

# **Meta-analysis of reliability generalisation of the FIT-Choice questionnaire (Factors Influencing Teaching Choice)<sup>1</sup>**

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

*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*

### **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

---

<sup>(1)</sup> Funding: Project nº RTI2018-099365-B-100 funded by the Ministry of Science and Innovation and the European Union in the 2018' National Programme for R+D+i Oriented to the Challenges of Society.

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

### **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 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

## Introduction

Research has shown that the best thing schools can do for students is to provide good teachers, since being taught by good teachers can put even the most disadvantaged students on the path to university while time spent with ineffective teachers can lead to school failure that students struggle to overcome (Jordan et al., 1997; Schleicher, 2018). The essential role of teachers in the quality of the education students receive has been demonstrated in a large number of empirical studies. Hattie (2008 and 2011), who performed meta-analyses of over 65,000 research works on the effects of hundreds of interventions on the learning of 250 million students, concludes that factors such as class size or level of investment, which are usually regarded as important, actually have little effect on students' learning. Instead, the quality of the teacher is the most decisive factor after controlling for other contextual effects such as the origin of the students, and increased teacher quality has a more effective impact than any other educational investment, even reducing class size (Goldhaber, 2010). The famous *McKinsey report* (Barber & Mourshed, 2007) also identifies teachers as a key factor in the educational systems that have the best results in international evaluations, to the extent that an educational system can never be better than the quality of its teachers.

This backdrop justifies the focus in research on establishing which characteristics of teachers or people who wish to become teachers relate most closely to their efficacy. One example of this interest can be found in the *Teaching and Learning International Survey* (TALIS) (INEE, 2020) and the *TALIS Initial Teacher Preparation study* (OECD, 2017), which focus on the importance of selecting the best candidates for initial teacher education (ITE) programmes. In this regard, the policies of countries such as England and Wales stand out, where Klassen and Dolan (2015), after reviewing 74 ITE programmes, confirmed that they all use procedures for selecting prospective teachers based on their cognitive and non-cognitive characteristics. Similarly, in Australia the standards set by the Australian Institute for Teaching and School Leadership (AITSL) for candidates include levels of command of language and maths equivalent to the top 30% of the population (AITSL, 2011).

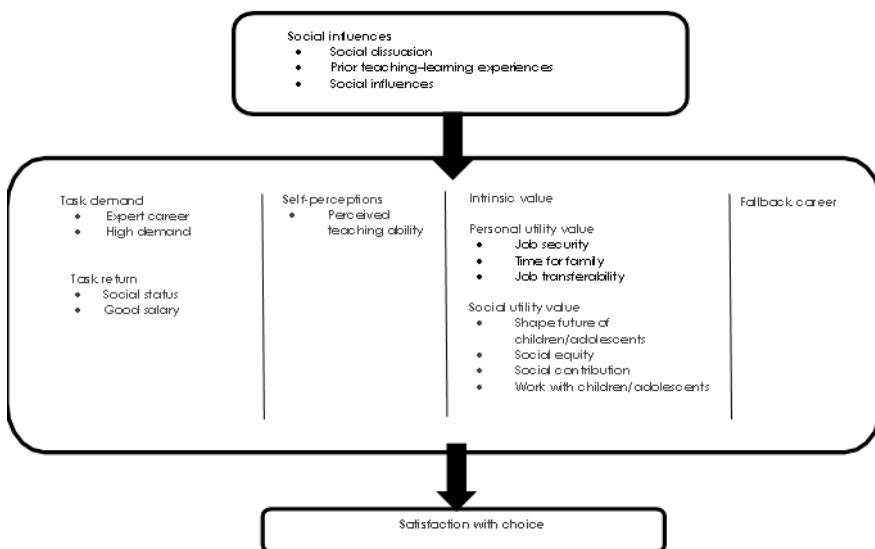
The volume of primary research on this topic has inspired relevant secondary research, such as the meta-analysis by Klassen and Kim (2017), which identifies a small but significant effect of teaching efficacy relating

to a set of individual academic attributes (disciplinary knowledge, cognitive skills and pedagogical knowledge) and non-academic attributes (personality, motivation, beliefs and dispositions), which in turn interact with contextual factors (cultural, social and prior teaching–learning experience). The systematic review by Coe et al. (2014) is also noteworthy. This analyses 200 works and identifies six components of what they call “great teaching”. These are principally pedagogical content knowledge and quality of instruction, as well as classroom management, classroom climate, beliefs about teaching and learning and other wider professional elements such as relationships or career development. With regards to the importance of non-cognitive factors, the meta-analysis by Brookhart and Freeman (1992) should be noted. This included works from the USA from between 1960 and 1990. The synthesis of these works suggests that altruism and intrinsic motivations along with service orientation were the main reasons for choosing a career in teaching. Heinz's work (2015) continues with analysis of research into non-cognitive factors, reviewing works from 1990 until 2014 in depth. She performed a systematic review of 41 studies, concluding that in 23 of them the reasons that led people to choose teaching as a profession are grouped around the same factors: intrinsic, extrinsic and altruistic motivations. Finally, the review by Fray and Gore (2018) of empirical studies published between 2007 and 2016 concludes that in most of them (63 of 70), intrinsic and altruistic teaching motivations have more weight than extrinsic ones.

