



VOL.24, Nº2 (Julio, 2020)

ISSN 1138-414X, ISSNe 1989-639X

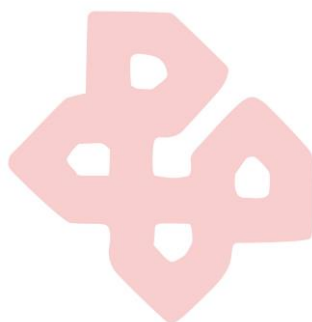
DOI: 10.30827/profesorado.v24i2.15040

Fecha de recepción: 14/09/2017

Fecha de aceptación: 12/11/2018

DISEÑO Y VALIDACIÓN DE UN INSTRUMENTO PARA ANALIZAR EL SISTEMA DE EVALUACIÓN DE LAS GUÍAS DOCENTES UNIVERSITARIAS EN LA FORMACIÓN INICIAL DEL PROFESORADO (IASEG)

Design and validation of an instrument to analyze the assessment system of the university's course guides at the initial teacher training (IASEG)



María Rosario Romero-Martín, Sonia Asún Dieste y Miguel Chivite Izco

Universidad de Zaragoza

E-mail: rromero@unizar.es; sonasun@unizar.es; mchivite@unizar.es

ORCID ID: <http://orcid.org/0000-0002-5019-4574>

<https://orcid.org/0000-0002-3761-2903>

<http://orcid.org/0000-0002-1218-5351>

Resumen:

La evaluación formativa ha sido considerada como un elemento altamente coherente en el diseño de programas docentes universitarios basados en competencias. A partir de esta concepción formativa de la evaluación, el objetivo de este estudio se ha centrado en alcanzar un consenso de expertos para crear y validar un instrumento que permita valorar la eficacia de los sistemas de evaluación, presentes en las guías universitarias de la formación inicial del profesorado, para detectar el grado de adquisición de competencias docentes. Se utilizó para ello el método Delphi. Inicialmente, a partir de una exhaustiva revisión bibliográfica y de un ciclo de juicios de expertos,

se identificaron las dimensiones y elementos constitutivos de los sistemas de evaluación, creándose un primer instrumento para el análisis de los mismos. Posteriormente, el instrumento se envió en dos rondas al grupo de expertos, mediante el que se llegó a consensos óptimos sobre el nivel de eficacia de los distintos elementos de los sistemas de evaluación para valorar competencias docentes. Finalmente, se efectuó una prueba piloto en la que tres evaluadores analizaron una muestra de guías docentes aplicando el instrumento. El grado de acuerdo entre expertos para la delimitación de los ítems y dimensiones, la compactibilidad hallada durante el análisis Delphi y la concordancia de evaluadores en la prueba piloto, arrojaron datos que permitieron concluir la validez del instrumento diseñado.

Palabras clave: competencias docentes; educación superior; evaluación educativa; evaluación formativa; técnica Delphi.

Abstract:

The formative assessment has been considered as an element highly coherent in the design of the teachers programs based on competencies at university. According to this formative assessment concept, the aim of this study has focused on reaching a consensus of experts to create and validate an instrument which provides to judge the effectiveness of assessment's systems, present in the courses guides of the Teaching training at University, to detect teaching skills. The Delphi method was used for this purpose. Initially, from a comprehensive literature review and a cycle of experts' judgement, they were identified the dimensions and elements of assessments' systems, creating a first instrument for the analysis of the same. Subsequently, the instrument was sent in two rounds to the Group of experts, which found an optimal consensus on the level of effectiveness of the various elements of the assessments' systems to assess teaching skills. Finally, a pilot test was carried out; three evaluators analyzed a sample of courses guides applying the instrument. The degree of agreement among experts for the delimitation of the items and dimensions, the compactness found during the Delphi analysis and concordance of evaluators in the pilot test, produced data of the direct evidence the validity of the instrument designed.

Key Words: Delphi technique; educational assessment; formative evaluation; higher education; teaching skills.

1. Presentación y justificación del problema

A lo largo de los últimos años, en los diversos foros y publicaciones sobre Educación Superior, han sido recurrentes las referencias a los cambios organizativos que, además de los estructurales y curriculares, se exigen a instituciones y a agentes educativos en la implantación del Espacio Europeo de Enseñanza Superior (Mateo, 2000; López, Pérez-García & Rodríguez, 2015). También se hace referencia a la necesidad de evaluar los resultados de los esfuerzos realizados por integrar dichos cambios en los sistemas universitarios utilizando mecanismos de evaluación del propio modelo (Aramburuzabala, Hernández-Castilla & Ángel-Ubide, 2013). Entre ellos, cabe destacar la importancia del desarrollo y afianzamiento de sistemas de evaluación para que sean garantes de una buena calidad pedagógica (Pérez, 2016).

Diversos autores, y en distintas etapas de este proceso, han coincidido en considerar la evaluación como la clave del cambio (Buscá, Pintor, Martínez & Peire, 2010; Medland, 2016; Nunziatti, 1990; Trillo & Porto, 1999; Zabalza, 2001). En particular, se ha considerado necesario evaluar el proceso, y como parte de él, los

programas y la docencia, para poder garantizar la calidad de la labor desempeñada por la institución y su profesorado (Goby & Nickerson, 2014; Muñoz, Ríos de Deus & Abalde, 2002).

Los sistemas de evaluación que se implementan en la educación superior aparecen hoy descritos en las guías docentes elaboradas por los profesores. Conviene subrayar que los planes de estudios de los grados y los másteres se desarrollan en unos documentos oficiales de las universidades denominados proyectos de titulación, y también es en ellos, donde se determina de modo sintético el sistema de evaluación de las asignaturas que pertenecen a dicho plan. Al mismo tiempo, y a partir siempre de este proyecto de titulación, los profesores de la universidad deben explicitar con mayor detalle los sistemas de evaluación en las guías docentes de sus asignaturas. Así pues, los elementos de los denominados proyectos docentes o programas de las asignaturas que diseñaban los profesores, se han venido integrando en las actuales guías docentes, es decir, en el documento que oficialmente debe ser publicado por las Universidades para informar a los estudiantes universitarios sobre su proceso de aprendizaje así como de su evaluación en cada una de las asignaturas. Sin embargo, y a pesar de que se ha detectado que los profesores universitarios valoran los sistemas de evaluación como muy importantes y que dedican muchos esfuerzos a su realización (Villa, Campo, Arranz, Villa & García, 2013), la literatura sobre este objeto de estudio es hoy limitada por lo que nos referiremos tanto a programas como a guías docentes para revisar el estado de la cuestión.

1.1. Evaluación de la calidad de la enseñanza universitaria: la evaluación de programas

A pesar de la importancia que se le atribuye a la programación de la enseñanza de la docencia universitaria, su investigación ha tenido una escasa presencia en la investigación desarrollada en nuestro país (Lorente, Montilla & Romero, 2013). Expósito, Olmedo y Fernández (2004) encontraron 208 estudios centrados en evaluación de programas educativos, y hallaron que tan sólo el 14.6% pertenecía al ámbito universitario. A pesar de ello, en el ámbito de la innovación docente universitaria, han ido surgiendo numerosos trabajos dedicados a la temática (Chivite, Asún & Romero, 2011; González-Fernández, 2007; Grau & Gómez, 2010; López-Pastor, 2009; Lorente & Kirk, 2013; Santos, Martínez & López-Pastor, 2009; Tortosa & Álvarez, 2009), lo que muestra un creciente interés en estudiar lo que realmente acontece en la enseñanza universitaria, intentando superar ese “escenario opaco”, que para Zabalza (2008, p.200), había venido existiendo en la universidad.

