creencias que corresponden a la estructura original del FIT-Choice.

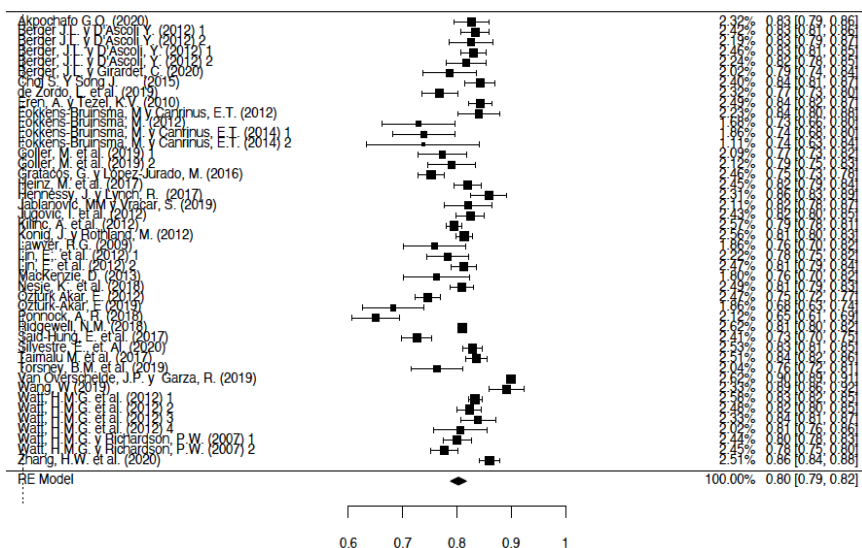
TABLA III. Resultados del meta-análisis de generalización de la fiabilidad por subdimensiones, dimensiones y constructos del FIT-Choice

	Fiabilidad					Heterogeneidad				Egger's Regression		
	k	F	E.T	Z	P	LI	LS	I ²	Q	P	Valor	P
M1 Percepción de habilidad	40	0.814	0.006	129	<.001	0.801	0.826	85.73%	273.256	<.001	-1.783	0.075
M2 Seguridad trabajo	35	0.829	0.008	98.9	<.001	0.813	0.846	96.09%	870.524	<.001	-0.293	0.77
M3 Tiempo en familia	35	0.83	0.007	125	<.001	0.817	0.843	91.31%	391.334	<.001	-0.508	0.611
M4 Movilidad en el puesto de trabajo	22	0.674	0.02	34.1	<.001	0.636	0.713	94.13%	357.953	<.001	-0.462	0.644
M5 Valor intrínseco	39	0.741	0.015	48.4	<.001	0.711	0.771	97.06%	1293.724	<.001	-2.14	0.032
M6 Carrera de segunda opción	32	0.654	0.018	36.2	<.001	0.618	0.689	90.89%	340.289	<.001	-0.624	0.533
Utilidad Personal (2° orden)	11	0.817	0.019	42.5	<.001	0.779	0.855	96.18%	261.71	<.001	0.82	0.412
M7 Influir en futuro	24	0.788	0.014	56.7	<.001	0.76	0.815	97.84%	1062.58	<.001	0.345	0.73
M8 Equidad Social	33	0.814	0.01	83.3	<.001	0.795	0.833	96.71%	973.112	<.001	-1.05	0.294
M9 Contribución social	32	0.796	0.009	84	<.001	0.778	0.815	96.07%	788.034	<.001	-0.926	0.354
M10 Trabajar con niños-niñas	37	0.891	0.005	189	<.001	0.882	0.901	95.83%	863.066	<.001	-1.921	0.055
M11 Experiencia académica previa	38	0.846	0.006	146	<.001	0.835	0.858	94.62%	687.562	<.001	-0.738	0.46
M12 Influencias sociales	40	0.837	0.011	73.4	<.001	0.814	0.859	97.18%	1384.538	<.001	0.895	0.371
Utilidad Social (2° orden)	12	0.824	0.018	45.9	<.001	0.788	0.859	97.37%	418.608	<.001	-0.629	0.529
MOTIVACIONES (constructo)	44	0.803	0.007	109	<.001	0.788	0.817	93.81%	694.197	<.001	-0.442	0.659
C13 Carrera para expertos	29	0.775	0.001	77.9	<.001	0.756	0.795	96.12%	722.167	<.001	-2.028	0.043
C14 Dificultad de la tarea	33	0.693	0.015	46.5	<.001	0.663	0.722	96.28%	859.354	<.001	-2.001	0.045
Exigencia de la profesión (2° orden)	8	0.728	0.037	19.7	<.001	0.655	0.801	97.03%	235.922	<.001	1.45	0.147
C15 Estatus Social	35	0.853	0.005	167	<.001	0.843	0.863	94.35%	601.848	<.001	-3.088	0.002
C16 Buen salario	32	0.878	0.008	105	<.001	0.862	0.894	98.51%	2087.421	<.001	1.817	0.069
Beneficios de retorno (2° orden)	5	0.755	0.044	17	<.001	0.668	0.842	98.56%	278.595	<.001	0.652	0.514
C17 Disuasión social	36	0.711	0.011	65.9	<.001	0.69	0.732	91.77%	425.129	<.001	0.237	0.813
C18 Satisfacción con la elección	31	0.879	0.007	124	<.001	0.865	0.893	98.43%	1913.813	<.001	-2.844	0.004
CREENCIAS (constructo)	37	0.783	0.008	98.1	<.001	0.767	0.799	95.06%	728.631	<.001	-1.755	0.079