The model Helen Watt and Paul Richardson propose for measuring the factors at play in choosing teaching as a profession includes all of these components and links the decision to choose a career in teaching with expectancy-value theory (Eccles, 2005). In this way, the choice is, on the one hand, affected by the value attributed to the profession, which depends on enjoyment or perceived usefulness, and, on the other hand, by expectations of success, which depend on beliefs about the profession and the perception of self-efficacy in the performance of the task, both of which in turn relate to prior experiences and their interpretation (Watt et al., 2012). Figure I shows the model, with a dimension of academic antecedents and social influences that can shape beliefs about the profession, perceived self-efficacy, motivations and the possibility of having chosen an alternative profession. The model distinguishes between intrinsic motivations, personal utility motivations and social value motivations. It also includes a factor relating to perceptions of the

profession relating to its difficulty and professional demands as well as possible benefits (salary, status, etc.). Finally, it considers a factor relating to self-perceived teaching ability. This model is the theoretical basis of the FIT-Choice (Factors Influencing Teaching Choice) instrument for measuring motivation.

FIGURE I. FIT-Choice Theoretical Model.



Source: Adapted from Watt & Richardson (2007)

FIT-Choice has a total of 18 subscales grouped into four second-order factors: personal utility value, social utility value, high demand and task return. The first two refer to motivational aspects and the second two to beliefs about the profession. In turn, the antecedents scales also refer to motivations (previous teaching–learning experiences and social influences) and beliefs (social dissuasion and satisfaction with choice).

The systematic reviews performed up to now justify the interest in studying this instrument, and the international scope of its use is especially interesting. FIT-Choice is used in 10 of the 41 articles in Heinz's (2015) review and in 17 of the 70 works reviewed by Fray and Gore (2018). Spain is one of the countries it has been used in, where Gratacós and López-

Jurado (2016) have validated it. The application of this questionnaire to very diverse populations has not only resulted in adaptations associated with translating it into other languages but also changes to the structure, items and original dimensions, changes that make a global study of its psychometric characteristics necessary. Accordingly, the present work focusses on the general reliability of FIT-Choice and seeks to summarise quantitatively the reliability results obtained in educational research that has applied this instrument as well as analysing the factors associated with it. As specific objectives, therefore, we intend to:

- Describe the characteristics of the instrument and the sample that the different studies have used.
- Estimate the mean reliability of the scales of the instrument across different studies
- Identify the effect of possible factors associated with reliability variance between different pieces of research (type of adaptation of the instrument, type of teacher evaluated, average age or percentage of women in the sample).

## Method

To achieve our objectives, we used the reliability generalisation (RG) meta-analytic method (Vacha-Haase, 1998). In general terms, reliability in classical test theory (CTT) is defined as the amount of variance in answers that does not result from measurement errors. Cronbach's alpha coefficient is one of the most commonly used procedures for estimating this.

RG meta-analysis is growing in importance and is a form of systemic review that overcomes the measurement problems present in primary studies (Greco et al., 2018). Graham et al. (2011) identify three types of application of this sort of meta-analysis: 1) estimating the average score reliability for the measurements used between different studies, 2) studying factors at play in the variance of reliability between different studies, and 3) answering the question of how reliable the scores obtained are in studies with different samples and characteristics.

This technique starts with the reliability figures of the different studies to calculate the central tendency and variability of the indices, which are analysed using a specific methodology since as Rodríguez

and Maeda (2006) note, the distribution of reliability coefficient scores does not have the same properties as the distribution of effect sizes used in traditional meta-analyses. Although there are various analysis strategies, in this work we have opted for a random effects model with a weighted transformation of the alpha coefficients, including variability data to consider the possible effect of the heterogeneity of the sample on the reliability estimation. This random effects model is recommended by Botella et al. (2010) and Sánchez-Meca et al. (2013) owing to its greater capacity for generalising conclusions and because it is the best way of drawing inferences about reliability as it makes estimating the effects of different sources of variation possible.

## Search strategy and selection of studies

We performed the search for articles on 6 July 2020 in the following international data bases: the main collection of *Web of Science* (WoS), the database of the *Education Resource Information Center* (ERIC), *Scopus*, *Google Scholar* and Dialnet. We used *FIT-Choice* and *Factors Influencing Teaching Choice* as search terms and we did not set a time limit. We searched in the “default” field of ERIC and Dialnet, the “topic” field of WoS, the “title, abstract and keywords” field of *Scopus* and the “title” field of Google Scholar. This last database only allows the search to be limited to the title field, and not the title, abstract and keywords field or default fields. The following search function was used<sup>2</sup>:

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

We also consulted the leading international repositories of doctoral theses: DART-Europe and *Open Access Theses and Dissertations*.

The literature search enabled us to retrieve 261 documents: 62 records from WoS, 50 from ERIC, 79 from *Scopus*, 5 from Dialnet, 53 from *Google Scholar*, 1 from DART-Europe and 11 from *Open Access Theses and Dissertations*. We also found two articles from other sources. After eliminating 100 duplicate texts, we performed an initial review of the