1.2. Evaluación de guías docentes

De las distintas dimensiones de la programación docente, Yániz (2006) diferencia entre los conceptos *proyecto docente*, entendido como un documento para el profesor y los equipos que comparten un proyecto en una titulación y, *guía de aprendizaje* (hoy comúnmente denominada *guía docente*), definida como documento informativo donde se muestran los compromisos y las exigencias, que se establecen con el alumno que cursa una titulación universitaria, consistente en una selección de

los elementos del proyecto más relevantes y significativos: competencias y resultados de aprendizaje, actividades detalladamente explicadas, contenidos, sistema de evaluación, guiones y consignas para los distintos trabajos y actividades que deben realizar a lo largo del curso y tiempo previsto para las tareas y calendario.

No obstante, la guía docente no se limita a la reunión de distintos elementos constitutivos sino que debe caracterizarse por la coherencia interna entre todos ellos (Gijbels, Watering, & Dochy, 2005; Segers & Dochy, 2001) de manera que las piezas encajen para que el engranaje funcione como un todo, si se pretende obtener buenos resultados como se traduce de la idea del alineamiento constructivo de Biggs (2005). Así, la constatación de incoherencias entre los componentes básicos, puede ser el punto de partida de la toma de decisiones orientadas a la mejora, en el contexto más amplio de la institución universitaria (Orden et al., 1997).

Diversos estudios han señalado, que considerando el momento de cambio de las universidades hacia un enfoque formativo en el que el alumno se convierte en el eje central y se programa a partir de la definición de las competencias, la evaluación, en coherencia, debe centrarse en el valor formativo y, no tanto, en la cultura de la evidencia con su capacidad sancionadora (Romero; Fraile, López-Pastor & Castejón, 2014; Shavelson, 2010; Trevitt, Berman & Stocks, 2012) ya que es la evaluación formativa la que orienta y reorienta el proceso de enseñanza-aprendizaje para que el alumnado logre mejores aprendizajes y resultados académicos (Torrance, 2012). Sánchez-Santamaría (2011) expresa la diferencia entre dos enfoques a la hora de evaluar las competencias: el primero, como evaluación superficial y el segundo como evaluación profunda. Este último enfoque es, sin duda, en el enmarcamos nuestra propuesta, es decir, y en palabras de este autor, una evaluación formativa (retroalimentación efectiva), compartida (proceso dialógico y toma de decisiones colectiva) y continua (producto por etapas vs finalista) y cuyo objeto es valorar lo que el alumno sabe, hace o sabe hacer en una situación auténtica.

1.3. Los sistemas de evaluación en las guías docentes

De todos los aspectos que deben recoger las guías docentes, el conjunto de precisiones sobre el qué y el cómo evaluar se denomina sistema de evaluación, el cual, para Ibarra y Rodríguez (2010), consiste en un conjunto de criterios, normas y procedimientos; comentan que en los últimos años son cada vez más frecuentes las publicaciones referidas al sistema de evaluación, siendo este asunto un elemento estratégico en la universidad (Gibbs, 2003; Partington & Brown, 2013). Toda la literatura parece indicar que en el nuevo contexto universitario es preciso diseñar procedimientos de evaluación adecuados para valorar el grado de adquisición de las competencias, aplicar técnicas diversas en función de los diferentes tipos de trabajo del estudiante, y aplicar otros procedimientos de evaluación además de los exámenes tradicionales (Muñoz-San Roque, Martín-Alonso, Prieto-Navarro & Urosa-Sanz, 2016; Rojo et al., 2009). Pero la cuestión sobre el modelo de evaluación que se considera óptimo para valorar y desarrollar competencias en los estudiantes universitarios es compleja (Medina, Domínguez & Sánchez, 2013). Para Hall y Burke (2003) o Kaftan, Buck y Haack (2006), entre otros, el modelo es la evaluación formativa, lo cual deberá tenerse en cuenta en los diseños de evaluación, para, como indica Bolívar

(2008), poder obtener información sobre producciones complejas que impliquen conocimiento, actitudes y pensamiento metacognitivo, así como estratégico, en los estudiantes.

1.4. Análisis de los sistemas de evaluación

Para analizar los sistemas de evaluación de las guías docentes existen diferentes modelos que podríamos situar en un continuum que iría desde lo más objetivo a lo más subjetivo. Según Vázquez, Fera, Moriel y Tijerina (2005), los modelos objetivos se corresponden con la medición y los subjetivos con la opinión de expertos; también se plantean un modelo intermedio en el que se pasa del criterio de un experto al acuerdo entre varios, siendo imprescindible, para asegurar la validez, que ese acuerdo sea coherente con el modelo teórico del objeto de estudio que habrá debido ser definido previamente. En el estudio de Expósito et al. (2004) sobre evaluaciones de programas educativos, el 46,1% se habían hecho desde modelos objetivistas; 21.6% con subjetivistas; el 10.1% con modelos críticos o no identificados; y el 22.1% con modelos eclécticos siendo estos últimos semejantes al modelo intermedio planteado posteriormente por Vázquez et al. (2005). También García-Bellido, Jornet y González-Such (2012) plantean su estudio con una complementariedad metodológica, cuantitativa/cualitativa; y en la misma línea se pronuncian Onwuegbuzie, Bustamante y Nelson (2010).

Por otro lado, Expósito et al. (2004) constataron que la mayoría de los estudios analizados sobre evaluación de programas educativos empleaban instrumentos que no estaban estandarizados sino que se elaboraban específicamente para el estudio en cuestión; y lo mismo expresaban Rodríguez-Gómez, Ibarra-Sáiz, Gallego-Noche, Gómez-Ruiz y Quesada-Serra (2012) en su estudio. Sea como fuere, para que la evaluación sea rigurosa estos instrumentos deberían cumplir criterios científicos de validez y confiabilidad (Casanova, 1995; Inciarte & Cánquiz, 2001).

En este sentido, cabe destacar el nuevo diseño y validación de la Escala de Evaluación de los aprendizajes (EEA) propuesta por Hebles, Dos Santos, Yaniz y Villardon-Gallego (2017) que corrobora esa necesidad de centrarse en el estudio de los sistemas de evaluación universitarios, en este caso desde la óptica de los propios estudiantes.

Por esto, el propósito del presente estudio fue crear y validar un instrumento que pudiese valorar de modo objetivo el sistema de evaluación presente en las guías docentes universitarias de la formación inicial del profesorado, bajo el prisma de la evaluación formativa y de las competencias docentes.

2. Método

2.1. Objetivos

El objetivo de este trabajo fue alcanzar un consenso de expertos sobre los elementos que convendría incluir en un Instrumento para analizar los sistemas de evaluación de las guías docentes universitarias, y valorar la capacidad de dichos elementos para evaluar las competencias docentes de la formación inicial del profesorado de una concepción de evaluación formativa, coherente con el diseño de programas basado en competencias; para finalmente, diseñar y validar el instrumento.

2.2. Diseño

Para la elaboración y validación del instrumento se utilizó un diseño no experimental basado en el método Delphi. Es una técnica que facilita el acuerdo entre expertos mediante retroalimentaciones anónimas efectuadas en distintas rondas y que se ha reconocido como adecuado para conseguir consensos en conocimientos en los que existe un cierto grado de novedad (Kenney, Hasson, & McKenna 2006; Skulmoski, Hartman & Krahn, 2007). Se realizó un método Delphi convencional (Yáñez y Cuadra, 2005), organizado en tres fases: preliminar, exploratoria y final (Astigarraga, s.f.; Bravo y Arrieta, 2005; Expósito et al., 2004, García-Bellido et al., 2012; Martínez, 2005; Onwuegbuzie et al., 2010; Vázquez et al., 2005). Además, se realizó una prueba piloto del instrumento.