k= número de estudios; F= fiabilidad promedio; E.T= error típico; LI= límite inferior; LS= límite superior; P= probabilidad asociada

Considerando las 12 subdimensiones que informan de las motivaciones de los docentes (de M1 a M12), 7 de ellas (58.33%) presentan fiabilidades buenas, 3 (25%) obtienen valores aceptables y la fiabilidad de las dos restantes (16.66%) es cuestionable, es decir, está por debajo de .7 (M4 y M6). En cualquier caso, conviene señalar que, en términos generales, tanto las dimensiones de “utilidad personal” y “utilidad social” como el constructo de “motivaciones” obtienen valores para la fiabilidad que pueden considerarse buenos (fiabilidad media de 0.8). En la figura III se muestran las fiabilidades promedio de cada estudio que ofrece datos relativos al constructo de motivaciones, junto con sus respectivos intervalos de confianza.

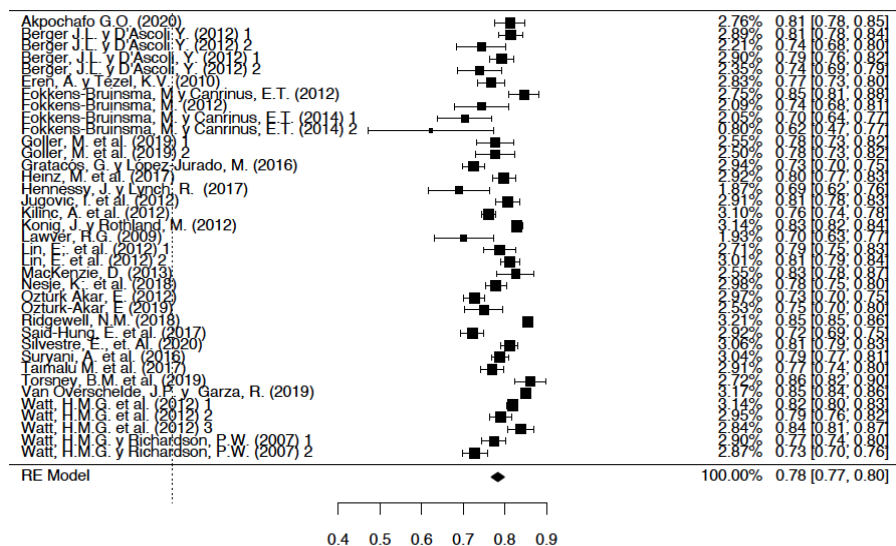
FIGURA III. Forest plot del constructo “motivaciones”



En el caso del constructo “creencias” la fiabilidad promedio es próxima a 0.8. En este sentido, la dimensión de “exigencia de la profesión” y, especialmente, una de sus subdimensiones (C14) es la que posee peores propiedades psicométricas. Por su parte, las fiabilidades de las subescalas C15 y C16 asociadas a la dimensión “beneficios de retorno” pueden considerarse buenas. En la figura IV se muestra la variabilidad

de la fiabilidad estimada para este constructo en los diferentes estudios que aportan datos al respecto.

FIGURA IV. Forest plot del constructo “creencias”



Haciendo referencia a la heterogeneidad en las propiedades psicométricas reportadas por los diferentes estudios, los valores de I^2 , superiores al 75% en el caso de todas las subdimensiones y dimensiones de la escala, y las probabilidades asociadas a los estadísticos Q, inferiores a 0.01, informan de diferencias significativas en los valores de las fiabilidades obtenidas al aplicar el FIT-Choice en diferentes grupos de sujetos. Esta heterogeneidad se ve reflejada en la Figuras III y IV, donde se identifican fácilmente de manera gráfica los intervalos para las fiabilidades promedio de las escalas.

Finalmente, es importante mencionar que el posible sesgo de publicación, evaluado a partir de la regresión de Egger, se produce en cinco subdimensiones: M5, C13, C14, C15, y C18 y en los factores de segundo orden. Es un problema que afecta principalmente a las escalas

de creencias y, en consecuencia, esos resultados deben interpretarse con cautela.