---

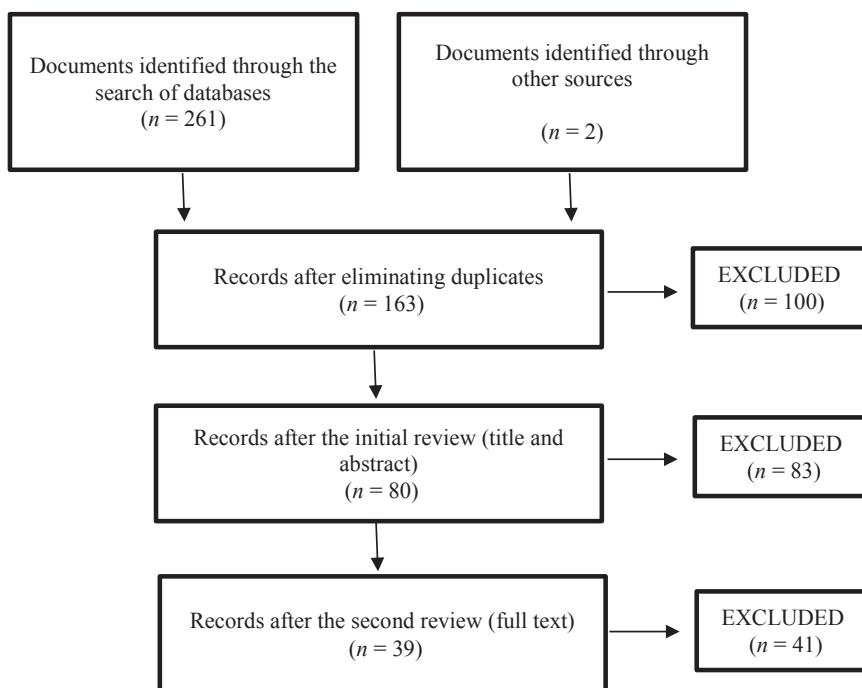
<sup>(2)</sup> When searching in *Google Scholar*, we eliminated truncation to avoid retrieving irrelevant documents.

remaining 163 works based on the title and abstract. This review enabled us to reject 83 texts that did not analyse the psychometric properties of FIT-Choice (different topic or theoretical studies) or were published in a language other than Spanish or English. This initial selection was done by two independent researchers, with a 94% agreement level. We evaluated the 80 works selected in this first phase by reading the full text. This enabled us to reject 41 works, primarily because they did not provide information about the reliability of FIT-Choice or because they referred to values reported by other works that we had already included. Figure II summarises the process described.

---

**FIGURE II.** Flow chart of the study selection process

---



On the basis of this, we analysed a total of 39 studies, which as a group provide information about the psychometric properties of FIT-

Choice after its application to 48 samples of different subjects. Table I shows the criteria used for coding these studies.

## Coding

As well as the descriptive information about the studies and samples, we characterised the FIT-Choice instrument by first considering the type of scale used, differentiating between use of the original instrument validated by Watt and Richardson (2007) and modified versions of it, whether these modifications involve reducing the scale, changing the construct or adapting it to another language. We also specify whether the study validates the scale as one of its main aims or as a preliminary study to justify its use, also noting whether they do not do so, do so through a reliability study alone or use exploratory and confirmatory factorial methods. In the case of studies that use translations of the instrument, the language in question is also noted. Finally, the information needed to be able to carry out the quantitative summary of reliability is given: a) Cronbach's alpha (in two studies they were figures about composite reliability obtained with the factorial scores in a confirmatory study), b) the number of items in each scale and c) the standard deviation of the sample.

Some of the works that adapt the model propose a new structure for organising dimensions. In these cases, the name of the new factor is added along with the previous information.

**TABLE I.** System for the coding of articles

	<b>Variable</b>	<b>Labels</b>	<b>Type</b>
1	Identifier		Chain
2	Author(s)		Chain
3	Year of publication		Numerical
4	Title of the document		Chain
5	Type of publication	1. Article 2. Book 3. Doctoral thesis 4. Research report 5. Others	Numerical
6	Title of publication		Chain
7	Language of publication		Chain
8	Keywords		Chain
9	FIT-Choice (type of scale)	1. Original (Watt & Richardson, 2007) 2. Scale that adapts the dimensionality of the construct 3. Translated scale 13. Original and translated 23. Adapted scale (translation + dimensionality)	Numerical
10	Validation study	1. No 2. Yes, reliability only 3. Yes, with exploratory factorial methods 4. Yes, with confirmatory factorial methods	Numerical
11	FIT-Choice Language		Chain
12	Type of teacher	1. Early years 2. Primary 3. Secondary 4. Language teacher 12. Early years and primary 13. Early years and secondary 23. Primary and secondary 123. Early years, primary and secondary	Numerical

I3	Country	In studies that compare various countries, each one is included as one sample.	Chain
I4	Sample size	Total number of cases	Numerical
I5	n women	Frequency or percentage	Numerical
I6	Age	Average age of sample	Numerical
I7	Teaching activity	1. Pre-service teacher 2. Practising teacher	Numerical

The final selection includes information from over 29,640 teachers (7.7%) and pre-service teachers (92.3%). The mean age is 26.18 years and 66% of them are female. These works were published in the 2007–2020 period, with 2012 and 2019 being the years with the most works (37.5% and 16.7% respectively). Most of the documents are articles (87.5%), followed by doctoral theses (8.3%) and books (4.2%). More than 93% of the works ( $n = 45$ ) are written in English, compared with 6% in Spanish ( $n = 3$ ). The studies were mainly carried out in the USA (18.8%), Switzerland (12.5%) and Turkey (10.4%).

## Data analysis plan

Firstly, as the normality of the distribution of the alpha coefficients across the different studies cannot be assumed, we applied the transformation proposed by Bonett (2010):

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

Where  $\alpha_i$  is the reliability coefficient of each study  $i$ . The variance of this transformation is estimated using formula 2:

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

Where  $q$  is the number of items on the scale in the study  $i$  and  $n$  the effect size used. The transformed results are then weighted by the inverse of the variance, which in a random effects model has two terms:  $Var(L_i)$  and  $\tau^2$ , that is, the intra- and inter-study variability. As variance between studies is an unknown parameter, it is estimated based on the information about reliability provided by the research works analysed by

the Dersimonian-Laird method (Sánchez-Meca et al., 2013). Finally, once weighted, to facilitate their interpretation as alphas, the transformation is reversed.

