

BORDÓN

Revista de Pedagogía



Volumen 73
Número, 2
2021

SOCIEDAD ESPAÑOLA DE PEDAGOGÍA

DISEÑO Y VALIDACIÓN DE UN INSTRUMENTO DE EVALUACIÓN DE TFG

The design and validation of an instrument for assessing undergraduate dissertations

CONSUELO GARCÍA⁽¹⁾, PAOLA PEROCHENA⁽²⁾ Y LARA ORCOS⁽¹⁾

⁽¹⁾ *Universidad Internacional de La Rioja (España)*

⁽²⁾ *Universidad Nacional de Educación a Distancia (UNED) (España)*

DOI: 10.13042/Bordon.2021.89015

Fecha de recepción: 20/04/2021 • Fecha de aceptación: 03/06/2021

Autora de contacto / Corresponding author: Consuelo García. E-mail: consuelo.garcia@unir.net

INTRODUCCIÓN. El trabajo final de grado (TFG) supone la culminación de los estudios universitarios en los que el alumno demuestra el dominio y la integración de los contenidos y las competencias desarrolladas. Por ello, su evaluación requiere de instrumentos completos y complejos. El objetivo de esta investigación es diseñar y validar una herramienta de evaluación de los TFG de los Grados de Maestro en Educación, a fin de soslayar las deficiencias y problemas observados en las rúbricas al uso. **MÉTODO.** Se formó un grupo de siete expertos, tanto del ámbito escolar como universitario, para la elaboración de este instrumento. Se analizaron diferentes herramientas de evaluación y trabajos para su elaboración. Posteriormente, se estudió la fiabilidad del instrumento resultante aplicando el coeficiente de correlación intraclase (ICC) a veinte trabajos por parte de seis expertos independientes. **RESULTADOS.** El resultado es un instrumento que combina las mejores propiedades de la rúbrica y de la lista de verificación, con un desglose concreto y preciso de todos los indicadores que configuran la parte escrita de un trabajo final de grado, y que valora sus elementos considerados centrales: los contenidos y su calidad, los aspectos de formato y de expresión. El análisis de correlación intraclase ofreció unos resultados excelentes en las escalas medidas, siendo todas $>.90$, lo que muestra su alta consistencia. **DISCUSIÓN.** Estos resultados permiten confiar en la fiabilidad de la herramienta. El instrumento puede ser un claro apoyo para directores de TFG y para miembros de tribunales de evaluación, por su claridad, objetividad y facilidad de uso. E, igualmente, puede ser una herramienta útil para el aprendizaje de los estudiantes y favorecer su autorregulación y autoevaluación, dado que comprende con detalle todos los indicadores por los que serán evaluados.

Palabras clave: *Rúbrica, Trabajo fin de carrera, Grado, Educación infantil, Educación primaria.*

Introducción

Tras la implantación del Espacio Europeo de Educación Superior (EEES), los nuevos planes de estudio de las titulaciones universitarias se centraron especialmente en el futuro profesional del egresado. Igualmente, este contexto generó una mayor atención a la transparencia de la evaluación, a fin de concienciar e informar a los estudiantes de forma clara sobre los objetivos y criterios de sus evaluaciones (Jonsson, 2014). En España, el Real Decreto 1393/2007, del 29 de octubre, por el que se establece la ordenación de las enseñanzas universitarias de carácter oficial, incorporó las competencias como elementos curriculares clave, vinculadas con el ámbito profesional. Como comentan Ortega *et al.* (2018): “las competencias profesionales que articulan los nuevos planes de estudio pretenden un acceso al conocimiento desde un enfoque socioconstructivista, concediendo especial relevancia al aprendizaje en y sobre la acción en contextos sociales y profesionales (integración teoría y acción práctica), más allá de los espacios y tiempos universitarios” (p. 36). En este sentido, se plantean los trabajos finales de estudios (TFE), según los cuales, un alumno debe “demostrar un dominio integrado de las distintas competencias desarrolladas a lo largo de sus estudios, y/o completar el desarrollo de las mismas” (Valderrama *et al.*, 2010, p. 107). De tal manera, que el trabajo fin de grado (TFG) se realiza al final del plan de estudios y se orienta a la evaluación del desarrollo global de la adquisición de competencias, tanto transversales como específicas, por parte del estudiante.

El contexto académico propiciado por el Plan Bolonia de adaptación y unificación de criterios educativos en todas las instituciones europeas universitarias ha promovido una mayor atención a la transparencia en la evaluación de los estudiantes, a fin de que sean conscientes de los objetivos y criterios de evaluación (Jonsson, 2014). Es justamente en estos aspectos, de adecuación y transparencia de la evaluación en los que se centra el proyecto de desarrollo de una herramienta de evaluación objetiva y fiable del

TFG en los grados de Maestro en Educación Infantil y Primaria.

Tipos de evaluación de los trabajos finales de grado

El proyecto *Tuning Educational Structures in Europe* (2009) fue el punto de partida en España para armonizar las pautas para la evaluación de competencias en el marco del Espacio Europeo de Educación Superior. Así, se considera que hay ciertas competencias de carácter transversal que han de tener los titulados universitarios tras la finalización de sus estudios y que son: la comunicación oral y escrita en la lengua nativa, la capacidad de buscar y gestionar información, así como la habilidad de organizar, planificar, analizar, sintetizar y resolver problemas. En los primeros años de funcionamiento de los TFG, los directores y los tribunales de evaluación aplicaban un criterio evaluativo más o menos libre a los trabajos presentados. Esta situación suscitó una serie de problemáticas debido a la pluralidad de criterios entre los responsables de la evaluación y la calificación, que se intentó corregir a través de diferentes instrumentos de evaluación (Pathirage *et al.*, 2004).

Para la evaluación de los trabajos de los estudiantes en muchos contextos se diseñan instrumentos que en inglés se denominan *scoring key* o clave de puntuación. Dos de los más comunes son las rúbricas y las listas de verificación (Cañadas, 2020). Ambas pueden ser usadas como instrumentos de autoevaluación (Bharuthram y Patel, 2017), y a la vez como instrumentos para la heteroevaluación, cumpliendo con la función certificadora de la evaluación. Aunque habitualmente se ha considerado a la rúbrica como una herramienta para evaluar el logro de estándares o describir la calidad de un trabajo, existen muchas otras de formas de evaluar que potencian la transparencia (Bearman y Ajjawi, 2018), entendida esta como el conocimiento del alumno sobre el propósito y los criterios de evaluación (Jonsson, 2014).