Para el logro del **tercer objetivo** específico, se estiman modelos que incluyen variables moderadoras vinculadas a las características del instrumento y a las muestras utilizadas. En primer lugar, se analiza el efecto de dispersión de las puntuaciones en los resultados de fiabilidad. Dicha dispersión afecta a la fiabilidad en M5 ($Z=2.93$; $p=0.003$), M8 ($Z=3.17$; $p=0.02$), M11 ($Z=2.30$; $p=0.021$), C13 ($Z=2.37$; $p=0.018$) y C15 ($Z=3.24$; $p=0.001$). A continuación se estudia el efecto del tipo de FIT-Choice y de las muestras. En el primer caso, los resultados se presentan a continuación por subdimensiones:

- En la subescala M2, la prueba traducida al alemán tiene efecto positivo ($Z=2.04$; $p=0.041$).
- En la M5, la utilización de una versión traducida del instrumento no parece problemática ($Z=2.48$; $p=0.013$), incluso la prueba en el idioma original (inglés) muestra un efecto negativo en la fiabilidad ($Z=-2.03$; $p=0.043$). En los trabajos que utilizan esta escala sin el propósito de hacer algún tipo de validación del instrumento tiene efecto negativo ($Z=-2.26$; $p=0.024$), incluso si el estudio pretende validar el instrumento solo con un estudio de fiabilidad ($Z=-2.10$; $p=0.036$). En cambio, en los estudios que aplican técnicas de análisis confirmatorio para el estudio de la validez el efecto es positivo ($Z=2.90$; $p=0.004$).
- En la subdimensión M6, los trabajos que no realizan validación tienen un efecto negativo en la fiabilidad ($Z=-2.23$; $p=0.026$).
- En la M7, el cambio de constructo respecto al modelo original tiene un efecto negativo ($Z=-2.26$; $p=0.024$) y el uso de la prueba en el idioma original muestra un efecto positivo en la fiabilidad ($Z=2.74$; $p=0.006$).
- En la M8, la prueba en el idioma original muestra un efecto positivo ($Z=2.07$; $p=0.039$).
- En la M10, el cambio de constructo respecto al modelo original tiene un efecto positivo ($Z=2.26$; $p=0.024$), si el estudio no es de validación el efecto es negativo ($Z=-4.61$; $p<0.001$) y también si se utiliza una versión traducida del instrumento ($Z=-2.03$; $p=0.043$).
- En la subescala C16, la traducción al alemán tiene efecto positivo ($Z=2.61$; $p=0.009$), en cambio, la traducción al español tiene un efecto negativo en la fiabilidad ($Z=-2.09$; $p=0.036$).

- Y en la C18, si el trabajo no tiene el propósito de validación del instrumento, se ha observado un efecto negativo ($Z=-3.57$; $p<0.001$). En cambio los estudios confirmatorios han mostrado un efecto positivo en la fiabilidad ($Z=3.78$; $p<0.001$).

Por último, en el factor de 2º orden de utilidad personal, la prueba traducida al alemán tiene efecto positivo ($Z=2.53$; $p=0.011$).

Por otra parte, para presentar las características muestrales que pueden afectar a la variabilidad, los resultados se agrupan por variables.

La especialidad de formación docente ha mostrado efectos significativos en la fiabilidad de algunas escalas. Los maestros y maestras de primaria tienen un efecto positivo en la fiabilidad en la escala de creencias ($Z=2.17$; $p=0.03$). La muestra de profesores de secundaria determina los resultados de fiabilidad de la subescala M4 ($Z=-2.37$; $p=0.018$), M10 ($Z=-2.52$; $p=0.01$) y C18 ($Z=-2.99$; $p=0.003$). La muestra de profesores de idiomas tiene efecto positivo en los M6 ($Z=2.69$; $p=0.007$), M10 ($Z=2.58$; $p<0.01$) y C17 ($Z=2.62$; $p=0.009$). En cambio, es negativo en el factor de segundo orden de utilidad social ($Z=-2.54$; $p=0.011$) y en la subdimensión C13 ($Z=-3.59$; $p<0.001$).

Utilizar una muestra mixta de profesores de infantil y de secundaria tiene efecto negativo en los resultados de la subescala M6 ($Z=-2.04$; $p=0.042$) y en el factor de 2º orden de la demanda de la tarea ($Z=-2.47$; $p=0.013$). Utilizar una muestra mixta de profesores de infantil y primaria tiene efecto negativo en M7 ($Z=-3.18$; $p=0.001$), M8 ($Z=-2.16$; $p=0.031$) y M9 ($Z=-2.58$; $p=0.01$). La muestra mixta de primaria y secundaria tiene efecto negativo en M6 ($Z=-2.18$; $p=0.030$) y positivo sobre en el factor de 2º orden de utilidad personal ($Z=1.98$; $p=0.048$) y de utilidad social ($Z=2.52$; $p<0.012$), en el constructo general de motivaciones ($Z=2.02$; $p<0.044$) y en la estimación de la fiabilidad total para la escala de creencias ($Z=2.96$; $p=0.003$) así como en la subescala C18 ($Z=3.11$; $p=0.002$).