We tested homogeneity using Cochran's Q test (Hedges & Olkin, 1985) and calculated the proportion of heterogeneity using  $I^2$  (Higgins & Thompson, 2002), which is classed as high above 0.75.

We calculated the effect of moderator variables using meta-regression. Finally, we tested for publication bias using Egger's regression test, where non-significant values indicate an absence of bias.

## Moderator variables

To study the effect of contextual variables on the variability of the reliability coefficients, we used the following variables from Table I as moderators: 9, 10 and 11 (relating to adaptations to the instrument) and 12, 15, 16 and 17 (relating to the sample). With the objective of verifying the independent effects of the groups of cases represented in the different moderator variables, we dichotomised them so that they could be included as separate predictors in the regression analysis. All of the analyses were done using the *Jamovi* software program (2020) and the *metafor* package (Viechtbauer, 2010) for R.

## Results

In response to the first specific objective, the selected studies and their characteristics are summarised in Table II

**TABLE II.** Studies finally included after the systemic review process

ID	Study	no. of samples	Type of scale	Validation study	Type of teacher	Teaching activity	Fit-Choice language
43	Alpaslan et al. (2018)	1	2	1	3	1	English
61	Akpochafo (2020)	1	23	1	23	1	Turkish
21	Berger & D'Ascoli (2012a)	2	3	4	3	2	German and French
105	Berger & D'Ascoli (2012b)	2	23	1	3	1	German and French
146	Berger & Girardet (2020)	1	3	4	3	2	English
81	Choi S., & Song J. (2015)	1	23	1	23	1	English
125	de Zordo et al. (2019)	1	3	1	12	1	English
59	Eren & Tezel (2010)	1	3	4	4	1	Turkish
13	Fokkens-Bruinsma & Canrinus (2012a).	1	23	4	2	1	Dutch
133	Fokkens-Bruinsma & Canrinus (2012b).	1	3	2	3	1	Dutch
86	Fokkens-Bruinsma & Canrinus (2014)	2	3	1	23	1	Dutch
63	Goller et al. (2019)	2	3	4	12	1	German and Finnish
143	Gratacós & López-Jurado (2016)	1	3	4	12	1	Spanish
23	Heinz et al. (2017)	1	1	1	3	1	English
2	Hennessy & Lynch (2017)	1	2	3	3	1	English
46	Jablanovic & Vracar (2019)	1	1	1	123	1	English
85	Jugovic et al. (2012)	1	3	4	2	1	Croatian
58	Kilinc et al. (2012)	1	3	4	23	1	Turkish
102	Konig & Rothland (2012)	1	3	4	23	1	German

51	Lawver (2009).	1	1	1	3	1	English
15	Lawver (2012)	1	1	1	3	1	English
76	Lin et al. (2012)	2	13	4	23	1	English and Chinese
26	MacKenzie (2013)	1	1	1	23	1	English
91	Nesje et al. (2018)	1	3	4	23	1	Norwegian
164	Ozturk-Akar (2012)	1	3	4	3	1	Turkish
14	Ozturk-Akar (2019)	1	23	1	3	1	Turkish
96	Ponnock (2018)	1	2	4	123	1	English
80	Ridgewell (2018)	1	1	4	2	2	English
48	Said-Hung et al. (2018)	1	3	1	23	1	Spanish
3	Silvestre et al. (2020)	1	3	4	123	1	Spanish
126	Suryani et al. (2016)	1	23	4	123	1	Indonesian
128	Taimalu et al. (2017)	1	23	4	123	1	Estonian
165	Topkaya & Uztosun (2012)	1	2	4	4	1	English
138	Torsney et al. (2019)	1	2	4	2	1	English
94	Van Overschelde & Garza (2019)	1	2	4	3	1	English
75	Wang (2019)	1	1	4	23	1	English
97	Watt & Richardson (2007)	2	13	4	23	1	English and German
101	Watt et al. (2012)	4	1	4	123	1	English
154	Zhang et al. (2020)	1	23	4	4	1	Chinese

The majority of the research works analyse a single sample (82%). The rest compare the results of 2 groups (15.4%) or in one case four different samples. With regards to educational stages, secondary education teachers participate in 30.8% of the studies, primary and secondary education teachers in 28.2%, teachers from all three stages (early years, primary and secondary) in 15.4% and primary education teachers in 10.3%. In

combination with this, around 7.7% of the studies refer to mixed samples of early years and primary and 7.7% to language teachers.

The texts use samples mainly comprising pre-service teachers (92.3%). Only three articles (7.7%) include active teachers.

In 20.5% of the works, the original version of Fit-Choice was used (Watt & Richardson, 2007). A further 5% used the original version along with a translation, compared with 15.4% where an adaptation of the dimensionality of the construct is carried out, 38.4% in which the language is adapted and 20.5% in which both dimensionality and language are adapted. Of the studies, 66.6% perform some kind of psychometric study, with the majority (61.5%) applying confirmatory factorial techniques. The selected research works cover the use of the scale in 12 different languages, with English (41.7%), German (10.4%), Turkish (10.4%) and Estonian (10.4%) being most frequent.

With regards to the **second specific objective**, Table III presents the results of the meta-analysis for the 18 subdimensions, the four second-order factors and the motivation and belief constructs that correspond to the original FIT-Choice structure.