2.3. Población y muestra

Se accedió a una muestra de 19 profesores de universidades españolas, mediante la técnica de muestreo no probabilístico intencional (Landeta, 2002, Tójar, 2006). La inexistencia de un listado completo y actualizado de profesores universitarios expertos en evaluación de guías docentes de universidad en España dificultó la realización de un muestreo probabilístico aleatorio (Caballero, 2014); por ello, se optó por un muestreo no probabilístico que a pesar de impedir calcular el error de la muestra permitió alcanzar estimaciones fiables (Rodríguez, 1990).

Dicha muestra se organizó en varios grupos humanos encargados de validar el instrumento en construcción (Bravo & Arrieta, 2005; Cruz, 2006; Luna, Infante & Martínez, 2005; Oñate, Ramos & Díaz, 1998). Concretamente se estableció un grupo coordinador de tres personas, entre los que se eligió un moderador (Varela, Díaz & García, 2012); otro grupo de 16 investigadores: 10 profesores universitarios de titulaciones de actividad física, miembros de la Red Nacional de Evaluación Formativa y Compartida; y 6 profesores, miembros destacados del Instituto de Ciencias de la Educación de la Universidad de Zaragoza. Diez de ellos fueron hombres y seis mujeres, con edades comprendidas entre los 42 y los 65 años. Todos ellos fueron elegidos atendiendo a su destacado perfil como investigadores, gestores y/o docentes innovadores en materia de educación superior y de todos se obtuvo un coeficiente de competencia experta medio y alto (el índice de competencia menor fue $K= 0,75$ y el mayor $K= 0,9$), siendo el valor de coeficiente de competencia experta alto el situado

entre los valores 0,8 y 1,0 y el índice de competencia medio el situado entre los valores 0,5 y 0,8 (Cabero & Barroso, 2013).

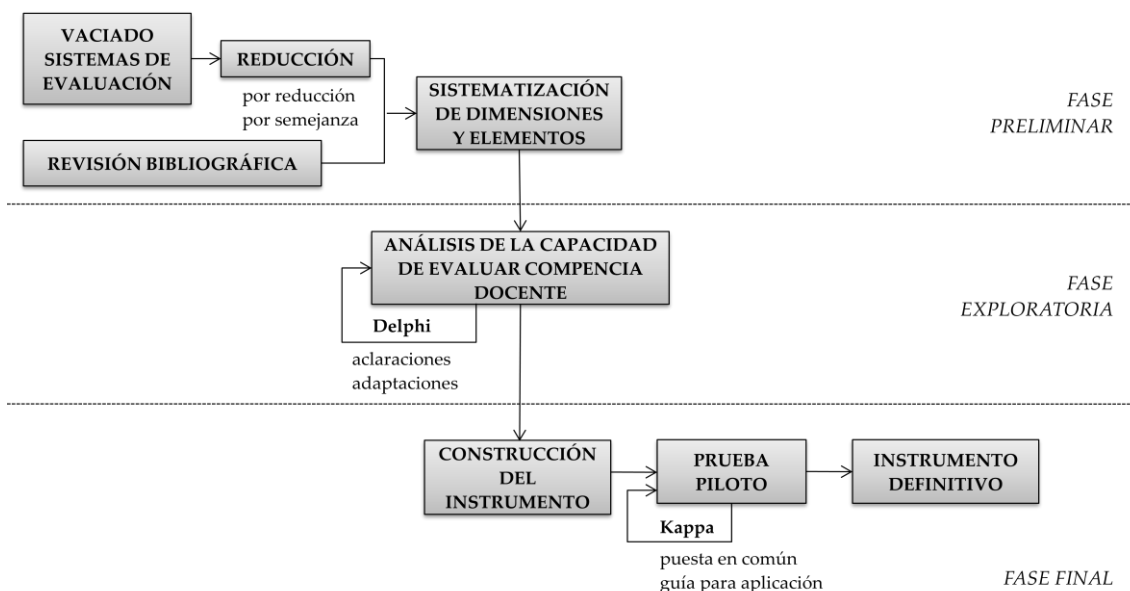
En lo que se refiere a la selección de las guías docentes, se tomaron 78 guías de 11 universidades españolas, obtenidas por accesibilidad a la muestra. El conjunto cumplía con el criterio de heterogeneidad en base a la titulación, materia y zona geográfica (Ibarra & Rodríguez, 2010).

Para la prueba piloto aplicada en la etapa final del estudio, se utilizó un muestreo estratificado estableciendo un reparto equitativo, y seleccionando, finalmente al azar, 9 guías docentes.

2.4. Procedimiento

El proceso de construcción del instrumento partió de la delimitación precisa del objeto de estudio (Vázquez et al., 2005); posteriormente se organizó el procedimiento en tres fases: (1) fase preliminar en la que se delimitaron los elementos y dimensiones del instrumento utilizando un consenso de expertos (Keeney et al., 2006); (2) fase exploratoria, en la que se preguntó a los expertos sobre la capacidad de cada elemento para identificar competencia docente, en dos rondas del método Delphi, y se analizó la consistencia entre sus opiniones y (3) fase final, en la que se dio forma al instrumento y protocolos de aplicación y se comprobó, mediante un estudio piloto, la aplicabilidad del instrumento diseñado (figura 1).

Figura 1. Esquema de la metodología empleada.



Fuente: Elaboración propia.

2.4.1. Fase preliminar: delimitación de dimensiones

El objetivo fue determinar las dimensiones y elementos constitutivos de los sistemas de evaluación. Se partió de dos fuentes de información: (1) el vaciado de los sistemas de evaluación hallados en las guías docentes; y (2) la revisión de bibliografía específica. Se revisaron 78 guías docentes hallando un total de 258 referencias a la evaluación, textualmente diferentes, aunque muchas de ellas con significados próximos. Se procedió a una simplificación de ese listado anidando aquellos elementos con significados semejantes, reduciéndose a un total de 109. En una siguiente fase se agruparon aquellos elementos diferenciados únicamente en aspectos puntuales o secundarios quedando reducida la lista a 97 referencias.

Paralelamente, la revisión bibliográfica (Castejón, Capllonch, González & López- Pastor, 2009; López-Pastor, 2012; Rodríguez, Ibarra & García, 2013; Hamodi, López- Pastor & López-Pastor, 2015) permitió una primera selección y sistematización de veintitrés elementos de evaluación en nueve dimensiones diferentes: (1) tareas; (2) objetos; (3) modalidad; (4) criterios; (5) retroalimentación; (6) medios; (7) técnicas; (8) participación; y (9) instrumentos.

Los miembros del equipo coordinador integraron las referencias procedentes de las guías y los elementos de evaluación extraídos de análisis de la bibliografía a las 9 dimensiones teóricas, mediante un proceso de análisis y registro de acuerdos existentes entre los tres componentes.

Con todo ello, se reordenó el conjunto componiendo un listado de 45 ítems agrupados en cuatro dimensiones: (1) modalidad de evaluación (final, continua o mixta); (2) agente evaluador (profesor, evaluador externo, autoevaluación, coevaluación, entre pares, por compañeros...); (3) retroalimentación; y (4) medios utilizados para la evaluación.

2.4.2. Fase exploratoria: determinación de la capacidad de los sistemas de evaluación para evaluar competencias

Concluida la delimitación de dimensiones y elementos constitutivos del instrumento, se procedió a averiguar su capacidad para evaluar competencia docente desde una perspectiva formativa.