El hecho de que la muestra esté compuesta por estudiantes tiene efecto negativo sobre la fiabilidad en las subescalas M6 ($Z=-2.78$; $p=0.05$), M8 ($Z=-2.05$; $p=0.041$), C16 ($Z=-2.09$; $p=0.036$) y C18 ($Z=-2.03$; $p=0.042$).

Por último, la edad de los participantes afecta en términos positivos a la variable M7 ($Z=2.41$; $p=0.016$), M8 ($Z=3.17$; $p=0.02$), M9 ($Z=2.11$; $p=0.034$) y C16 ($Z=2.03$; $p=0.043$); y de forma negativa en el factor de 2º orden de utilidad social ($Z=-2.50$; $p<0.012$) y de exigencia de la profesión ($Z=-2.50$; $p=0.012$).

Discusión y conclusiones

En el panorama actual de preocupación por la eficacia de los profesores es de interés conocer si los factores que explican la elección de la carrera docente son similares en todos los países. Para ello, es necesario disponer de instrumentos de medida que acrediten su fiabilidad en estudios internacionales y que, por tanto, permitan esa comparación. El FIT-Choice, diseñado en 2007 e inicialmente validado para una muestra de 1653 estudiantes australianos (Watt y Richardson, 2007), tras su gran historial de aplicaciones, se ha confirmado como una herramienta útil para dicho objetivo.

El presente estudio ha perseguido analizar, comparar y sintetizar cuantitativamente las fiabilidades de la escala FIT-Choice, tras su aplicación en diversas muestras internacionales para medir las motivaciones y creencias de los aspirantes a maestro, así como de docentes en ejercicio. Dado que se trata de un instrumento de uso extensivo, es oportuno reflexionar sobre su consistencia interna, a través de una revisión sistemática y un meta-análisis de generalización de su fiabilidad. Los resultados obtenidos indican una gran heterogeneidad de la fiabilidad que, en su conjunto, es adecuada, aunque se observa un comportamiento psicométrico diferencial en algunas de las dimensiones y subdimensiones. En términos generales, tanto las dimensiones de utilidad personal y utilidad social como el constructo de motivaciones obtienen valores para la fiabilidad que pueden considerarse buenos (fiabilidad media de 0.8). Por otra parte, tanto el constructo de creencias como las subdimensiones que lo constituyen logran una fiabilidad media próxima a 0.8, exceptuando la subescala C14, dificultad de la tarea, que está por debajo de 0.6.

Los creadores del cuestionario argumentan que el instrumento permite una comparación internacional de las motivaciones para acceder a la carrera docente. Watt et al. (2012) analizan los resultados de Australia, EE.UU, Noruega y Alemania y concluyen que la estructura de dimensiones es estable entre países, si bien las creencias vinculadas con la profesión varían de forma considerable. Sus hallazgos les llevaron a descartar, por su baja fiabilidad, las subescalas de “movilidad en el trabajo” y “carrera de segunda opción” (M4 y M6). Los resultados de este meta-análisis muestran que la escala de utilidad personal y el buen salario tienen una mayor fiabilidad en la versión alemana del instrumento. Sin embargo,

la consideración del buen salario de la profesión es menos fiable en la versión española empleada también en Latinoamérica.

Los resultados señalan que la escala de elección de la profesión para ayudar a la equidad social tiene más fiabilidad en la versión original en inglés, aunque también se ve más afectada por la dispersión de las puntuaciones. Ello podría informar de que esta escala cuenta con significados distintos en función del contexto, por lo que debería adaptarse a cada situación. Esta limitación aparece ya en los meta-análisis previos y una posible explicación, como señalaran los autores de la teoría del valor de las expectativas (Klassen et al., 2011), es que los instrumentos cuantitativos puedan enmascarar diferencias culturales en estas motivaciones porque asumen que aquellas que son dominantes en países occidentales también lo son en otros contextos. Como prospectiva cabría hacer una revisión sistemática de los trabajos atendiendo a la teoría que justifica esas diferencias entre culturas o tipos de docentes. Incluso convendría meta-analizar las diferencias de medias entre los grupos (occidente frente a otros), etapa educativa (infantil, primaria, secundaria), estudiantes frente a docentes en ejercicio o etapa formativa. Además, cabría ampliar la muestra de estudios incluyendo trabajos que aplican otro tipo de instrumentos.