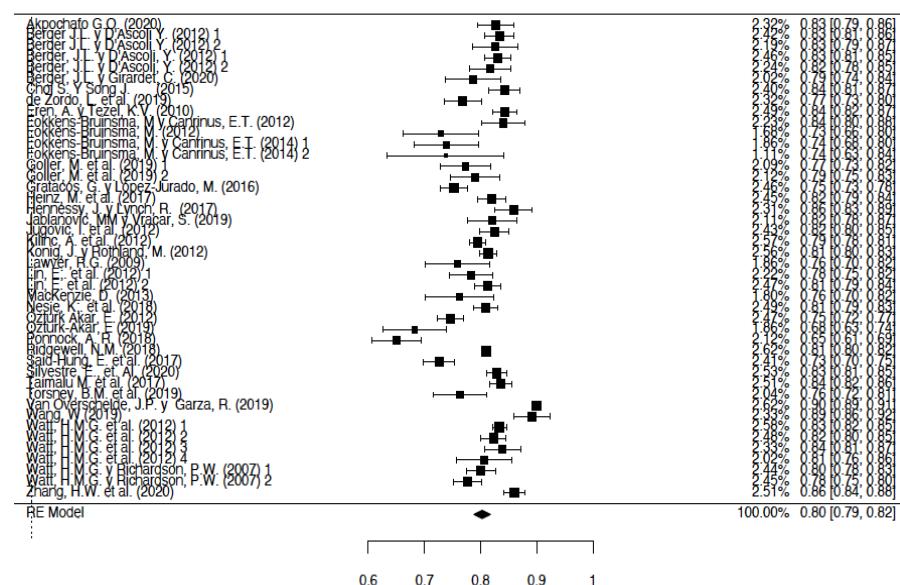
TABLE III. Results of the meta-analysis of generalisation of reliability by subdimensions, dimensions and constructs of FIT-Choice

	<b>k</b>	<b>F</b>	<b>SE</b>	<b>Z</b>	<b>P</b>	<b>LL</b>	<b>UL</b>	<b>I<sup>2</sup></b>	<b>Q</b>	<b>p</b>	Reliability			Heterogeneity			Egger's Regression		
											<b>Value</b>	<b>P</b>	<b>Value</b>	<b>P</b>	<b>Value</b>	<b>P</b>			
M1 Perceived teaching ability	40	0.814	0.006	129	<.001	0.801	0.826	85.73%	273.256	<.001	-1.783	0.075							
M2 Job security	35	0.829	0.008	98.9	<.001	0.813	0.846	96.09%	870.524	<.001	-0.293	0.77							
M3 Time for family	35	0.83	0.007	125	<.001	0.817	0.843	91.31%	391.334	<.001	-0.508	0.611							
M4 Job transferability	22	0.674	0.02	34.1	<.001	0.636	0.713	94.13%	357.953	<.001	-0.462	0.644							
M5 Intrinsic value	39	0.741	0.015	48.4	<.001	0.711	0.771	97.06%	1293.724	<.001	-2.14	0.032							
M6 Fallback career	32	0.654	0.018	36.2	<.001	0.618	0.689	90.89%	340289	<.001	-0.624	0.533							
<b>Personal Utility (2nd order)</b>	<b>11</b>	<b>0.817</b>	<b>0.019</b>	<b>42.5</b>	<b>&lt;.001</b>	<b>0.779</b>	<b>0.855</b>	<b>96.18%</b>	<b>261.71</b>	<b>&lt;.001</b>	<b>0.82</b>	<b>0.412</b>							
M7 Shape future	24	0.788	0.014	56.7	<.001	0.76	0.815	97.84%	1062.58	<.001	0.345	0.73							
M8 Social equity	33	0.814	0.01	83.3	<.001	0.795	0.833	96.71%	973.112	<.001	-1.05	0.294							
M9 Social contribution	32	0.796	0.009	84	<.001	0.778	0.815	96.07%	788034	<.001	-0.926	0.354							
M10 Working with children	37	0.891	0.005	189	<.001	0.882	0.901	95.83%	863066	<.001	-1.921	0.055							
M11 Prior teaching-learning experience	38	0.846	0.006	146	<.001	0.835	0.858	94.62%	687562	<.001	-0.738	0.46							
M12 Social influences	40	0.837	0.011	73.4	<.001	0.814	0.859	97.18%	1384.538	<.001	0.895	0.371							
<b>Social Utility (2nd order)</b>	<b>12</b>	<b>0.824</b>	<b>0.018</b>	<b>45.9</b>	<b>&lt;.001</b>	<b>0.788</b>	<b>0.859</b>	<b>97.37%</b>	<b>418608</b>	<b>&lt;.001</b>	<b>-0.629</b>	<b>0.529</b>							
<b>MOTIVATIONS (construct)</b>	<b>44</b>	<b>0.803</b>	<b>0.007</b>	<b>109</b>	<b>&lt;.001</b>	<b>0.788</b>	<b>08.17</b>	<b>93.81%</b>	<b>694197</b>	<b>&lt;.001</b>	<b>-0.442</b>	<b>0.659</b>							
C13 Expert career	29	0.775	0.001	77.9	<.001	0.756	0.795	96.12%	722167	<.001	-2028	0.043							
C14 High demand	33	0.693	0.015	46.5	<.001	0.663	0.722	96.28%	859354	<.001	-2001	0.045							
<b>High demand (2nd order)</b>	<b>8</b>	<b>0.778</b>	<b>0.037</b>	<b>19.7</b>	<b>&lt;.001</b>	<b>0.655</b>	<b>0801</b>	<b>97.03%</b>	<b>235922</b>	<b>&lt;.001</b>	<b>1.45</b>	<b>0.147</b>							
C15 Social status	35	0.853	0.005	167	<.001	0.843	0.863	94.35%	601848	<.001	-3088	0.002							
C16 Good salary	32	0.878	0.008	105	<.001	0.862	0894	98.51%	2087.421	<.001	1817	0.069							
<b>Task return (2nd order)</b>	<b>5</b>	<b>0.755</b>	<b>0.044</b>	<b>17</b>	<b>&lt;.001</b>	<b>0.668</b>	<b>0842</b>	<b>98.56%</b>	<b>278595</b>	<b>&lt;.001</b>	<b>0652</b>	<b>0.514</b>							
C17 Social dissuasion	36	0.711	0.011	65.9	<.001	0.69	0.732	91.77%	425129	<.001	0.237	0.813							
C18 Satisfaction with choice	31	0.879	0.007	124	<.001	0.865	0893	98.43%	1913.813	<.001	-2844	0.004							
<b>BELIEFS (construct)</b>	<b>37</b>	<b>0.783</b>	<b>0.008</b>	<b>98.1</b>	<b>&lt;.001</b>	<b>0.767</b>	<b>0799</b>	<b>95.06%</b>	<b>728631</b>	<b>&lt;.001</b>	<b>-1755</b>	<b>0.079</b>							