Una rúbrica es un conjunto de elementos de evaluación coherentes respecto al trabajo de los estudiantes, que incluye descripciones para cada nivel de desempeño asociados a cada criterio (Brookhart, 2013). Así, la rúbrica se considera una herramienta valiosa para los estudiantes porque pone el énfasis en la comprensión en lugar de en la memorización, en el aprendizaje “profundo” en vez del aprendizaje “superficial” (Pausch y Popp, 1997), y las personas generalmente aprenden cuando se les dice qué es lo que se supone que deben aprender. Con el fin de evaluar el logro de las competencias en el desarrollo del TFG, surgieron algunas rúbricas de evaluación (Valderrama *et al.*, 2010; Reddy y Andrade, 2010).

Los beneficios de la evaluación con la rúbrica son relevantes tanto para los estudiantes como para los profesores (Bohórquez *et al.*, 2019). Comenta Oakleaf (2009) que a los estudiantes la rúbrica les permite comprender las expectativas de sus profesores, estar mejor capacitados para cumplir con lo que se les solicita, recibir comentarios directos sobre su aprendizaje actual y lo que necesitan en el futuro. Para los profesores que participan en la evaluación, la rúbrica permite acordar la manera de otorgar una calificación bastante fiable al estudiante, obtener información cualitativa de los resultados y un enfoque en la educación basada en estándares (Oakleaf, 2009). Además, la rúbrica es una herramienta que apoya la comprensión mutua de las tareas entre docentes y estudiantes, haciendo de soporte para el aprendizaje (Mauri, *et al.*, 2014). El estudio de Pegalajar (2021) recoge una valoración positiva por de los estudiantes del Grado de Maestro en Educación Infantil de la Universidad de Jaén sobre el uso de una rúbrica. Por su parte, el estudio de Zornoza-Gallego y Vercher (2019) aborda el diseño y aplicación de una rúbrica para la valoración de competencias generales en la elaboración de un TFG.

Existen varios tipos de rúbricas. En las denominadas analíticas, cada criterio se evalúa por

separado, mientras que en las holísticas los criterios se evalúan de manera simultánea (Dawson, 2017). El problema que presentan las primeras es que no representan interacciones complejas (Tomas *et al.*, 2019), y son poco operativas cuando el número de características a evaluar es grande. Y en el caso de las holísticas, el problema es que su validez de contenido es baja (Tomas *et al.*, 2019) y no permite identificar con detalle qué aspectos han motivado una puntuación u otra, como veremos en nuestro análisis posterior. En su estudio de validación de la homogeneidad de un modelo de rúbrica de evaluación del TFG en el grado de Enfermería, Sánchez (2021) encontró una homogeneidad baja tras aplicar el coeficiente de correlación intraclass (ICC). A pesar de la incertidumbre asociada a la definición de la rúbrica y sus usos, esta ha ido ganando popularidad a lo largo de los años como una herramienta de evaluación importante en relación con el aprendizaje, al brindar mayor transparencia en las evaluaciones (Bharuthram y Patel, 2017).

Las listas de verificación (*checklist*) se definen como instrumentos que comprueban si algunos comportamientos o características se presentan y si se dan las evidencias de los aprendizajes (sean estos trabajos escritos u observación de conductas o demostraciones). En el ámbito universitario estas herramientas se han aplicado para evaluar planes y programas de estudios (Merrill, 2020), materiales didácticos para determinar si cumplen o no con ciertas características (Baharuddin *et al.*, 2013), en cursos *online* (Shin y Cheon, 2019) y para la evaluación de estudiantes bilingües emergentes (SAEBL *check list*) (McClearyCale y Furtak, 2020).

Hay que considerar que las listas de verificación, en sus formas más elaboradas, destacan las relaciones entre las dimensiones particulares del proceso a evaluar (Kang *et al.*, 2014). Por ejemplo, en una investigación sobre el aprendizaje del inglés como segunda lengua (McCollum y Reed, 2020), la lista de verificación permitió a los estudiantes establecer sus propias

metas y centrarse más en las tareas, y a los docentes mejorar el currículo y comunicar mejor los resultados. El uso de listas de verificación resulta particularmente útil en la metodología de aprendizaje basado en problemas (An, 2013) y también en la enseñanza de las lenguas (Gori, 2014). Algunas críticas al uso de las listas de verificación son las de Meola (2014), que considera que puede ser un instrumento que, en manos de los estudiantes, favorece una evaluación mecánica que no facilita el pensamiento crítico. Igualmente, Kayapınar (2014) en su estudio no encontró buenos resultados de interfiabilidad en la aplicación de listas de verificación y escalas de evaluación de textos escritos.

Si comparamos las ventajas y desventajas entre las rúbricas y las listas de verificación, encontramos que las primeras tienden a ser más generalizables que las últimas (Uzun *et al.*, 2019), aunque las listas tienen un diseño más ágil y ofrecen resultados más consistentes. Desde el punto de vista de los inconvenientes, en el caso de la rúbrica, la descripción de los indicadores no es necesariamente homogénea en todos los niveles, y en el caso de la lista de verificación, puede dar sensación de rigidez y poca flexibilidad. La lista de verificación es similar a la rúbrica en que enumera los “criterios o lo que cuenta”, pero no describe los “niveles de calidad de excelente a pobre”, que es una característica definitoria de una rúbrica. Por ello, se plantean herramientas que combinan lo mejor de cada instrumento (Bharuthram y Patel, 2017), para abarcar todos los elementos, que, desde la perspectiva académica, se deben verificar para evaluar un trabajo de manera más cualitativa. Por su parte, Reddy y Andrade (2010) insisten en que, además del contenido, la claridad en el lenguaje es un aspecto fundamental para determinar la validez de una prueba de evaluación, ya que la ambigüedad hace que los evaluadores puedan interpretar el criterio de maneras diferentes.

De acuerdo con Pathirage *et al.* (2004), hay una serie de elementos relevantes a evaluar en

los TFG que comprenden: 1) la introducción (resumen, objetivos, justificación y contexto); 2) el conocimiento en la disciplina (fuentes, teorías y objeto de estudio); 3) la metodología (métodos experimentales, diseño y aspectos éticos); 4) el análisis y discusión de los resultados (presentación, claridad y argumentos lógicos); 5) la conclusión y recomendaciones; 6) la presentación y comunicación del trabajo (estructura, organización, referencias y lenguaje) y, por último, 7) la relevancia, originalidad, contribución, trabajo futuro, impacto y dificultad. El propósito de esta investigación ha sido diseñar un nuevo instrumento de evaluación de todos los elementos de un TFG en los estudios en Maestro en Educación. Partimos del marco jurídico que regula los TFG (Real Decreto 1393/2007), la Orden ECI/3857/2007 y la orden ECI/3854/2007; así como de la normativa de la universidad, la guía específica de TFG de los Grados en Maestro y la rúbrica anterior que había mostrado serias limitaciones. Nos planteamos una herramienta evaluativa objetiva, rigurosa y completa, que reduzca ambigüedades y diferencias de interpretación entre evaluadores y que pueda orientar a estudiantes, directores y tribunales a la hora de desarrollar y evaluar las competencias adquiridas en el grado, a través de un amplio conjunto de indicadores. Que los estudiantes dispongan de los instrumentos con los que se evaluará su trabajo les permite autorregular su aprendizaje. El fin último es mejorar la evaluación como elemento fundamental en el aprendizaje de los estudiantes.