Un factor de complejidad en el estudio de la fiabilidad global del FIT-Choice está en la dimensionalidad del constructo. Las diferencias contextuales podrían estar en la base de las variaciones en la estructura de dimensiones observada, lo que precisaría de un mayor estudio. Estas modificaciones normalmente están relacionadas con la unión de algunas escalas y otras veces suponen la incorporación de dimensiones nuevas. Por ejemplo, el trabajo de Suryani et al. (2016) incorpora una nueva escala de influencias religiosas y otra para identificar la disuasión provocada por los medios de comunicación. Van Overschelde y Garza (2019) incluyen una dimensión de motivación de elección de la profesión para contribuir al cambio social, juntando ítems de las escalas del valor de utilidad social; y Zhang et al. (2020) definen una escala de valor extrínseco. En el trabajo de Akpochafo (2020) se agrupan, por un lado, las dimensiones de habilidad y motivaciones intrínsecas y, por otro, los antecedentes contextuales con escalas de motivaciones (como las experiencias académicas previas con elegir la profesión para hacer una contribución social o la influencia social con la seguridad de la profesión). En los trabajos de Berger y D'Ascoli (2012a y 2012b) se unen

escalas pertenecientes a la utilidad social, como contribuir a moldear el futuro o hacer una contribución social, al tiempo que se generan nuevas dimensiones como las creencias constructivistas o la transmisión directa. Superar la limitación que supone la heterogeneidad observada en la dimensionalidad del constructo plantea como prospectiva un meta-análisis de validación del constructo.

A pesar de las dificultades existentes, con el meta-análisis realizado, complementado con el estudio del efecto de las variables moderadoras seleccionadas, este trabajo supone un paso importante hacia el esclarecimiento de las propiedades métricas de un instrumento de alcance internacional, como es el FIT-Choice, y hace una aportación en la línea que se deriva del meta-análisis de Heinz (2015), en el que se recomienda profundizar en los estudios comparados de las motivaciones de elección de la carrera docente ya que pueden variar en función del perfil. Los datos de fiabilidad promedio conseguidos avalan la utilidad del cuestionario y su valor no solo como herramienta para el estudio comparado de las motivaciones y creencias implicados en la elección de la profesión docente en estudios internacionales sino también como instrumento apto para la investigación acerca de la dimensionalidad de un constructo tan complejo como interesante.

Referencias bibliográficas

- Australian Institute for Training and School Leadership (AITSL). (2011). National Professional Standards for Teachers. Recuperado de: <https://www.aitsl.edu.au/docs/default-source/national-policy-framework/australian-professional-standards-for-teachers.pdf>
- Akpochafo, G.O. (2020). Factors influencing undergraduates' choice of teaching as a career (Fit-choice) in Nigeria. *International Journal of Education and Practice*, 8, 121-133. <https://doi.org/10.18488/journal.61.2020.81.121.133>
- Alpaslan, M. M., Ulubey, O. y Yildirim, K. (2018). Examining the contributions of support and class belonging to preservice teachers' career motivation in Turkey, *Turkish Journal of Education*. 7(4), 196-209. <https://doi.org/10.19128/turje.458535>

- Barber, M. y Mourshed, M. (2007). *How the world's best-performing school systems come out on top*. McKinsey y Co.
- Berger J.-L. y D'Ascoli Y. (2012a). Becoming a VET teacher as a second career: Investigating the determinants of career choice and their relation to perceptions about prior occupation. *Asia-Pacific Journal of Teacher Education*, 40(3), 317-341. <https://doi.org/10.1080/1359866X.2012.700046>
- Berger J.-L. y D'Ascoli Y. (2012b). Motivations to Become Vocational Education and Training Educators: A Person-Oriented Approach. *Vocations and Learning*, 5, 225-249. <http://dx.doi.org/10.1007/s12186-012-9075-z>
- Berger, J.-L. y Girardet, C. (2020). Vocational teachers' classroom management style: the role of motivation to teach and sense of responsibility. *European Journal of Teacher Education*, (online first) <https://doi.org/10.1080/02619768.2020.1764930>
- Bonett, D. G. (2010). Varying coefficient meta-analytic methods for alpha reliability. *Psychological methods*, 15(4), 368-385. <https://doi.org/10.1037/a0020142>
- Botella, J., Suero, M. y Gambará, H. (2010). Psychometric inferences from a meta-analysis of reliability and internal consistency coefficients. *Psychological Methods*, 15(4), 386-397. <https://doi.org/10.1037/a0019626>
- Brookhart, S. M. y Freeman, D. J. (1992). Characteristics of entering teacher candidates. *Review of Educational Research*, 62(1), 37-60. <https://doi.org/10.3102%2F00346543062001037>
- Choi, S. y Song, J. (2015). Korean pre-service English teachers' motivations for choosing a teaching career. *International Information Institute (Tokyo)*, 18(3), 921-925.
- Coe, R., Aloisi, C., Higgins, S. y Major, L. E. (2014) *What makes great teaching? Review of the underpinning research. Research Report*. Recuperado de: <https://www.suttontrust.com/wp-content/uploads/2014/10/What-Makes-Great-Teaching-REPORT.pdf>
- Eccles, J. S. (2005). Subjective task value and the Eccles et al. Model of achievement-related choices. En A. J. Elliot y C. S. Dweck (Eds.). *Handbook of competence and motivation*. Guilford.
- Eren, A. y Tezel, K.V. (2010). Factors influencing teaching choice, professional plans about teaching, and future time perspective: A mediational analysis. *Teaching and Teacher Education*, 26, 1416-1428. <https://doi.org/10.1016/j.tate.2010.05.001>