Para ello, aplicamos la primera ronda de la técnica Delphi al grupo de los expertos seleccionado para esta fase, aislados entre sí, pidiéndoles que valoraran de 0 a 5 puntos cada uno de los 45 ítems en función de su capacidad para evaluar competencia docente desde un punto de vista formativo, siendo 0 el valor mínimo y 5 el máximo. Además, para cada ítem se ofreció la posibilidad de que el experto incluyera, de manera abierta, cuantas observaciones considerara oportunas.

Recibidas las respuestas de los expertos se calcularon para cada ítem: la media aritmética (\bar{x}), la desviación típica (DT) y el coeficiente de variación (cv) dado que este permite relacionar el tamaño de la muestra con la variabilidad de la variable; cuando el valor es 0 el acuerdo entre expertos es máximo y mínimo si es 1. Además, las

observaciones aportadas se transcribieron a un cuadro de doble entrada en el que se recogieron dichas observaciones para cada ítem y cada experto, de modo que todos los participantes pudieran conocer la lógica empleada por el resto de expertos, y siempre respetando el anonimato de todos ellos (Varela et al., 2012).

También se redactó un documento en el que se introdujeron y comentaron de forma explícita las adaptaciones realizadas en el cuestionario y se aportaron las explicaciones a las preguntas formuladas por los expertos.

Reformulado el cuestionario, se efectuó la segunda ronda de la técnica Delphi, volviendo a enviar el instrumento a los expertos (segunda vuelta), junto con el cuadro de sus respuestas y las medias resultantes, el cuadro de observaciones y las explicaciones aportadas por los autores, solicitando una nueva contestación, en forma análoga a la anterior, pero ahora conociendo las respuestas del resto de expertos y las adaptaciones surgidas; todo ello para garantizar que el proceso fuera iterativo, clave en la técnica Delphi (Varela et al., 2012).

Recibidas las respuestas de la segunda ronda se calcularon las medias aritméticas (\bar{x}), desviaciones típicas (DT) y coeficientes de variación de Pearson (cv); así como las diferencias entre ambas rondas. Los valores de los coeficientes de variación descendieron en 42 de los 45 ítems, pasando de un valor medio del 30% en la primera ronda a uno del 18% en la segunda; ascendieron muy ligeramente en otros dos (0.06 y 0.05, respectivamente), apareciendo un resultado de error en el otro ítem, dado que en ese caso, la media de la segunda vuelta era 0 (todos los expertos determinaron el valor 0 en su respuesta). Con estos resultados se consideró que el acuerdo entre los expertos mostraba un alto grado de compactibilidad y, por tanto, no se vio la necesidad de seguir con una tercera ronda.

Todo ello permitió la obtención de un *índice* para cada uno de los elementos, producto de la media por 100, para obtener valores enteros y conseguir una mejor visibilidad, quedando así entre 0 y 500. También se resolvió eliminar los ítems que, a juicio de los expertos, resultaban irrelevantes quedando finalmente 34 ítems.

2.4.3. Fase final: construcción definitiva del instrumento y prueba piloto

Concretados los ítems y dimensiones del instrumento y obtenida la valoración de cada uno de ellos, se procedió a darle formato de hoja de registro. El instrumento se transcribió a una hoja Excel del paquete Microsoft Office 2007.

A efectos prácticos, los 34 ítems organizados en las cuatro dimensiones, pasaron a ser 36 ya que, de cara a la transcripción de los resultados a una hoja de cálculo, resultaba más sencillo el desdoble de dos de ellos (trabajos individuales y trabajos colectivos). Esta información se muestra en las tres primeras columnas de la hoja de registro (figura 2). La cuarta columna se diseñó para que el evaluador marcara con un 1 cada entrada, es decir, cada vez que encontrara en la guía uno de los ítems. La columna quinta contenía el *índice* o valor de la media (Astigarraga, s.f.), resultante de las decisiones aportadas por los expertos a través de la técnica Delphi. Para conseguir que quien revisara la guía docente se limitara a colocar los datos en las

casillas correspondientes, en la sexta columna, el sistema realizaba los cálculos necesarios para cada una de las dimensiones. Se protegió la hoja Excel y se bloquearon todas las casillas excepto las que debería rellenar el usuario: (1) columna "registro", en la que el evaluador indicaba la presencia de ese elemento y (2) columna "observaciones" preparada para que el observador escribiera, en su caso, los comentarios que considerara oportunos.

Figura 2. Instrumento para analizar el sistema de evaluación de las guías docentes universitarias en la formación inicial de profesorado (IASEG), desde un punto de vista formativo.

			registro	índice	valor	observaciones	
Modalidad	evaluación continua			475			
	prueba global / final			200	0	Modalidad	
Agente (quién evalúa)	externo			189			
	profesor			456	0	Agente ¿quién evalúa?	
	autoevaluación			422			
	entre pares, por compañeros...			422			
	coevaluación			467			
Retroalimentación	explícita			489	0	Retroali- mentación	
	implícita			222			
Medios	examen o prueba	test		156			
		preguntas cortas		222			
		prácticas motrices		322			
		preguntas de desarrollo		378			
		tipo ensayo		422			
		aspectos procedimentales		478			
	sesiones	diseño			500	0	Medios: sesiones o partes de sesión
		impartición			489		
	trabajos	en grupo / colectivo	observación		478		
			evaluación / reflexión / análisis		489		
práctico				489	0	Medios: trabajos en grupo	
individual		teórico		333			
		autónomo		422			
		dirigido		356			
Medios: obras fuentes		práctico		489	0	Medios: trabajos individuales	
		teórico		333			
		autónomo		422			
		dirigido		356			
		Carpeta de aprendizaje / portafolio / dossier / mochila didáctica			500		
		Guaderno de campo			422		
		Debates			400		
		Entrevista			400		
		Estudio / análisis / resolución de caso(s)			478		
		Exposición / defensa de trabajos, memorias...			444		
Informes / análisis sobre actividades, contenidos, tareas, sesiones...			444				
Lecturas y análisis de contenido			411				
Producción / representación artística expresiva o de otra índole			389				

Fuente: Elaboración propia.

El último paso para la validación del instrumento consistió en aplicarlo en una muestra de nueve guías tomadas al azar mediante un muestreo estratificado.

Cada guía fue revisada por dos componentes del grupo de expertos. Tras las revisiones de las guías se procedió a conocer el grado de concordancia interobservadores, calculando el coeficiente *kappa de Cohen*, al tratarse de variables categóricas y utilizando para ello el programa *SPSS Statistics*, versión 22. El bajo nivel de acuerdo obtenido en una primera ronda (una media de 0.556, en una horquilla que iba de 0.175 a 0.916) aconsejaron una reconsideración del procedimiento aplicado, resolviendo que el mero enunciado de las categorías no era suficiente, puesto que con frecuencia se encontraban en las guías docentes expresiones que podrían ser interpretadas de formas diferentes.

A partir de una puesta en común entre los expertos participantes, exponiendo las situaciones encontradas y la discusión en torno a ellas, se tomaron los siguientes acuerdos: (1) En la dimensión "agente", señalar la categoría "profesor" en todos los casos; para el resto de agentes, marcar únicamente cuando hubiera mención explícita en la guía; (2) la constatación de que existía retroalimentación siempre que se mencionase de forma explícita, pero también en los casos en los que se utilizara evaluación entre iguales, evaluación compartida o se aludiese a la existencia de

tutorías, dado que se daba por seguro que en todos estos casos el estudiante recibía información sobre su labor; y (3) la asociación de expresiones "prueba objetiva" como examen o prueba tipo test siempre que apareciese de ese modo en la guía docente.