Método

Consta de dos fases: el diseño del instrumento de evaluación y la validación del mismo, mediante la aplicación del coeficiente intraclass (ICC), que mide el grado de acuerdo o consistencia entre medidas cuya varianza puede ser atribuida a variaciones entre experimentadores.

Estudio I. Diseño del instrumento

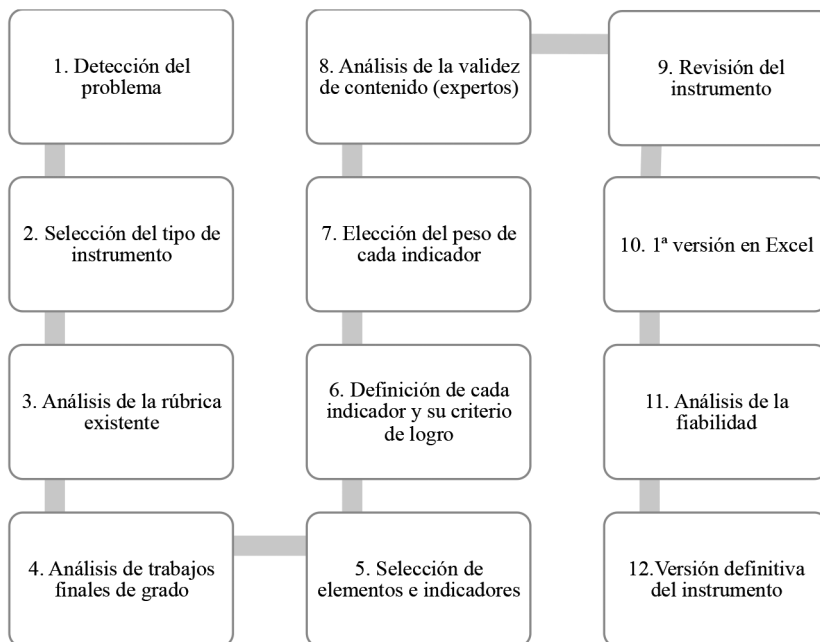
En la Universidad Internacional de La Rioja (UNIR), los TFG de Maestro en Educación Infantil y Primaria suponen la demostración de la adquisición de las competencias desarrolladas para el desempeño de esta profesión mediante la presentación de un trabajo con una propuesta educativa.

Para diseñar el nuevo instrumento de evaluación, seguimos las directrices de Stufflebeam (2000): centrar el objetivo de la herramienta a partir de la detección del problema, establecer las dimensiones, seleccionar y definir las categorías de análisis, formatear, aplicar y revisar herramienta. Se constituyó un grupo de trabajo de siete expertos del ámbito escolar y universitario. En la figura 1 se pueden ver las fases por las que atravesó el diseño de la herramienta, entre los meses de octubre 2018 a mayo 2020.

En primer lugar, partimos del problema detectado (fase 1) para identificar los objetivos del

trabajo. El actual modelo de rúbrica de los grados en Maestro en Educación ha presentado limitaciones tras varios años de uso, especialmente por la ambivalencia en muchas de sus definiciones. En ocasiones los alumnos han manifestado que sus directores les alaban sus trabajos y, sin embargo, sus calificaciones finales no se corresponden con las expectativas generadas. También sucede que los propios directores de los estudiantes, que no otorgan la calificación final, realizan informes con propuestas de notas de sus tutorados a los tribunales a partir de la aplicación de la rúbrica tradicional, que luego no se ven correspondidas con la calificación dada por los tribunales e incluso generan alguna tensión entre los docentes por la disparidad en la evaluación. Por otra parte, por su planteamiento holístico, encontramos para cada criterio varios descriptores en un mismo nivel, donde unos podían cumplirse y otros no, por lo que valorar la adquisición de un determinado logro terminaba siendo algo subjetivo. En la tabla 1 podemos observar un ejemplo de esta problemática:

FIGURA 1. Proceso de elaboración del instrumento de evaluación de TFG



Fuente: elaboración propia.

TABLA 1. Ejemplo de indicadores y niveles de logro en la rúbrica de evaluación de TFG de los Grados en Maestro de Educación Infantil y Primaria

Indicador	Niveles de logro	
	Nivel 3 (notable)	Nivel 4 (sobresaliente)
Marco teórico	<ul style="list-style-type: none"> • Está relacionado el tema y los objetivos • Hay unidad discursiva dentro de cada apartado, pero no entre apartados para el nivel educativo previsto • La mayoría de las fuentes son relevantes y la mayoría son primarias 	<ul style="list-style-type: none"> • Está perfectamente relacionado con los objetivos • Están definidos todos los conceptos y justificadas todas las afirmaciones • Los contenidos dentro y entre apartados están enlazados y siguen una unidad discursiva • Todas las fuentes son relevantes

Fuente: elaboración propia.

Como se puede observar, un trabajo podría presentar que los objetivos estuvieran perfectamente relacionados con el contenido del marco teórico (nivel 4), pero no tener unidad discursiva entre apartados (nivel 3). Al realizar un análisis pormenorizado de cada característica e indicador, se detectaron múltiples incoherencias como esta.

En la segunda fase de diseño se estuvieron valorando diferentes alternativas para superar los problemas y dificultades encontrados en la rúbrica tradicional. La alternativa para resolver el problema consistía en elegir entre elaborar una rúbrica analítica, que podría llegar a ser atomística y poco eficiente, aunque práctica a la hora de cumplimentarla, o elaborar una lista de verificación o control, fácil de cumplimentar, pero pobre en la descripción de los niveles de logro alcanzados. Siguiendo a Bharuthram y Patel (2017), se decidió optar por un diseño mixto basado en una rúbrica-*checklist* con diversas dimensiones e indicadores relacionados. Al combinar la rúbrica con la lista de verificación, se facilita la consecución de una información evaluativa completa y a la vez cualitativa, con referencia la calidad del trabajo.