- Fray, L. y Gore, J. (2018). Why people choose teaching: A scoping review of empirical studies, 2007–2016. *Teaching and Teacher Education*, 75, 153-163. <https://doi.org/10.1016/j.tate.2018.06.009>
- Fokkens-Bruinsma, M. y Canrinus, E. (2012a). Adaptive and Maladaptive Motives for Becoming a Teacher. *Journal of Education for Teaching: International Research and Pedagogy*, 38(1), 3-19. <https://doi.org/10.1080/02607476.2012.643652>
- Fokkens-Bruinsma, M. y Canrinus, E. (2012b). The Factors Influencing Teaching (FIT)-Choice scale in a Dutch teacher education program. *Asia-Pacific Journal of Teacher Education*, 40(3), 249-269. <https://doi.org/10.1080/1359866X.2012.700043>
- Fokkens-Bruinsma, M. y Canrinus, E.T. (2014). Motivation for becoming a teacher and engagement with the profession: Evidence from different contexts. *International Journal of Educational Research*, 65, 65-74. <http://dx.doi.org/10.1016/j.ijer.2013.09.012>
- Goldhaber, D. (2010). *Teacher Pay Reforms: The Political Implications of Recent Research*. Working paper no. 2010-4.0. Center for Education Data and Research (CEDR), University of Washington. Recuperado de [http://www.cedr.us/papers/working/CEDR%20WP%202010-4_Teacher%20Pay%20Reforms%20\(8-23-10\).pdf](http://www.cedr.us/papers/working/CEDR%20WP%202010-4_Teacher%20Pay%20Reforms%20(8-23-10).pdf)
- Goller, M., Ursin, J., Vahasantanan, K., Festner, D. y Harteis, C. (2019). Finnish and German student teachers' motivations for choosing teaching as a career. The first application of the FIT-Choice scale in Finland. *Teaching and Teacher Education*, 85, 235-248. <https://doi.org/10.1016/j.tate.2019.06.023>
- Graham, J. M., Diebels, K. J. y Barnow, Z. B. (2011). The reliability of relationship satisfaction: A reliability generalization meta-analysis. *Journal of Family Psychology*, 25(1), 39-48. <https://doi.org/10.1037/a0022441>
- Gratacós, G. y López-Jurado, M. (2016). Validación de la versión en español de la escala de los factores que influyen en la elección de los estudios de educación (FIT-choice). *Revista de Educación*, 372, 87-110. <https://doi.org/10.4438/1988-592X-RE-2015-372-316>
- Greco, L. M., O'Boyle, E.H., Cockburn, B.S. y Yuan, Z. (2018). Meta-analysis of coefficient alpha: A reliability generalization study. *Journal of Management Studies*, 55(4), 583-618. <https://psycnet.apa.org/doi/10.1111/joms.12328>

- Hattie, J. A. (2008). *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. Routledge.
- Hattie, J. A. (2011). *Visible Learning for Teachers: Maximizing Impact on Learning*. Routledge.
- Hedges, L. V. y Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press.
- Heinz, M. (2015). Why choose teaching? An international review of empirical studies exploring student teachers' career motivations and levels of commitment to teaching. *Educational Research and Evaluation*, 21(3), 258-297. <https://doi.org/10.1080/13803611.2015.1018278>
- Heinz, M., Keane, E. y Foley, C. (2017). Career motivations of student teachers in the republic of Ireland continuity and change during educational reform and 'boom to bust' economic times. In H.M.G. Watt, P.W. Richardson, y K. Smith (Eds.) *Global Perspectives on Teacher Motivation* (pp. 22-54). Cambridge University Press.
- Hennessy, J. y Lynch, R. (2017). "I Chose to Become a Teacher Because". Exploring the Factors Influencing Teaching Choice Amongst Pre-Service Teachers in Ireland. *Asia-Pacific Journal of Teacher Education*. 45(2), 106-125. <https://doi.org/10.1080/1359866X.2016.1183188>
- Higgins, J. P. y Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in medicine*, 21(11), 1539-1558. <https://doi.org/10.1002/sim.1186>
- INEE (2020). *TALIS 2018. Marco conceptual*. Ministerio de Educación y Formación Profesional.
- Jablanovic, M.M. y Vracar, S. (2019). Exploring Serbian teachers' motivation for teaching with the application of FIT-choice scale. *Zbornik Instituta za Pedagoska Istrazivanja*, 51(1), 7-45. <https://doi.org/10.2298/ZIPI1901007M>
- Jamovi project (2020). jamovi. (Version 1.6) [Computer Software]. Recuperado de <https://www.jamovi.org>.
- Jordan, H. R., Mendro, R. L. y Weerasinghe, D. (1997). *Teacher effects on longitudinal student achievement: A preliminary report on research on teacher effectiveness*. National Evaluation Institute, Indianapolis, IN. Recuperado de <https://www.dallasisd.org/cms/lib/TX01001475/Centricity/Shared/evalacct/research/articles/Jordan-Teacher-Effects-on-Longitudinal-Student-Achievement-1997.pdf>