Consecuentemente, se redactó una minuciosa guía para el usuario que fuera a proceder a la revisión de guías docentes, asegurando que en todo momento estuviera perfectamente informado sobre cómo debía rellenar el instrumento, incluyendo las oportunas explicaciones sobre la naturaleza y sentido de las diferentes dimensiones y elementos presentes en la hoja de registro, así como las indicaciones sobre cómo debería interpretarse la información recogida dentro del apartado de evaluación de las guías docentes.

Un segundo ensayo, en el que los revisores pudieron utilizar la ayuda de esta guía, mostró unos resultados de concordancia muy elevados, medidos a través del coeficiente *kappa de Cohen*: .967 de media, en un abanico entre .883 y 1.000.

3. Resultados

A partir de las referencias existentes en los sistemas de evaluación contenidos en las 78 guías docentes consultadas, se consiguió una importante reducción, al identificar las coincidencias y analizar las semejanzas encontradas; por su parte, la revisión bibliográfica amplió las referencias y permitió una sistematización en dimensiones y elementos. La conjunción de ambos procedimientos permitió pasar de las 258 referencias iniciales, a 45 elementos, y posteriormente a 36 como se ha indicado, agrupados en cuatro dimensiones: (1) modalidad; (2) agente; (3) retroalimentación; y (4) medios utilizados para la evaluación. El volumen de información contenido en esta cuarta dimensión aconsejó distribuirla en cuatro apartados: (1) exámenes o pruebas; (2) prácticas docentes; (3) trabajos; y (4) otros.

Naturalmente esta nueva disposición permitió una revisión de guías docentes notablemente más ágil y precisa.

La averiguación de la capacidad de cada elemento de evaluación para detectar mayor o menor competencia docente, fue realizada por 16 expertos. Se utilizó la técnica Delphi a doble vuelta. Tras la segunda vuelta, las medias y el coeficiente de variación de Pearson (acuerdo entre expertos, compactibilidad) se muestran en la tabla 1:

Tabla 1
Índice de capacidad para captar competencia docente y coeficiente de variación de Pearson.

		\bar{x} (índice)	cv Pearson	DT
Modalidad	<i>evaluación continua</i>	4.75	0.10	0.46
	<i>prueba global</i>	2.00	0.46	0.93
	<i>ambas</i>	4.00	0.22	0.87
Agente (quién evalúa)	<i>externo</i>	1.89	0.41	0.78

		<i>profesor</i>	4.56	0.12	0.53
		<i>autoevaluación</i>	4.22	0.16	0.67
		<i>entre "pares", por compañeros...</i>	4.22	0.10	0.44
		<i>coevaluación</i>	4.67	0.11	0.50
Retroalimentación		<i>explícita</i>	4.89	0.07	0.33
		<i>implícita</i>	2.22	0.49	1.09
		<i>no hay</i>	0.00	0.00	0.00
Examen o prueba	<i>estructura formal</i>	<i>tipo test / preguntas cerradas</i>	1.56	0.65	1.01
		<i>preguntas cortas</i>	2.22	0.30	0.67
		<i>preguntas de desarrollo</i>	3.78	0.22	0.83
	<i>basado en la</i>	<i>tipo ensayo</i>	4.22	0.10	0.44
		<i>reproducción</i>	1.44	0.36	0.53
		<i>resolución (problemas, situaciones...)</i>	4.78	0.09	0.44
	<i>con</i>	<i>producción</i>	4.56	0.12	0.53
		<i>apoyo de las TIC</i>	3.89	0.15	0.60
		<i>documentos</i>	4.11	0.08	0.33
	<i>sobre, referido a...</i>	<i>aspectos conceptuales / teóricos</i>	3.33	0.26	0.87
		<i>aspectos procedimentales / prácticos</i>	4.78	0.09	0.44
		<i>prácticas motrices / pruebas físicas</i>	3.22	0.30	0.97
Trabajos	<i>componentes</i>	<i>en grupo / colectivo</i>	4.33	0.16	0.71
		<i>individual</i>	3.89	0.15	0.60
	<i>contenido</i>	<i>práctico</i>	4.89	0.07	0.33
		<i>teórico</i>	3.33	0.30	1.00
	<i>decisiones</i>	<i>autónomo</i>	4.22	0.16	0.67
		<i>dirigido</i>	3.56	0.29	1.01
	<i>elección</i>	<i>obligatorio</i>	3.33	0.30	1.00
<i>opcional / optativo</i>		3.33	0.21	0.71	
Sesiones o partes: ejercicios, tareas, situaciones...		<i>voluntario</i>	3.67	0.24	0.87
		<i>diseño</i>	5.00	0.00	0.00
		<i>impartición</i>	4.89	0.07	0.33
		<i>observación</i>	4.78	0.09	0.44
Medios		<i>evaluación / reflexión / análisis</i>	4.89	0.07	0.33
		<i>Carpeta de aprendizaje / portafolio / dossier / mochila didáctica</i>	5	0.00	0.00
		<i>Cuaderno de campo</i>	4.22	0.10	0.44
		<i>Debates</i>	4.00	0.13	0.50
		<i>Entrevista</i>	4.00	0.13	0.50

<i>Estudio / análisis / resolución de caso(s)</i>	4.78	0.09	0.44
<i>Exposición / defensa de trabajos, memorias...</i>	4.44	0.12	0.53
<i>Informes / análisis sobre actividades, contenidos, tareas, sesiones...</i>	4.44	0.16	0.73
<i>Lecturas y análisis de contenido</i>	4.11	0.08	0.33
<i>Producción / representación artístico expresiva o de otra índole</i>	3.89	0.09	0.33

Fuente: Elaboración propia.

Entre estos resultados destacamos dos: (1) el diseño de prácticas docentes y (2) la carpeta de aprendizaje, portafolio, mochila didáctica..., ambos con (\bar{x})=5.00; $cv=0.00$), por ser dos medios sobre los que los expertos se muestran totalmente de acuerdo en indicar su máxima capacidad para detectar competencia docente desde un plano formativo. Otros cuatro elementos: (1) presencia explícita de la retroalimentación; (2) trabajos con contenido práctico; (3) impartición de prácticas docentes; y (4) evaluación de esas prácticas, presentan un elevadísimo grado de acuerdo ($\bar{x}=4.89$; $cv=0.07$) entre los expertos para indicar, a su vez, su altísima sensibilidad para captar competencia docente; y finalmente cuatro elementos con ($\bar{x}=4.78$; $cv=0.09$): (1) exámenes basados en la resolución de problemas o situaciones; (2) exámenes sobre aspectos procedimentales; (3) observación de prácticas docentes; y (4) estudio de casos, se encuentran en una situación muy próxima, tanto por su compactibilidad como por su sensibilidad didáctica, a los anteriores.

Para dar una visión global de los resultados, se cruzaron los datos de ambas dimensiones, es decir del índice de capacidad para captar competencia docente y del coeficiente de variación. La tabla 2 muestra que 36 de los 45 elementos presentan una compactibilidad alta o superior y 38 resultan, en opinión de los expertos, con alta -o superior- capacidad para captar competencia docente. Su combinación marca que casi cuatro de cada cinco elementos (77.77%) están en alguno de los tres tramos superiores en ambas variables. El resto se encuentra disperso en categorías inferiores; de entre ellos merece la pena comentar el caso de la “ausencia de retroinformación” sobre la que los expertos se muestran totalmente de acuerdo en valorar como nula su capacidad para identificar competencia docente.

Tabla 2
Visión global de los resultados (en nº de casos por opción).