A continuación, se revisó la rúbrica existente para extraer aquellos elementos que podían ser útiles para la nueva herramienta. Se seleccionaron aquellos relacionados con los componentes

estructurales del trabajo (los relativos al índice, formato, ilustraciones) y con los contenidos de la propuesta educativa (la introducción, objetivos, marco teórico, propuesta didáctica, conclusiones y referencias bibliográficas).

En la cuarta fase se escogieron por conveniencia cuatro TFG (dos de Infantil y dos de Primaria) para realizar un análisis pormenorizado del contenido, estructura y apartados de cada trabajo y determinar todos los elementos relevantes a evaluar.

En la fase 5 se debatió extensamente la relación de elementos encontrados en la fase anterior para, finalmente realizar una selección de los elementos susceptibles de evaluación de un TFG, tal como se recoge en la tabla 2, destacando la importancia de evaluar no solo el contenido del trabajo, sino todo lo relativo al formato y a la calidad del texto escrito.

A lo largo de la fase 6 se analizó el conjunto de apartados y se fueron definiendo los indicadores que conformarían la rúbrica-*checklist* final, junto con el criterio cualitativo de logro. En la tabla 3 se muestran algunos ejemplos de redacción de los indicadores de cada apartado.

En la fase 7 y conforme a la opinión consensuada de los expertos, se estableció qué peso específico, en el conjunto de la calificación, debía tener

TABLA 2. Elementos a evaluar de un trabajo final de grado en Maestro de Educación Infantil y Primaria

Contenidos	Aspectos de formato	Aspectos expresivos
<ul style="list-style-type: none"> • Portada • Resumen y palabras clave • Índice • Introducción • Justificación • Referencias legislativas • Competencias • Objetivos • Marco teórico • Propuesta (metodología, actividades, medios, recursos, temporalización, atención a la diversidad, evaluación) • Conclusiones • Consideraciones finales • Referencias 	<ul style="list-style-type: none"> • Tipo de letra en título y texto • Espaciado y sangrías • Numeración correlativa • Justificación y márgenes • Elementos visuales (tablas, gráficos, imágenes) 	<ul style="list-style-type: none"> • Corrección lingüística: ortografía, corrección gramatical, expresión escrita, uso correcto de signos de puntuación • Lenguaje: uso indistinto del lenguaje impersonal o plural mayestático; utilización de vocabulario académico (uso de sinónimos, términos adecuados según el contexto), uso de terminología propia del grado y área de conocimiento específica • Coherencia expositiva (uso de conectores), relación entre ideas, expresión fluida (frases cortas, no abuso de subordinadas), concordancia y claridad en la redacción • Citación correcta de referencias (normativa APA) y de la legislación vigente

Fuente: elaboración propia.

TABLA 3. Ejemplos de indicadores de la rúbrica-checklist y criterios

Apartado	Indicador	Criterio
Justificación	Aclara el porqué o razón de ser de la propuesta dentro del proceso de enseñanza-aprendizaje	Coherencia con la legislación vigente
Objetivos	Señala los objetivos que el estudiante debe alcanzar	Adecuación de competencias, objetivos y contenidos al nivel y destinatarios
Medios y recursos	Especifica todos los recursos y medios que va a utilizar, indicando la sesión o actividad vinculada	Adecuación de la metodología con los espacios y recursos y las medidas de atención a la diversidad
Temporalización	Presenta la distribución temporal de la propuesta (sesiones/actividades) dentro del calendario escolar	Secuencia lógica de las actividades
Estándares de aprendizaje evaluables	Indica en cada estándar de aprendizaje qué competencia y objetivo permite evaluar	Coherencia entre los criterios, estándares, los instrumentos de evaluación y los momentos de la evaluación

Fuente: elaboración propia.

cada uno de los apartados, considerando que, de los 100 puntos de la calificación final del trabajo, el 70% de ellos debía corresponder a la parte escrita, según lo establecido en la normativa de la universidad y en la propia memoria de la titulación. Se decidió dar más peso, dentro de la parte escrita, a los apartados donde el alumno debía

mostrar el dominio de las competencias adquiridas durante su grado, como son el marco teórico y la propuesta educativa. Esta distribución se muestra en la tabla 4, resolviendo así otro de los problemas que tenía el instrumento anterior, donde la parte de la propuesta didáctica tenía menos peso en la nota final.

TABLA 4. Distribución del peso de cada apartado escrito en el conjunto de la calificación

Apartado	Peso
Formato	4%
Expresión y corrección lingüística	6%
Referencias bibliográficas	10%
Portada, resumen y palabras claves	1%
Introducción	4%
Objetivos	5%
Marco teórico	15%
Propuesta	20%
Conclusiones	4%
Consideraciones finales	1%

Fuente: elaboración propia.

Con esta versión preliminar del instrumento, definidos los elementos, indicadores de logro y peso ponderado, se procedió a realizar un análisis de la validez del contenido mediante la técnica de entrevistas cognitivas (Estefanía y Zalazar-Jaime, 2018) a un panel de expertos, evaluadores habituales de trabajos finales de grado en Maestro en Educación Infantil y Primaria. Estos aplicaron esta primera versión a 4 TFG y después cada experto verbalizó las puntuaciones que había dado a cada apartado. Se expusieron las diferencias de interpretación de los indicadores y se reformularon algunas definiciones que se consideraron imprecisas. Una vez modificado y revisado el instrumento, se repitió el proceso con cuatro nuevos TFG para confirmar la validez de contenido entre los expertos. Posteriormente, se elaboró una primera versión aplicable de la rúbrica-checklist en formato Excel.

Estudio II. Análisis de la fiabilidad de la herramienta

Para valorar la fiabilidad interobservadores de la rúbrica-checklist, se aplicó el coeficiente de correlación intraclase (ICC), apropiado para cuantificar la concordancia entre distintas mediciones de una variable numérica y cuando se disponen de más de dos jueces. Este coeficiente mide el grado de acuerdo o consistencia entre medidas cuya varianza puede ser atribuida a variaciones

entre observadores, cuando más de dos experimentadores realizan medidas sobre una misma muestra. Cuando los valores de ICC $> .90$ se considera que existe una muy buena concordancia (Pita Fernández *et al.*, 2003).

Para el cálculo se utilizó el paquete estadístico SPSS v26. Esta investigación cumple con los criterios establecidos sobre Reglamento General de Protección de Datos de la Unión Europea (<https://www.boe.es/doue/2016/119/L00001-00088.pdf>).