k = number of studies; F = average reliability; SE = standard error; LL = lower limit; UL = upper limit; P = associated probability

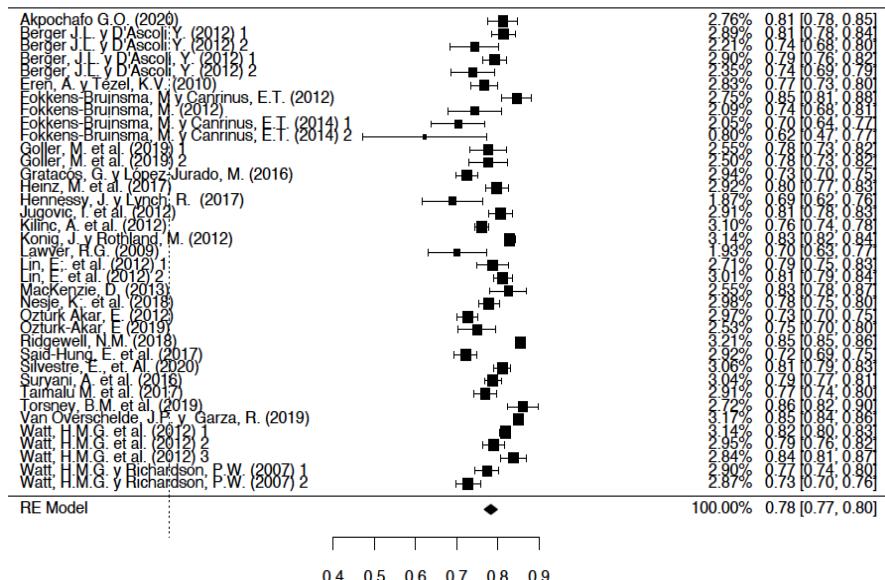
Of the 12 subdimensions that relate to teacher motivation (from M1 to M12), 7 (58.33%) display good reliability, 3 (25%) have acceptable values and the reliability of the other two (16.66%) is questionable, that is to say, it is below .7 (M4 and M6). In any case, it is worth noting that in general terms the “personal utility” and “social utility” dimensions and the “motivations” construct have reliability values that can be regarded as good (mean reliability of 0.8). Figure III shows the average reliabilities of each study that provides data relating to the motivations construct, along with their respective confidence intervals.

**FIGURE III.** Forest plot of the “motivations” construct



In the case of the “beliefs” construct, average reliability is close to 0.8. The “high demand” dimension and, in particular, one of its subdimensions (C14) have the worst psychometric properties. For their part, the reliabilities of the C15 and C16 subscales associated with the “task return” dimension can be regarded as good. Figure IV shows the variability of the estimated reliability for this construct in the different studies that provide data in this regard.

**FIGURE IV.** Forest plot of the “beliefs” construct



Regarding the heterogeneity in the psychometric properties reported by the different studies, the  $I^2$  values, which are greater than 75% in all of the subdimensions and dimensions of the scale, and the probabilities associated with the Q statistics, which are lower than 0.01, reflect significant differences in the reliability values obtained when using FIT-Choice with different groups of subjects. This heterogeneity is shown in Figures III and IV, where the intervals for the average reliabilities of the scales are easily identifiable.

It is also important to note that possible publication bias, evaluated using Egger's regression test, occurs in five subdimensions (M5, C13, C14, C15, and C18) and in the second-order factors. This problem primarily affects the beliefs scales and so these results should be interpreted with caution.