	Compactibilidad (cv)						total
	Máxima (0,01- 0,09)	Muy alta (0,1- 0,19)	Alta (0,2- 0,29)	Moderada (0,3-0,39)	Aceptable (0,4-0,5)	Baja (más de 0,5)	
Máxima (5)	2						2
Muy alta (4,9-4)	10	14	1				25
Alta (3,9- 3)	1	2	5	3			11
							<u>38 (84%)</u>

Capacidad para captar competencia docente	Moderada (2,9-2)				1	2	3		
	Aceptable (1,9-1)				1	1	1	3	
	Nula (0)	1						1	
Total		14	16	6	5	3	1	45	
		36 (80%)						35 (77.7%)	

Fuente: Elaboración propia.

La aplicabilidad del instrumento se comprobó mediante su implementación en una prueba piloto consistente en la revisión por pares de nueve guías docentes universitarias y el consiguiente análisis del grado de acuerdo calculando el coeficiente kappa de Cohen para cada caso. Todos los valores muestran un índice de concordancia excelente (Landis & Koch, 1977) siendo máximo en cinco de los nueve casos (tabla 3).

Tabla 3
Índice de concordancia.

Asignatura	Kappa	Asignatura	Kappa	Asignatura	kappa
A01	1.000	B01	1.000	C01	.929
A02	.883	B02	.941	C02	1.000
A03	.939	B03	1.000	C03	1.000

Fuente: Elaboración propia.

Finalmente la obtención de un índice para cada uno de los elementos permite la valoración de cada dimensión en su conjunto e igualmente un índice final global del conjunto del sistema de evaluación de la asignatura, lo que permite reflexionar sobre el significado del análisis de cara a comparar entre distintos sistemas de evaluación, analizar coherencias con otros elementos programáticos, etc.

4. Discusión

Se ha desarrollado un instrumento mediante un procedimiento en el que se han aplicado técnicas y cálculos para asegurar la validez y confiabilidad del mismo (Onwuegbuzie et al., 2010; Vázquez et al., 2005); todo ello con el objetivo de poder analizar la adecuación de los sistemas de evaluación de las guías docentes universitarias para evaluar competencias docentes desde un prisma formativo.

En su diseño se han conseguido amplios acuerdos entre los grupos de expertos implicados en las distintas fases. Ello se ha podido refrendar a partir de los resultados de las pruebas efectuadas con el método Delphi. Finalmente en el estudio piloto los valores del índice de concordancia entre evaluadores pueden calificarse de excelentes cumpliendo así los presupuestos de Landis y Koch (1977), lo que nos indica su aplicabilidad, en cuanto a su comprensión.

Además se han elaborado documentos que facilitan y complementan la utilización del instrumento, como la guía de acompañamiento que por motivos de espacio no se adjuntan en este trabajo. También cabe destacar que el instrumento presenta un formato de fácil aplicación dado que los cálculos se efectúan de manera automática por lo que se facilita la toma de decisiones del evaluador.

Ambas cuestiones permiten decir que el instrumento es útil para su objetivo y que el procedimiento ha seguido estándares sobre la necesidad de controlar la confiabilidad, fiabilidad avalados por diversos autores (Casanova, 1995; Inciarte & Canquiz, 2001) y dispone de una considerable aplicación práctica. Una vez obtenidos los valores de una determinada guía para cada una de las dimensiones

Por otro lado hemos construido un elemento *ad hoc*, en la línea que comentaban Expósito et al. (2004) o Rodríguez-Gómez et al. (2012) como característico de este tipo de estudios.

No obstante, consideramos que las guías docentes son un instrumento sintético y comprensible y podrían no contener toda la información programada por el docente (García & Morillas, 2011), por lo que su análisis podría proporcionar una visión incompleta del sistema de evaluación. Este problema se solucionaría disponiendo del proyecto docente del profesor o equipo de profesores, si existiese (Yániz, 2006), al ser un documento más descriptivo. A pesar de ello, al igual que Stlington y Coetzer (2015), hemos comprobado la utilidad de la técnica Delphi como una buena práctica para el estudio y posible renovación del diseño curricular.

En resumen, el instrumento que presentamos tiene cualidades por las que concluimos que es una excelente aportación para la investigación educativa en la universidad:

(1) El tópico sobre la evaluación formativa y por competencias en la universidad está siendo desarrollado en los últimos tiempos de manera notable, pero escasean los instrumentos consensuados que posibiliten el registro mediante evidencia de lo que realmente está aconteciendo en los sistemas de evaluación universitaria.

(2) El instrumento es novedoso en cuanto que permite evaluar el valor formativo de los sistemas de evaluación de las guías docentes, teniendo en cuenta que a partir de la implantación del ECTS la programación se diseña a partir de competencias; y que los expertos coinciden en que la evaluación formativa es el sistema de evaluación más alineado con la programación por competencias.

(3) Por otro lado, se ha llegado por consenso a destacar y definir los elementos y dimensiones de los sistemas de evaluación de las guías docentes que los expertos han considerado pertinentes, es decir, puede haber otros pero no se les ha atribuido suficiente importancia desde un plano formativo.

(4) Además, una gran aportación consiste en mostrar una ponderación entre las dimensiones y también entre los elementos, es decir, una ordenación de mayor o menor valor desde un prisma formativo.

(5) Este instrumento además, debido al exhaustivo proceso llevado a cabo para su construcción, permite corroborar, a tenor de la opinión de los expertos, afirmaciones quizá extendidas pero no suficientemente comprobadas sobre la pertinencia de determinados procesos evaluativos y su función formativa. Asimismo, este registro permitirá evidenciar tanto lo que acontece de forma global como por factores o dimensiones, radiografiando los sistemas de evaluación con la detección de los puntos fuertes y débiles de los mismos en relación a su carácter competencial y formativo.

Finalmente, concluimos que este instrumento es una aportación de gran utilidad para la investigación educativa en el ámbito universitario en particular, dado que permite llevar a cabo análisis tanto de aspectos programáticos de las asignaturas, como de las titulaciones en su conjunto, contribuyendo a su análisis riguroso y, en consecuencia, facilitando los procesos de mejora de la calidad a nivel didáctico e institucional.

Referencias bibliográficas

- Astigarraga, E. (s.f.) *El método Delphi. Material del Curso de Prospectiva y Estrategia Empresarial*. Universidad de Deusto. Facultad de Ciencias Económicas y Empresariales. Recuperado de http://www.prospectiva.eu/zaharra/Metodo_delphi.pdf
- Aramburuzara, P., Hernández-Castilla, R. y Ángel-Urbe, I. C. (2013). Modelos y tendencias de la formación docente universitaria. *Profesorado. Revista de Currículum y Formación del Profesorado*, 17(3), 345-357. Recuperado de <http://www.ugr.es/local/recfpro/rev173COL9.pdf>
- Biggs, J. (2005). *Calidad del aprendizaje universitario*. Narcea.
- Bolívar, A. (2008). *Ciudadanía y competencias básicas*. Fundación ECOEM.
- Bravo, M. y Arrieta, J. (2005). El Método Delphi. Su implementación en una estrategia didáctica para la enseñanza de las demostraciones geométricas. *Revista Iberoamericana de Educación*, 35(3). Recuperado de http://www.rieoei.org/inv_edu38.htm
- Buscà, F., Pintor, P., Martínez-Mínguez, L. y Peire, T. (2010). Sistemas y procedimientos de Evaluación Formativa en docencia universitaria: Resultados de 34 casos aplicados durante el curso académico 2007-2008. *Estudios sobre Educación*, 18, 255-276.
- Caballero Martínez, F. (2014). Modificación de un método de consenso experto para reducir la variabilidad de la práctica clínica. Tesis doctoral. Recuperado de: https://repositorio.uam.es/bitstream/handle/10486/662738/caballero_martinez_fernando.pdf?sequence=1