Seis evaluadores independientes aplicaron la rúbrica-checklist a veinte trabajos (anónimos), diez TFG de Infantil y diez TFG de Primaria escogidos mediante un muestreo por cuotas entre los trabajos presentados al finalizar el grado con varias horquillas de calificación, por lo que la rúbrica-checklist fue aplicada 120 veces. Los evaluadores eran profesores estables del claustro de los grados en Maestro con más de cinco años de experiencia en la dirección de TFG y que habían sido miembros de tribunales.

Resultados

La herramienta finalmente diseñada se encuentra en el anexo 1 con todas las dimensiones, subdimensiones e ítems. Podemos observar un detalle en el ejemplo de la tabla 5 sobre el marco

TABLA 5. Ejemplo de una parte de la rúbrica-checklist diseñada

Marco teórico: fundamentación de la propuesta en fuentes bibliográficas primarias (fuentes fiables) para sostener conceptualmente el texto - 15 puntos	1. Los contenidos se relacionan con los objetivos teóricos del apartado <i>Objetivos</i>
	2. Existe relación entre la propuesta educativa y su fundamentación teórica
	3. Hay una aproximación de lo más general a lo más particular (progresión de ideas)
	4. El planteamiento de los elementos teóricos está definido de manera clara y concisa y acorde con el contenido de la parte práctica
	5. Hay una relación coherente entre los distintos apartados del marco teórico
	6. La información o ideas del marco teórico están citadas o justificadas desde al menos algún autor
	7. Se señalan, comentan y analizan investigaciones similares

Fuente: elaboración propia.

TABLA 6. Coeficientes de correlación intraclase (ICC)

	Correlación intraclase Medidas promedio	95% de intervalo de confianza		Prueba F con valor verdadero 0			
		Límite inferior	Límite superior	Valor	gl1	gl2	Sig
Escala Estructura	.908	.828	.959	10.909	19	95	0
Escala Contenido	.946	.899	.976	18.499	19	95	0
Escala Global	.961	.926	.982	25.36	19	95	0

Fuente: elaboración propia.

teórico, con la definición del apartado, la lista de indicadores, así como la definición de cada uno de ellos, de manera que el evaluador únicamente debe señalar si el indicador se ha logrado o no de acuerdo a la descripción contenida.

En la tabla 6 se presentan los resultados de fiabilidad de la rúbrica-checklist. Se han calculado tres índices distintos, uno para la parte de estructura, otro para el apartado de contenido y otro para la escala global que aglutina la estructura y el contenido.

La escala de estructura valora la coincidencia en tres aspectos que tienen un peso del 20% en la calificación final y que se refieren al formato, la

expresión y corrección lingüística y las referencias bibliográficas. En este caso, la correlación es de .908 por lo que la coincidencia entre evaluadores es alta.

La escala de contenido valora la coincidencia en tres aspectos que tienen un peso del 50% en la calificación final, y que se refieren al resumen, la introducción, los objetivos, el marco teórico, la propuesta, las conclusiones y las consideraciones finales. En este caso, la correlación es de .946, por lo que la coincidencia entre evaluadores es alta.

La escala global valora la coincidencia de las puntuaciones totales obtenidas en la parte

escrita de TFG. La correlación es de .961, lo que indica una alta coincidencia entre los jueces sobre la nota final del estudiante.

Discusión y conclusiones

El Plan Bolonia —que rige actualmente los planes de estudio en las universidades europeas— puso un fuerte énfasis en las competencias profesionales y en la necesidad de demostrar la integración de todas ellas (Valderrama *et al.*, 2010) en los TFG. En este proyecto nos planteamos el desafío de disponer de una herramienta para una evaluación objetiva, completa y fiable de la adquisición de esas competencias. En esta investigación hemos combinado los principios de dos herramientas de evaluación, conjuntando los niveles de logro de una rúbrica con la especificación que comporta una lista de verificación. En este instrumento se valoran los elementos que, de acuerdo con Pathirage *et al.* (2004), se deben recoger en un instrumento de evaluación de un TFG: resumen, objetivos, justificación, contexto, conocimiento de la disciplina (a través del marco teórico), metodología y propuesta, análisis, resultados, conclusión y recomendaciones, relevancia, originalidad, contribución, trabajo futuro, impacto y dificultad.

La herramienta resultante supone una innovación sobre las rúbricas tradicionales por varios motivos. En primer lugar, recoge todos los indicadores que debe contener un TFG en Maestros de Educación, de tal manera que el trabajo final de un estudiante y la rúbrica-*checklist* son las dos caras de una misma moneda. Un estudiante será capaz de llevar adelante con éxito su trabajo teniendo a la vista todos los elementos que constituyen su evaluación. La ganancia en transparencia y orientación es evidente. Con esta herramienta conseguimos uno de los grandes retos de la evaluación, que esta suponga realmente una ayuda para el estudiante y que no sirva solamente para calificar. Además, al establecer cada indicador parejo a los apartados del TFG que se deben entregar en el informe,

brinda al estudiante la evidencia de si cumple o no un criterio en particular y este puede detectar perfectamente cómo se encuentra su trabajo y, por tanto, es una herramienta que permite y apoya la autorregulación y la autoevaluación. La evaluación y la enseñanza no deben ser tratadas como elementos separados o independientes, sino que están estrechamente entrelazadas. Por ello, una herramienta como la que presentamos supone dar apoyo tanto a la instrucción como a la propia autoevaluación de las competencias de los estudiantes. De hecho, la rúbrica-*checklist* se puede considerar un “andamio” en el sentido anglosajón de la palabra, ya que supone una palanca para la autorregulación del estudiante sobre su trabajo y su progreso. Con este apoyo, la calidad de las decisiones a la hora de plantear el trabajo final es mucho mayor (Bharuthram y Patel, 2017). En este sentido, esta herramienta potencia mucho más la transparencia de la evaluación (Bearman y Ajjawi, 2021) al disponer el estudiante de criterios concretos y precisos (Jonsson, 2014). Wiggins (1998) sostiene que la autoevaluación es inseparable de cualquier evaluación que tenga como objetivo mejorar el aprendizaje, y hay que intentar que estas herramientas evaluativas de TFG sirvan para estos fines y no solamente con propósitos de clasificación o calificación (Bharuthram, 2015).