To achieve the **third specific objective**, we estimated models that include moderator variables linked to the characteristics of the instrument and to the samples used. Firstly, the effect of the dispersion of the scores on the reliability results is analysed. This dispersion affects reliability

in M5 ( $Z=2.93$ ;  $p=0.003$ ), M8 ( $Z=3.17$ ;  $p=0.02$ ), M11 ( $Z=2.30$ ;  $p=0.021$ ), C13 ( $Z=2.37$ ;  $p=0.018$ ) and C15 ( $Z=3.24$ ;  $p=0.001$ ). Next, the effect of the type of FIT-Choice and of the samples is considered. In the first case, the results are presented below by subdimensions:

- In subscale M2, the test translated into German has a positive effect ( $Z=2.04$ ;  $p=0.041$ ).
- In M5, use of a translated version of the instrument does not appear to be problematic ( $Z=2.48$ ;  $p=0.013$ ). Even the test in its original language (English) displays a negative effect in reliability ( $Z=-2.03$ ;  $p=0.043$ ). In the works that use this scale without the aim of performing some type of validation of the instrument, there is a negative effect ( $Z=-2.26$ ;  $p=0.024$ ), including when the work attempts to validate the instrument using only a reliability study ( $Z=-2.10$ ;  $p=0.036$ ). In contrast, in studies that use confirmatory analysis techniques to study validity, the effect is positive ( $Z=2.90$ ;  $p=0.004$ ).
- In subdimension M6, in works that do not include validation there is a negative effect on reliability ( $Z=-2.23$ ;  $p=0.026$ ).
- In M7, the change of construct with regards to the original model has a negative effect ( $Z=-2.26$ ;  $p=0.024$ ) and use of the test in the original language displays a positive effect on reliability ( $Z=2.74$ ;  $p=0.006$ ).
- In M8, using the test in the original language displays a positive effect ( $Z=2.07$ ;  $p=0.039$ ).
- In M10, the change of construct with regards to the original model has a positive effect ( $Z=2.26$ ;  $p=0.024$ ). If the study is not a validation study, the effect is negative ( $Z=-4.61$ ;  $p<0.001$ ). This is also the case if a translated version of the instrument is used ( $Z=-2.03$ ;  $p=0.043$ ).
- In subscale C16, the translation into German has a positive effect ( $Z=2.61$ ;  $p=0.009$ ). In contrast, the Spanish translation has a negative effect on reliability ( $Z=-2.09$ ;  $p=0.036$ ).
- And in C18, a negative effect is observed ( $Z=-3.57$ ;  $p<0.001$ ) if the work does not have the aim of validating the instrument. In contrast, confirmatory studies have shown a positive effect on reliability ( $Z=3.78$ ;  $p<0.001$ ).

Finally, in the personal utility second order factor, the test translated into German has a positive effect ( $Z=2.53$ ;  $p=0.011$ ).

In addition, to present the sample characteristics that can affect the variability, the results are grouped by variables.

The teacher training specialism displays significant effects on the reliability of some scales. Primary school teachers have a positive effect on the reliability of the beliefs scale ( $Z=2.17$ ;  $p=0.03$ ). The sample of secondary school teachers determines the reliability results for subscale M4 ( $Z=-2.37$ ;  $p=0.018$ ), M10 ( $Z=-2.52$ ;  $p=0.01$ ) and C18 ( $Z=-2.99$ ;  $p=0.003$ ). The sample of language teachers has a positive effect on M6 ( $Z=2.69$ ;  $p=0.007$ ), M10 ( $Z=2.58$ ;  $p<0.01$ ) and C17 ( $Z=2.62$ ;  $p=0.009$ ). In contrast, the effect is negative on the second order factor of social utility ( $Z=-2.54$ ;  $p=0.011$ ) and on subdimension C13 ( $Z=-3.59$ ;  $p<0.001$ ).

Using a mixed sample of teachers from early years and secondary has a negative effect on the results of subscale M6 ( $Z=-2.04$ ;  $p=0.042$ ) and the second order factor of task demand ( $Z=-2.47$ ;  $p=0.013$ ). Using a mixed sample of teachers from early years and primary has a negative effect on M7 ( $Z=-3.18$ ;  $p=0.001$ ), M8 ( $Z=-2.16$ ;  $p=0.031$ ) and M9 ( $Z=-2.58$ ;  $p=0.01$ ). The mixed sample of primary and secondary has a negative effect on M6 ( $Z=-2.18$ ;  $p=0.030$ ) and a positive effect on the second order factors of personal utility ( $Z=1.98$ ;  $p=0.048$ ) and social utility ( $Z=2.52$ ;  $p<0.012$ ), on the general motivations construct ( $Z=2.02$ ;  $p<0.044$ ) and on the estimation of the total reliability for the beliefs scale ( $Z=2.96$ ;  $p=0.003$ ) as well as on subscale C18 ( $Z=3.11$ ;  $p=0.002$ ).

The sample comprising students has a negative effect on reliability in subscales M6 ( $Z=-2.78$ ;  $p=0.05$ ), M8 ( $Z=-2.05$ ;  $p=0.041$ ), C16 ( $Z=-2.09$ ;  $p=0.036$ ) and C18 ( $Z=-2.03$ ;  $p=0.042$ ).

Finally, the age of participants has a positive effect on variables M7 ( $Z=2.41$ ;  $p=0.016$ ), M8 ( $Z=3.17$ ;  $p=0.02$ ), M9 ( $Z=2.11$ ;  $p=0.034$ ) and C16 ( $Z=2.03$ ;  $p=0.043$ ); it negatively affects the second-order social utility factor ( $Z=-2.50$ ;  $p<0.012$ ) and the high demand factor ( $Z=-2.50$ ;  $p=0.012$ ).

## Discussion and conclusions

Given the current concern about teachers' efficacy, it is of interest to establish whether similar factors explain the choice of teaching as a career in all countries. To do this, measurement instruments are needed that have validated reliability in international studies and therefore allow this comparison. FIT-Choice, designed in 2007 and initially validated for

a sample of 1653 Australian students (Watt & Richardson, 2007) has been confirmed to be a useful tool for this objective following its long history of applications.