- Cabero, J. & Barroso, J. (2013). La utilización del juicio de experto para la evaluación de TIC: el coeficiente de competencia experta. *Bordón Revista de Pedagogía*, vol. 65(2), 25-38.
- Casanova, M. A. (1995). *Manual de evaluación educativa*. La Muralla.
- Castejón, J., Capllonch, M., González, N. y López-Pastor, V. M. (2009). Técnicas e instrumentos de evaluación formativa y compartida para la docencia universitaria. En V. M. López-Pastor (Coord.), *Evaluación formativa y compartida en educación superior* (pp. 65-91). Narcea.
- Chivite, M, Asún S. & Romero, M. R. (Coords.) (2011). *Actas del “VI Congreso Internacional de Evaluación Formativa y Compartida en Docencia Universitaria”*. Universidad de Zaragoza.
- Cruz, M. (2006). *El Método Delphi en las Investigaciones Educativas*. Material del Proyecto Ramal Estadística para la Educación, Cuba. Recuperado de [file:///D:/Desktop/Miguel%20Cruz%20\(Libro\)%20Delphi.pdf](file:///D:/Desktop/Miguel%20Cruz%20(Libro)%20Delphi.pdf)
- Expósito, J., Olmedo, E. & Fernández, A. (2004). Patrones metodológicos en la investigación española sobre evaluación de programas educativos. *RELIEVE*, 10(2), 185-209. Recuperado de http://www.uv.es/RELIEVE/v10n2/RELIEVEv10n2_2.htm
- García, M. P. & Morillas, L. R. (2011). La planificación de evaluación de competencias en Educación Superior. *REIFOP*, 14(1), 113-124. Recuperado de <http://www.aufop.com>
- García-Bellido, R., Jornet, J. M. & González-Such, J. (2012). Una aproximación conceptual para el diseño de instrumentos de evaluación de la competencia aprender a aprender en los profesionales de la educación. *Revista Iberoamericana de Evaluación Educativa*, 5(1). Recuperado de http://www.rinace.net/riee/numeros/vol5-num1_e/art14.pdf
- Gibbs, G. (2003). Uso estratégico de la evaluación en el aprendizaje. En S. Brown & A. Glasner, *Evaluar en la universidad. Problemas y nuevos enfoques* (pp. 61-70). Narcea.
- Gijbels, D., Wattering, G. van de & Dochy, F. (2005). Integrating assessment tasks in a problem-based learning environment. *Assessment and evaluation in higher education*, 30, 73-86.
- Goby, V. P. & Nickerson, C. (2014). Accreditation and assessment of learning in the UAE. *Quality Assurance in Education*, 22(3), 212-225.
- González-Fernández, N. (2007) (Coord.). *Desarrollo y evaluación de competencias a través del portafolio del estudiante*. Universidad de Cantabria.

- Grau, S. & Gómez, C. (2010). La evaluación, un proceso de cambio para el aprendizaje. En S. Grau & C. Gómez (Coords.), *Evaluación de los aprendizajes en el Espacio Europeo de Educación Superior* (pp. 17-32). Marfil.
- Hamodi, C., López-Pastor, V. L; López-Pastor, A. T. (2015). Medios, técnicas e instrumentos de evaluación formativa y compartida en Educación Superior. *Perfiles Educativos*, 147(37), 146-161. Recuperado de http://www.iisue.unam.mx/perfiles/perfiles_articulo.php?clave=2015-147-146-161
- Hall, K. & Burke, W. (2003). *Making formative assessment work - Effective practice in the primary classroom*. Open University Press.
- Hebles, M., Dos Santos, M. A., Yaniz, C. y Villardon-Gállego, L. (2017). Diseño y validación de la escala de evaluación de los aprendizajes (EEA). *Profesorado. Revista de Curriculum y Formación del Profesorado*, nº extraordinario, 107-126.
- Ibarra, M. S. & Rodríguez, G. (2010). Aproximación al discurso dominante sobre la evaluación del aprendizaje en la universidad. *Revista de Educación*, 351, 385-407.
- Inciarte, A. & Cánquiz, L. (2001). *Informe de investigaciones educativas*, 15(1-2), 1-16.
- Kaftan, J.; Buck, G. & Haack, A. (2006). Using Formative Assessments to Individualize Instruction and Promote Learning. *Middle School Journal*, 37(4), 44-49.
- Keeney, S., Hasson, F. & McKenna, H. P. (2006). Views on the ground: teaching and clinical assessors' views on vocational training for health care assistants in Ireland. *Journal of Clinical Nursing*, 53(2) 205-212.
- Landeta, J. (2002). *El método Delphi: una técnica de previsión del futuro*. Barcelona: Ariel.
- Landis J. R. & Koch G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33, 159-174.
- López-Pastor, V. M. (Coord). (2009). *Evaluación formativa y compartida*. Narcea.
- López-Pastor, V. M. (2012). Evaluación formativa y compartida en la universidad: clarificación de conceptos y propuestas de intervención desde la Red Interuniversitaria de Evaluación Formativa. *Psychology, Society & Education*, 4(1), 117-130. Recuperado de <http://www.psyse.org/articulos.php?id=86>
- López, M. C.; Pérez-García, M. P. y Rodríguez, M. J. (2015). Concepciones del profesorado universitario sobre la formación en el marco del espacio europeo de educación superior. *Revista de Investigación Educativa*, 33(1), 179-194. doi: <http://dx.doi.org/10.6018/rie.33.1.189811>

- Lorente, E. & Kirk, D. (2013). Alternative democratic assessment in PETE: an action-research study exploring risks, challenges and solutions. *Sport, Education and Society*, 18(1), 77-96. Recuperado de <http://dx.doi.org/10.1080/13573322.2012.713859>
- Lorente, E., Montilla, M. & Romero, M. (2013). Grado de definición y coherencia de los programas de expresión corporal en las titulaciones universitarias de educación física. *Revista de evaluación educativa*, 2(1). Recuperado de <http://revalue.mx/revista/index.php/revalue/article/view/47/113>
- Luna, P., Infante, A. Y Martínez, F. J. (2005). Los Delphi como fundamento metodológico predictivo para la investigación en Sistemas de Información y Tecnologías de la Información (IS/IT). *Pixel-Bit. Revista de Medios y Educación*, 26, 89-112.
- Martínez, M. (2005). Instrumentos de diagnóstico. *GestioPolis.com [en línea]*. Recuperado de <http://www.gestiopolis.com/recursos4/docs/ger/instrudiag.htm>
- Mateo, J. (2000). *La evaluación educativa, su práctica y otras metáforas*. Barcelona: ICE-Horsori. Recuperado de <https://web.ua.es/es/ice/documentos/recursos/materiales/la-evaluacion-educativa.pdf>
- Medina, A.; Domínguez, M. C. y Sánchez, C. (2013). Evaluación de las competencias de los estudiantes: modelos y técnicas para la valoración. *Revista de Investigación Educativa*, 31(1), 239-255. doi: <http://dx.doi.org/10.6018/rie.31.1.157601>
- Medland, E. (2016). Assessment in higher education: drivers, barriers and directions for change in the UK. *Assessment & Evaluation in Higher Education*, 41(1), 81-96.
- Muñoz, J. M., Ríos de Deus, M. P. & Abalde, E. (2002). Evaluación Docente vs. Evaluación de la Calidad. *Revista Electrónica de Investigación y Evaluación Educativa*, 8(2). Recuperado de http://www.uv.es/RELIEVE/v8n2/RELIEVEv8n2_4.htm
- Muñoz-San Roque, I., Martín-Alonso, J. F., Prieto-Navarro, L. & Urosa-Sanz, B. (2016). Autopercepción del nivel de desarrollo de la competencia de aprender a aprender en el contexto universitario: propuesta de un instrumento de evaluación. *Revista de Investigación Educativa*, 34(2), 369-383. doi: <http://dx.doi.org/10.6018/rie.34.2.235881>
- Nunziatti, G. (1990). Pour construire un dispositif d'évaluation formatrice. (Dossier du formateur). *Cahiers pédagogiques*, 280, 47-64.
- Onwuegbuzie, A. J., Bustamante, R. M. & Nelson, J. A. (2010). Mixed Research as a tool for developing quantitative instruments. *Journal Mixed Methods Research*, 4(1), 56-78. doi: <http://dx.doi.org/10.1177/1558689809355805>