Otra de las grandes ventajas de este instrumento es su facilidad, claridad y homogeneidad de criterio a la hora de aplicarlo por parte de los profesores. El desarrollo conjunto de las dimensiones e indicadores, a partir de los trabajos, ha conllevado que no se deje de lado ninguno de los criterios a evaluar y que además se puedan concretar las relaciones entre las diferentes dimensiones, como plantean Kang *et al.* (2014). Este formato de rúbrica-*checklist* invita a los evaluadores a razonar mejor sobre cómo se produce la relación entre las distintas dimensiones (contenidos y su calidad; aspectos de formato y aspectos expresivos) sin generar ambigüedades interpretativas, pero también incorporando todos los elementos a evaluar en un TFG y ponderándolos de

acuerdo con su importancia. Los datos de fiabilidad interjueces revelan alta concordancia, por lo que el riesgo de subjetividad se reduce considerablemente. Y eso es importante porque las rúbricas tradicionales normalmente establecen criterios de logro en los que la frontera que separa uno y otro suele ser confusa. Esta rúbrica-*checklist* demuestra que se puede disponer de una herramienta de evaluación completa y fiable, útil para estudiantes y profesores. Esta alta concordancia ofrece garantías a la herramienta, máxime cuando cada vez más los sistemas y agencias de acreditación de la calidad solicitan evidencias de los niveles competenciales alcanzados. Por otra parte, desde el punto de vista de su usabilidad, esta rúbrica-*checklist* es un instrumento ágil y fácil de aplicar, ya que el propio diseño en Excel facilita su cumplimentación, así como el cálculo de la calificación.

En tercer lugar, consideramos que el potencial de esta herramienta para adaptarse a otros programas o áreas es alto. Lo interesante de esta rúbrica-*checklist* no es solo disponer de una herramienta de estas características para evaluar los TFG, sino su diseño y el consenso de los expertos sobre 73 indicadores que cubren un amplio espectro competencial (la comunicación escrita,

la capacidad de buscar y gestionar información, la habilidad de analizar, sintetizar y resolver problemas, etc.) y de contenidos curriculares (propuesta educativa, marco teórico) que se espera de un estudiante que se gradúa como Maestro en Educación Infantil o Maestro en Educación Primaria. Esta rúbrica-*checklist* rompe con dos de los problemas habituales de las rúbricas en uso: la falta de homogeneidad que suele darse en los diferentes niveles de logro y la imprecisión o generalidad con que suelen definirse los indicadores. Por otra parte, también supera alguna de las limitaciones de las clásicas listas de verificación que no suelen incluir criterios de logro. Estas ventajas del nuevo instrumento de evaluación pueden beneficiar a muchos planteamientos evaluativos de diferentes trabajos de grado y máster.

Aunque la muestra de TFG no es amplia, consideramos que es suficiente para confirmar la fiabilidad de la herramienta. Apreciamos como línea de investigación futura la elaboración de herramientas con parámetros y planteamientos similares para otros grados o másteres universitarios. Por otra parte, el análisis sobre la satisfacción de esta rúbrica-*checklist* como herramienta de autoevaluación y andamiaje para los estudiantes supondría también un estudio de interés.

Referencias bibliográficas

- An, Y. J. (2013). Systematic design of blended PBL: Exploring the design experiences and support needs of PBL novices in an online environment. *Contemporary Issues in Technology and Teacher Education*, 13(1), 61-79.
- Baharuddin, A., Gharbaghi, A., Ahmad, M. H. y Rosli, M. S. (2013). A Check List for Evaluating Persuasive Features of Mathematics Courseware. *International Education Studies*, 6(9). <http://dx.doi.org/10.5539/ies.v6n9p125>
- Bearman, M. y Ajjawi, R. (2018). Actor-network theory and the OSCE: formulating a new research agenda for a post-psychometric era. *Advances in Health Sciences Education*, 23(5), 1037-1049. <https://doi.org/10.1007/s10459-017-9797-7>
- Bharuthram, S. (2015). Lecturers' perceptions: the value of assessment rubrics for informing teaching practice and curriculum review and development. *Africa Education Review*, 12(13), 415-428. <https://doi.org/10.1080/18146627.2015.1110907>
- Bharuthram, S. y Patel, M. (2017). Co-constructing a rubric checklist with first year university students: A self-assessment tool. *Apples: Journal of Applied Language Studies*, 11(4), 35-55. <http://dx.doi.org/10.17011/apples/urn.201708073430>

- Bohórquez Gómez-Millán, M. y Checa Esquiva, I. (2019). Desarrollo de competencias mediante ABP y evaluación con rúbricas en el trabajo en grupo en Educación Superior. *REDU. Revista de Docencia Universitaria*, 17(2), 197-210.
- Brookhart, S. M. (2013). *How to Create and Use Rubrics for Formative Assessment and Grading*, ASCD.
- Cañadas, L. (2020). Evaluación formativa en el contexto universitario: oportunidades y propuestas de actuación. *Revista Digital de Investigación en Docencia Universitaria*, 14(2), e1214. <https://doi.org/10.19083/ridu.2020.1214>
- Dawson, P. (2017). Assessment rubrics: towards clearer and more replicable design, research and practice. *Assessment & Evaluation in Higher Education*, 42(3), 347-360. <https://doi.org/10.1080/02602938.2015.1111294>
- Gori, F. (2014). The European Language Portfolio and Languages for Specific Purposes: A project to develop “can do” descriptors focused on students' interests and motivation. *Language Learning in Higher Education*, 3(2), 305. <https://doi.org/10.1515/cercles-2013-0016>
- Jonsson, A. (2014). Rubrics as a way of providing transparency in assessment. *Assessment and Evaluation in Higher Education*, 39(7), 840-852.
- Kang, H., Thompson, J. y Windschitl, M. (2014). Creating opportunities for students to show what they know: The role of scaffolding in assessment tasks. *Science Education*, 98(4), 674-704. <https://doi.org/10.1080/02602938.2013.875117>
- Kayapınar, U. (2014). Measuring essay assessment: Intra-rater and inter-rater reliability. *Eurasian Journal of Educational Research*, 57, 113-136. <http://dx.doi.org/10.14689/ejer.2014.57.2>
- Mauri, T., Colomina, R. y De Gispert, I. (2014). Transformando las tareas de escritura colaborativa en oportunidades para aprender: ayuda educativa y uso de rúbricas en la Educación Superior. *Cultura y Educación*, 26(2), 298-348. <https://doi.org/10.1080/11356405.2014.935111>
- McClearyCale, C. G. y Furtak, E. M. (2020). The SABEL Checklist Science classroom assessments that work for emergent bilingual learners. *Science Teacher*, 87(9), 38-48. <https://www.nsta.org/science-teacher/science-teacher-julyaugust-2020/saeb1-checklist>
- McCollum, R. M. y Reed, E. T. (2020). Developing a Badge System for a Community ESL Class Based on the Canadian Language Benchmarks. *Canadian Journal of Applied Linguistics*, 23(2), 228-236. <https://doi.org/10.37213/cjal.2020.30438>
- Meola, M. (2004). Chucking the Checklist: A Contextual Approach to Teaching Undergraduates Web-Site Evaluation portal: *Libraries and the Academy*, 4(3), 331-344. <https://doi.org/10.1353/pla.2004.0055>.
- Merrill, M. D. (2020). A Syllabus Review Check-List to Promote Problem-Centered Instruction. *TechTrends* 64, 105-123. <https://doi.org/10.1007/s11528-019-00411-4>
- Oakleaf, M. (2009). Using rubrics to assess information literacy: An examination of methodology and interrater reliability. *Journal of the American Society for Information Science and Technology*, 60(5), 969-983. <https://doi.org/10.1002/asi.21030>
- Orden ECI/3857/2007 [Ministerio de Educación y Ciencia]. Por la que se establecen los requisitos para la verificación de los títulos universitarios oficiales que habiliten para el ejercicio de la profesión de Maestro en Educación Primaria. 27 de diciembre de 2007.
- Orden ECI/3854/2007 [Ministerio de Educación y Ciencia]. Por la que se establecen los requisitos para la verificación de los títulos universitarios oficiales que habiliten para el ejercicio de la profesión de Maestro en Educación Infantil. 27 de diciembre de 2007.
- Ortega, D., Carcedo, B. y Blanco, P. (2018). El Trabajo Fin de Grado en Didáctica de las Ciencias Sociales: líneas, materias y temáticas. *Revista de Investigación en Didáctica de las Ciencias Sociales*, 3, 35-51.