- Jugovic, I., Marusic, I., Ivanec, T.P. y Vidovic, V.V. (2012). Motivation and personality of preservice teachers in Croatia. *Asia-Pacific Journal of Teacher Education*, 40(3), 271-287. <https://www.doi.org/10.1080/1359866X.2012.700044>
- Kilinc, A., Watt, H.M.G. y Richardson, P.W. (2012). Factors Influencing Teaching Choice in Turkey. *Asia-Pacific Journal of Teacher Education*, 40(3), 199-226. <https://www.doi.org/10.1080/1359866X.2012.700048>
- Klassen, R. M. y Dolan, R. (2015, September). *Selection for teacher education in the UK and the Republic of Ireland: A proposal for innovation*. Presented at the meeting of the European Conference on Educational Research, Budapest, Hungary.
- Klassen, R. M. y Kim, L. E. (2017) Assessing critical attributes of prospective teachers: Implications for selection into initial teacher education programs. *British Journal of Educational Psychology*. Monograph Series II, 12, 5-22.
- Klassen, R. M., Al-Dhafri, S., Hannok, W. y Betts, S.M. (2011). Investigating pre-service teacher motivation across cultures using the Teachers' Ten Statements Test. *Teaching and Teacher Education*, 27(3), 579-588. <https://doi.org/10.1016/j.tate.2010.10.012>
- Konig, J. y Rothland, M. (2012). Motivations for choosing teaching as a career: effects on general pedagogical knowledge during initial teacher education. *Asia-Pacific Journal of Teacher Education*, 40(3), 289-315. <https://doi.org/10.1080/1359866X.2012.700045>
- Lawver, R. G. (2009). *Factors influencing agricultural education students' choice to teach* (Doctoral Dissertation). University of Missouri. <https://doi.org/10.32469/10355/6168>
- Lawver, R. G. (2012). An Analysis of Post-Secondary Agricultural Education Students' Choice to Teach. *Journal of Agricultural Education*. 53(2), 28-42. <https://doi.org/10.5032/jae.2012.02028>
- Lin, E., Shi, Q.M., Wang, J., Zhang, S.A. y Hui, L. (2012). Initial motivations for teaching: comparison between preservice teachers in the United States and China. *Asia-Pacific Journal of Teacher Education*, 40(3), 227-248. <https://doi.org/10.1080/1359866X.2012.700047>
- MacKenzie, D. (2013). *Choosing a career: a study of motivational factors and demographics that influence P-12 pre-service teachers* (Doctoral Dissertation). University of Louisville. <https://ir.library.louisville.edu/etd/877/>

- Nesje, K., Brandmo, C. y Berger, J.L. (2018). Motivation to Become a Teacher: a Norwegian Validation of the Factors Influencing Teaching Choice Scale. *Scandinavian Journal of Educational Research*, 62, 813-831. <https://doi.org/10.1080/00313831.2017.1306804>
- OECD (2017). Do new teachers feel prepared for teaching? *Teaching in Focus*, 17. OECD Publ.. <http://dx.doi.org/10.1787/980bf07d-en>.
- Ozturk-Akar, E. (2012). Motivations of Turkish Pre-service Teachers to Choose Teaching as a Career. *Australian Journal of Teacher Education*, 37(10) <http://dx.doi.org/10.14221/ajte.2012v37n10.7>
- Ozturk-Akar, E. (2019). Alternative Teacher Certification Students' Motivations Of Teaching. *Australian Journal of Teacher Education*. 44(11), <http://dx.doi.org/10.14221/ajte.2019v44.n11.3>
- Ponnock, A. R. (2018). Motivational Differences throughout Teachers' Preparation and Career. *New Waves-Educational Research and Development Journal*, 21(2), 26-45. <https://files.eric.ed.gov/fulltext/EJ1211293.pdf>
- Richardson, P. W. y Watt, H. M. (2006). Who chooses teaching and why? Profiling characteristics and motivations across three Australian universities. *Asia-Pacific Journal of Teacher Education*, 34(1), 27-56. <http://dx.doi.org/10.1080/13598660500480290>
- Ridgewell, N.M. (2018). *Is It the Right "Fit?" Florida Teachers' Motivations for Choosing a Teaching Career Using the FIT-Choice Scale* (Doctoral Dissertation). University of Florida. <https://ufdc.ufl.edu/UFE0052721>
- Rodriguez, M. C. y Maeda, Y. (2006). Meta-analysis of coefficient alpha. *Psychological Methods*, 11(3), 306-322. <https://doi.org/10.1037/1082-989X.11.3.306>
- Said-Hung, E., Gratacós, G. y Valencia, J. (2017). Factores que influyen en la elección de las carreras de pedagogía en Colombia. *Educação e Pesquisa*, 43, 31-48. <https://doi.org/10.1590/s1517-9702201701160978>
- Sánchez-Meca, J., López-López, J. A. y López-Pina, J. A. (2013). Some recommended statistical analytic practices when reliability generalization studies are conducted. *British Journal of Mathematical and Statistical Psychology*, 66(3), 402-425. <https://doi.org/10.1111/j.2044-8317.2012.02057.x>
- Schleicher, A. (2018). *World Class: How to Build a 21st-Century School System*, Strong Performers and Successful Reformers in Education, OECD Publishing, Paris, <https://doi.org/10.1787/9789264300002-en>.