- Oñate, N., Ramos, L. & Díaz, A. (1988). Utilización del Método Delphi en la pronosticación: Una experiencia inicial. *Economía Planificada*, 3(4), 9-48.
- Orden, A. de la, Asensio, I., Carballo, R., Fernández-Díaz, J., Fuentes, A., García-Ramos, J. M. & Guardia, S. (1997). Desarrollo y validación de un modelo de calidad universitaria como base para su evaluación. *RELIEVE*, 3(1). Recuperado de http://www.uv.es/RELIEVE/v3n1/RELIEVEv3n1_2.htm
- Partington, P. & Brown, G. (2013). Quality assessment, staff development and cultural change. *Quality Assurance in Education*, 5(4), 208-217.
- Pérez, R. (2016). ¿Quo vadis, evaluación? Reflexiones pedagógicas en torno a un tema tan manido como relevante. *Revista de Investigación Educativa*, 34(1), 13-30. doi: <http://dx.doi.org/10.6018/rie.34.1.239381>
- Rodríguez, G., Ibarra, M. S. & García, E. (2013). Autoevaluación, evaluación entre iguales y coevaluación: conceptualización y práctica en las universidades españolas. *Revista de Investigación en Educación*, 11(2), 198-210. Recuperado de <http://webs.uvigo.es/reined/>
- Rodríguez-Gómez, G. Ibarra-Sáiz, M. S.; Gallego-Noche, B., Gómez-Ruiz, M. A. & Quesada-Serra, V. (2012). La voz del estudiante en la evaluación del aprendizaje: un camino por recorrer en la universidad. *RELIEVE*, 18(2), 2. doi: <http://dx.doi.org/10.7203/relieve.18.2.1985>.
- Rodríguez, J. (1990). La muestra: teoría y aplicación. En García Ferrando, Ibáñez y Alvira (coords). *El análisis de la realidad social. Métodos y técnicas de investigación* (287-320). Alianza Editorial.
- Rojó, V. Á., García, M. G., Muñoz, I. A., López, J. C., Arbizu, R. D. F., Noche, B. G.,... González, S. G. (2009). Evaluación del diseño de las "guías FORCOM de autoevaluación de competencias docentes" para el espacio europeo de educación superior (EEES). *Revista de Docencia Universitaria*, 4.
- Romero, M. R., Fraile, A., López-Pastor, V. M. & Castejón, J. (2014). Relación entre sistemas de evaluación formativa, rendimiento académico y carga de trabajo del profesor y del alumno en la docencia universitaria. *Infancia y aprendizaje*, 37(2), 310-341. Recuperado de <http://dx.doi.org/10.1080/02103702.2014.918818>doi:10.1080/02103702.2014.918818
- Sánchez-Santamaría, J. (2011). Evaluación de los aprendizajes universitarios: una comparación sobre sus posibilidades y limitaciones en el Espacio Europeo de Educación Superior. *Revista de Formación e Innovación Educativa Universitaria*, 4(1), 40-54. Recuperado de http://refiedu.webs.uvigo.es/Refiedu/Vol4_1/REFIEDU_4_1_4.pdf
- Santos, M. L., Martínez, F. & López-Pastor, V. M. (Coords.) (2009). *La innovación docente en el EEES*. Universidad de Almería.

- Shavelson, R. J. (2010). Assessing Student Learning Responsibly: From History to an Audacious Proposal. *Change: The Magazine of Higher Learning*, 39(1), 26-33.
- Segers, M. & Dochy F. (2001). New assessment forms in problem-based learning: the value-added of the students' perspective. *Studies in higher education*, 26, 327-339.
- Sitlington, H. B. & Coetzer, A. J. (2015). Using the Delphi technique to support curriculum development. *Education & Training*, 57(3), 306-321.
- Skulmoski, G. J.; Hartman, F. T.; Krahn, J. (2007). The Delphi method for graduate research. *Journal of Information Technology Education*, 6, 1-21.
- Tójar, J. C. (2006). *Investigación cualitativa: comprender y actuar*. La Muralla.
- Torrance, H. (2012). Formative assessment at the crossroads: conformance, deformative and transformative assessment. *Oxford Review of Education*, 38(3), 323-342. doi: <http://dx.doi.org/10.1080/03054985.2012.689693>.
- Tortosa, M. T. & Álvarez, J. D. (Coords.). (2009). *Investigaciones colaborativas en el ámbito universitario: Propuestas para el cambio*. Universidad de Alicante.
- Trevitt, C.; Breman, E. & Stocks, C. (2012). Assessment and learning: is it time to rethink student activities and academic roles? *Revista de Investigación Educativa*, 30 (2), 253-267. doi: <http://dx.doi.org/10.6018/rie.30.2.153441>
- Trillo, F. & Porto, M. (1999). La percepción de los estudiantes sobre su evaluación en la universidad. Un estudio en la Facultad de Ciencias de Educación. *Revista de Innovación Educativa*, 9, 55-75.
- Varela, M., Díaz, García, L. R. (2012). *Metodología de investigación en educación médica. Descripción y usos del método Delphi en investigaciones del área de la salud*. Departamento de Investigación en Educación Médica. Secretaría de Educación Médica. Facultad de Medicina, UNAM. México D.F., México.
- Vázquez, F. D., Fera, A., Moriel L. & Tijerina, L (2005). La validez de la evaluación de los programas de educación superior. Requisito para su internacionalización. *Revista Iberoamericana de Educación (Sección de los lectores)*, 35(6). Recuperado de <http://rieoei.org/evaluacion15.htm>
- Villa, A., Campo, L., Arranz, S., Villa, O. y García, A. (2013). Valoración del profesorado de magisterio sobre el aprendizaje basado en competencias implantado. *Profesorado. Revista de currículum y formación del profesorado*, vol 17(3), 36-55. Recuperado de <http://www.ugr.es/local/recfpro/rev173ART5.pdf>
- Yániz, C. (2006). Planificar la enseñanza universitaria para el desarrollo de competencias. *Educatio Siglo XXI*, 24, 17-34.

Yáñez, R. & Cuadra, R. (2008). La técnica Delphi y la investigación en los servicios de salud. *Ciencia y enfermería*, XIV(1), 9-15. Recuperado de http://www.scielo.cl/scielo.php?Script=sci_arttext&pid=S0717-95532008000100002

Zabalza, M. Á. (2001). Evaluación de los aprendizajes en la universidad. En A. García-Valcárcel (Ed.), *Didáctica Universitaria* (pp. 261-291). La Muralla.

Zabalza, M. Á. (2008). Innovación en la Enseñanza Universitaria: el proceso de convergencia hacia un Espacio Europeo de Educación Superior. *Educação*, 31(3), 199-209.

Cómo citar este artículo:

Romero-Martín, M.R., Asún Dieste, S. & Chivite Izco, M. (2020). Diseño y validación de un instrumento para analizar el sistema de evaluación de las guías docentes universitarias en la formación inicial del profesorado (IASEG). *Profesorado. Revista de Currículum y Formación de Profesorado*, 24(2), 346-367. DOI: 10.30827/profesorado.v24i2.15040