- Pathirage, C. P., Haigh, R., Amaratunga, D., Baldry, D. y Green, C. M. (2004). *Improving Dissertation Assessment*. [Resumen de presentación de la conferencia]. Education in a Changing Environment 13th-14th, Universidad de Salford. <http://eprints.hud.ac.uk/id/eprint/22716/>
- Pausch, L. M. y Popp, M. P. (1997, April). *Assessment of information literacy: Lessons from the higher education assessment movement*. Paper presented at the meeting of the Association of College and Research Libraries, Nashville, TN.
- Pegalajar, M. C. (2021). La rúbrica como instrumento para la Evaluación de Trabajos Fin de Grado. *REICE. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 19(3), 77-96. <https://revistas.uam.es/reice/article/view/13134>
- Pita Fernández, S., Pértega Díaz, S. y Rodríguez Maseda, E. (2003). La fiabilidad de las mediciones clínicas: el análisis de concordancia para variables numéricas. *Cad Aten Primaria*, 10(4), 290-296.
- Proyecto Tuning (2009). *Una introducción a Tuning Educational Structures in Europe*. Universidad de Deusto.
- Real Decreto 1393 de 2007. Por el que se establece la ordenación de las enseñanzas universitarias oficiales. 29 de octubre de 2007. BOE-A-2007-18770.
- Reddy, Y. M. y Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, 35(4), 435-448. <https://doi.org/10.1080/02602930902862859>
- Sánchez, I. D. (2021). Propuesta de validación de la homogeneidad de un modelo de rúbrica de evaluación del TFG en el grado de enfermería de la UCM [tesis de doctorado, Universidad Complutense de Madrid].
- Shin, S. y Cheon, J. (2019). Assuring Student Satisfaction of Online Education: A Search for Core Course Design Elements. *International Journal on E-Learning*, 18(2), 147-164.
- Stufflebeam, D. L. (2000). *Guidelines for developing evaluation checklists: the checklists development checklist (CDC)*. The Evaluation Center Retrieved on January, 16, 2008.
- Tomas, C., Whitt, E., Lavelle-Hill, R. y Severn, K. (2019, September). *Modeling holistic marks with analytic rubrics*. In *Frontiers in Education* (vol. 4, p. 89). Frontiers. <https://doi.org/10.3389/educ.2019.00089>
- Uzun, N. B., Alici, D. y Aktas, M. (2019). Reliability of the analytic rubric and checklist for the assessment of story writing skills: g and decision study in generalizability theory. *European Journal of Educational Research*, 8(4), 169-180. <https://doi.org/10.12973/euler.8.1.169>
- Valderrama, E., Rullán, M., Sánchez, F., Pons, J., Mans, C., Giné, F., Seco, G., Jiménez L., Peig, E., Carrera, J., Moreno, A., García J., Pérez, J., Vilanova, R., Cores, F., Renau, J. M., Tejero, J. y Bisbal, J. (2010). La evaluación de competencias en los Trabajos Fin de Estudios. *IEEE-RITA*, 5(3), 107-114.
- Wiggins, G. (1998). *Educative assessment*. Jossey-Bass.
- Zornoza-Gallego, C. y Vercher, N. (2021). Evaluación de competencias genéricas... *Cuadernos Geográficos*, 60(1), 119-138.

Anexo 1. Rúbrica-checklist de evaluación

ESCALA ESTRUCTURA		
FORMATO (4 puntos)	Seguimiento de la plantilla (estructura y partes) - 1 punto	<ol style="list-style-type: none"> 1. Contiene todos los apartados. 2. Cumple la extensión del trabajo.
	Maquetación del trabajo y del texto - 2 puntos	<ol style="list-style-type: none"> 1. En general se deja el espacio correcto entre palabras. 2. En general se usan adecuadamente las sangrías. 3. En general se deja el espacio correcto entre párrafos y apartados. 4. La numeración de tablas/figuras es correlativa (orden correcto). 5. En general los textos y márgenes están correctamente justificados. 6. Los epígrafes nuevos se inician dónde deben (no se inician en la última línea de una página). 7. Se indican los anexos en el texto donde corresponde.
	Uso de elementos visuales (tablas, gráficos, imágenes) - 1 puntos	<ol style="list-style-type: none"> 1. Se incorporan elementos visuales (tablas y/o figuras) y de calidad (con buena resolución de imagen). 2. Se explican las tablas, figuras y gráficos en el contenido del texto. 3. Los elementos visuales son adecuados al texto (aportan valor a la explicación del contenido, no son meros elementos decorativos).
EXPRESIÓN Y CORRECCIÓN LINGÜÍSTICA (6 puntos)	Expresión escrita global/ corrección gramatical global <i>La valoración realizada por cada indicador de expresión escrita debe evaluarse de forma global dicotómica (sí cumple o no cumple).</i> - 3,5 puntos	<ol style="list-style-type: none"> 1. En general se usan estructuras gramaticales adecuadas. 2. En general se usan correctamente los signos de puntuación. 3. En general se relacionan correctamente las ideas (tanto en párrafos como entre apartados). 4. Hay coherencia expositiva (uso de conectores). 5. La escritura es fluida (frases cortas, no abuso de subordinadas). 6. En general la redacción es clara (comprensión de las ideas expuestas).
	Lenguaje - 2,5 puntos	<ol style="list-style-type: none"> 1. En general se redacta de forma académica (pasiva refleja o plural mayestático indistintamente) evitando la primera persona del singular. 2. En general se utiliza un vocabulario académico (evitando coloquialismos, usando sinónimos, términos formales y adecuados según temática ...). 3. En general se hace un uso de la terminología propia del grado y del área de conocimiento del TFG.
REFERENCIAS BIBLIOGRÁFICAS (10 puntos)	Citación <i>Sólo se trabajará con referencias (no bibliografía ni webgrafía consultada)</i> - 10 puntos	<ol style="list-style-type: none"> 1. Se utilizan al menos 10 fuentes primarias publicadas en revistas y/o libros de carácter académico (sin incluir las leyes recogidas en el marco legal). 2. Se incluye al menos, dos citas actualizadas de los últimos 5 años. 3. Al menos, una referencia proviene de un idioma que no es el castellano.