The present study has set out to analyse, compare and synthesise quantitatively the reliability of the FIT-Choice scale following its use with various international samples to measure the motivations and beliefs of pre-service teachers and practising teachers. Given that it is an extensively used instrument, it is appropriate to reflect on its internal consistency, through a systematic review and a meta-analysis of generalisation of its reliability. The results obtained reflect significant heterogeneity in its reliability, which is adequate on the whole, although differential psychometric behaviour is observed in some of its dimensions and subdimensions. In general terms, the personal utility and social utility dimensions and the motivations construct have reliability values that can be regarded as good (mean reliability of 0.8). In contrast, the beliefs construct and its subdimensions have a mean reliability close to 0.8, except for subscale C14, high demand, which is below 0.6.

The creators of the questionnaire argue that the instrument allows for international comparison of motivations for pursuing a career in teaching. Watt et al. (2012) analyse results from Australia, the USA, Norway and Germany and conclude that the dimensions structure is stable between countries, although beliefs linked to the profession vary considerably. Their findings led them to reject the “job transferability” and “fallback career” subscales (M4 and M6) owing to their low reliability. The results of this meta-analysis show that the personal utility scale and good salary have greater reliability in the German version of the instrument. Nonetheless, the estimation of the good salary of the profession is less reliable in the Spanish version, which is also used in Latin America.

The results indicate that the scale relating to choosing this career to help social equity is more reliable in the original English version, although it is also more affected by the dispersion of scores. This could suggest that the scale has different meanings according to the context, and so it should be adapted to each situation. This limitation already appears in the previous meta-analyses. One possible explanation, as the authors of expectancy-value theory (Klassen et al., 2011) noted, is that quantitative instruments can conceal cultural differences in these motivations because they assume that the ones that are dominant in Western countries are in other settings as well. In future, it would be

advisable to perform a systematic review of works that consider the theory that explains these differences between cultures or types of teacher. It would also be desirable to perform a meta-analysis of differences of means between groups (the West compared with others), educational level (early years, primary, secondary), pre-service teachers compared with practising teachers and educational stage. Furthermore, the sample of studies could be expanded by including works that apply other types of instrument.

One complicating factor when studying the overall reliability of FIT-Choice is the dimensionality of the construct. Contextual differences could be the root of the variations in the structure of dimensions observed, which would require further study. These modifications normally relate to the combination of scales and at other times involve the addition of new dimensions. For example, the work of Suryani et al. (2016) incorporates a new religious influences scale and another to identify dissuasion caused by the media. Van Overschelde and Garza (2019) include a motivational dimension of the choice of the profession to contribute to social change, combining items from the social utility value scales; and Zhang et al. (2020) define an extrinsic value scale. Akpochafo's work (2020) groups together the skill and intrinsic motivations dimensions on the one hand and contextual antecedents with scales of motivations on the other (such as prior teaching-learning experiences with choosing the profession to make a social contribution or the social influence with job security). The works by Berger and D'Ascoli (2012a & 2012b) combine scales that relate to social utility, such as helping shape the future or making a social contribution, while also adding new dimensions such as constructivist beliefs and direct transmission beliefs. Overcoming the limitation of the heterogeneity observed in the dimensionality of the construct raises the possibility of a meta-analysis of construct validity as future work.

Despite the existing difficulties, with the meta-analysis carried out, complemented by the study of the effect of the selected moderator variables, this work represents an important step towards elucidating the metric properties of FIT-Choice, which is an instrument with international scope. It also makes a contribution on the line deriving from the meta-analysis by Heinz (2015), which recommends greater depth in comparative studies of motivations for choosing a career in teaching as these could vary by profile. The average reliability figures found support the usefulness of the questionnaire and its value not only

as a tool for comparative study of the motivations and beliefs involved in the choice of teaching as a profession in international studies but also as an appropriate instrument for research into the dimensionality of a construct that is as complex as it is interesting.

## References

- 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>
- Akpochafio, 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. & 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. & Mourshed, M. (2007). *How the world's best-performing school systems come out on top*. McKinsey y Co.
- Berger J.-L. & 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. & 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. & 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 Gambara, 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. & 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. & 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. & 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. & 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. & 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. & 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. & 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. & 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., Vahasantanen, K., Festner, D. & 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. & 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. & López-Jurado, M. (2016). Validation of the Spanish version of the Factors Influencing Teaching (FIT)-Choice scale. *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. & 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. & 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. & 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. & 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. & 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. & 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. & 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. & 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. & 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. & 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. & 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. & 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. & 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. & 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. & 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. & 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. & Valencia, J. (2017). Factors affecting the choice of teaching programs in Colombia. *Educação e Pesquisa*, 43, 31-48. <https://doi.org/10.1590/s1517-9702201701160978>
- Sánchez-Meca, J., López-López, J. A. & 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. & Díaz, J.V. (2020). Why ISFODOSU Pedagogical University students choose the education career? *Ciencia y Educación*, 1, 47-69. <https://doi.org/10.22206/cyed.2020.v4i1.pp47-69>
- Suryani, A., Watt, H.M.G. & 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. & Täht, K. (2017). Teaching motivations and perceptions during the first year of teacher education in Estonia. In H.M.G. Watt, P.W. Richardson, & K. Smith (Eds.) *Global Perspectives on Teacher Motivation* (pp. 189-219). Cambridge University Press.
- Topkaya, E. & 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. & 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. & 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. & 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. & 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. & 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 & 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>

**Contact address:** 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