- Silvestre, E., Figueroa, V. y Díaz, J.V. (2020). ¿Por qué los estudiantes de la Universidad Pedagógica ISFODOSU escogen la carrera de Educación? *Ciencia y Educación*, 1, 47-69. <https://doi.org/10.22206/cyed.2020.v4i1.pp47-69>
- Suryani, A., Watt, H.M.G. y Richardson, P.W. (2016). Students' motivations to become teachers: FIT-Choice findings from Indonesia. *International Journal of Quantitative Research in Education*, 3(3), 179-203. <http://dx.doi.org/10.1504/ijqre.2016.077802>
- Taimalu M., Luik P. y Täht, K. (2017). Teaching motivations and perceptions during the first year of teacher education in Estonia. In H.M.G. Watt, P.W. Richardson, y K. Smith (Eds.) *Global Perspectives on Teacher Motivation* (pp. 189-219). Cambridge University Press.
- Topkaya, E. y Uztosun, M. S. (2012). Choosing Teaching as a Career: Motivations of Pre-service English Teachers in Turkey. *Journal of Language Teaching and Research*, 3(1), 126-134. <http://dx.doi.org/10.4304/jltr.3.1.126-134>
- Torsney, B.M., Lombardi, D. y Ponnock, A. (2019). The role of values in pre-service teachers' intentions for professional engagement. *Educational Psychology*, 39(1), 19-37. <https://doi.org/10.1080/01443410.2018.1504892>
- Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement*, 58(1), 6-20. <http://dx.doi.org/10.1177/0013164498058001002>
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36, 1-48. <https://doi.org/10.18637/jss.v036.i03>
- Van Overschelde, J.P. y Garza, R. (2019). Motivational and perceptual differences between Hispanic and White preservice teachers. *Journal of Latinos and Education*, 19(3), 277-293. <https://doi.org/10.1080/15348431.2018.1510331>
- Wang, W. (2019) 'I really like teaching, but...' A mixed methods study exploring pre-service teachers' motivations for choosing teaching as a career (Doctoral Dissertation). University of Glasgow. <http://theses.gla.ac.uk/id/eprint/72474>
- Watt, H. M. G. y Richardson, P.W. (2007). Motivational factors influencing teaching as a career choice: development and validation of the FIT-Choice scale. *Journal of Experimental Education*, 75(3), 167-202. <https://doi.org/10.3200/JEXE.75.3.167-202>

- Watt, H. M., Richardson, P. W., Klusmann, U., Kunter, M., Beyer, B., Trautwein, U. y Baumert, J. (2012). Motivations for choosing teaching as a career: An international comparison using the FIT-Choice scale. *Teaching and teacher education*, 28(6), 791-805. <https://doi.org/10.1016/j.tate.2012.03.003>
- Zhang, H.W., Wu, J.J. y Zhu, Y. (2020). Why do you choose to teach Chinese as a second language? A study of pre-service CSL teachers' motivations. *System*, 91, 102242. <https://doi.org/10.1016/j.system.2020.102242>
- Zordo, L. de, Hagenauer, G y Hascher, T. (2019). Student teachers' emotions in anticipation of their first team practicum. *Studies in Higher Education*, 44(10), 1758-1767. <https://doi-org.bucm.idm.oclc.org/10.1080/03075079.2019.1665321>

Información de contacto: Enrique Navarro Asencio. Universidad Complutense de Madrid. Facultad de Educación. Departamento Investigación y Psicología en Educación. Rector Royo Villanova, s/n. 28040 Madrid. E-mail: enriquenavarro.a@gmail.com