ESCALA CONTENIDO		
PORTADA, RESUMEN Y PALABRAS CLAVES (1 punto)	Portada – 0,25 puntos	1. El título refleja inequívocamente el contenido del trabajo y su contexto y/o destinatarios
	Resumen – 0,50 puntos	2. El resumen expone de forma adecuada y comprensible las ideas fundamentales del trabajo 3. El resumen está debidamente estructurado (recoge la relevancia/actualidad del tema, el objetivo general, las acciones de la propuesta y sus posibles conclusiones)
	Palabras clave – 0,25 puntos	4. Las palabras clave son adecuadas para el tema escogido y la etapa educativa (se indica la etapa educativa explícitamente)
INTRODUCCIÓN (4 puntos)	Introducción – 4 puntos	1. El tema del que trata del trabajo se presenta de manera explícita. 2. Se justifica la motivación personal para abordar el tema. 3. Se introduce una justificación conceptual o empírica (en relación con algún autor, estudio, etc.) 4. Se presenta la contextualización y a quién va dirigido.
OBJETIVOS (Cumplir las condiciones A y B es obligatorio. Si no estuvieran se considera que el nivel sería 0). – (5 puntos)	Condición necesaria: A) Señalar el objetivo general que se persigue con la elaboración del TFG. B) Señalar los objetivos específicos que desglosan el objetivo general. – 5 puntos	1. Señala un objetivo general y un mínimo de 3 y un máximo de 5 objetivos específicos. 2. Los objetivos específicos están vinculados tanto a aspectos teóricos como prácticos y legislativos. 3. Los objetivos están redactados en infinitivo. 4. En general los objetivos representan una idea clara. 5. En general los objetivos son inherentes al tema, alcanzables, operativos y coherentes. 6. Los objetivos hacen referencia al nivel educativo y/o destinatarios.
MARCO TEÓRICO (15 puntos)	Marco teórico: fundamentación de la propuesta en fuentes bibliográficas primarias (fuentes fiables) para sostener conceptualmente el texto- 15 puntos	1. Los contenidos se relacionan con los objetivos teóricos del apartado <i>Objetivos</i> . 2. Existe relación entre la propuesta educativa y su fundamentación teórica. 3. Hay una aproximación de lo más general a lo más particular (progresión de ideas). 4. El planteamiento de los elementos teóricos está definido de manera clara y concisa y acorde con el contenido de la parte práctica 5. Hay una relación coherente entre los distintos apartados del marco teórico. 6. La información o ideas del marco teórico están citadas o justificadas desde al menos algún autor. 7. Se señala, comenta y analiza investigaciones similares

<p>PROPUESTA (20 puntos)</p>	<p>Unidad didáctica (UD) / Proyecto de Intervención (PI) *</p> <p>Al menos el 80% del contenido de cada epígrafe cumple con el indicador. Si el contenido es insuficiente en calidad, se considera no cumplido – 20 puntos</p>	<ol style="list-style-type: none"> 1. Introducción: se presenta la temática de UD/PI y se vincula con el marco teórico del TFG. 2. Justificación: se aclara el por qué o la razón de ser de UD/PI dentro del proceso de enseñanza-aprendizaje. 3. Referencias legislativas: se indica la legislación nacional y autonómica que le afecta. 4. Competencias: se concreta qué competencias va a desarrollar con la UD [no aplicable a Infantil ni a PI]. 5. Objetivos: se señala el objetivo general y los objetivos específicos [operativos] que el estudiante debe alcanzar al finalizar UD/PI 6. Contenidos: se indica los contenidos que se van a trabajar en la UD/PI (conocimientos, habilidades, destrezas y actitudes). 7. Metodología: se describe las estrategias metodológicas que se van a poner en marcha en UD/PI. 8. Actividades: se presenta de formas concreta y clara las sesiones y/o actividades que va a realizarse para conseguirse los objetivos (con suficiente descripción, tiempo aproximado y materiales requeridos) 9. Medios y recursos: se especifica todos los recursos/medios que van a utilizarse indicándose la sesión o actividad vinculada. 10. Temporalización: se presenta la distribución temporal de UD/PI (sesiones/actividades) dentro del calendario escolar. 11. Medidas de atención a la diversidad / Diseño universal del aprendizaje: se proporciona una manera de optimizar el aprendizaje para enseñar de manera eficaz a un grupo diverso de estudiantes. 12. Evaluación: se describe el proceso de evaluación diseñado (evaluación del alumnado y autoevaluación de UD/PI). 13. Estándares de aprendizaje evaluables: se indica en cada estándar de aprendizaje qué competencia y objetivo de la UD te permite evaluar [no aplicable a Infantil ni a PI]. 14. Criterios de evaluación: se indica los criterios que va a utilizar para saber si se están consiguiendo o no los objetivos y competencias. 15. Instrumentos de evaluación: se presenta instrumentos adecuados y pertinentes para evaluar los objetivos y competencias. 16. La UD/PI presenta suficiente nivel de coherencia entre sus partes. 17. La UD/PI es potencialmente aplicable según etapa, desarrollo psico-evolutivo y contexto.
	<p>CONCLUSIONES (4 puntos)</p>	<p>Conclusiones – 4 puntos</p>

Abstract

The design and validation of an instrument for assessing undergraduate dissertations

INTRODUCTION. The dissertation represents the fulfillment of undergraduate studies in which the student demonstrates mastery and integration of the content and skills they have developed. Therefore, its evaluation requires complete and complex tools. The aim of this research is to design and validate a reliable instrument to evaluate the final dissertation of the degree of Bachelor Degree in Primary Education Teaching to avoid the deficiencies and problems commonly observed in the traditional rubrics. **METHOD.** A group of seven experts, both from the school and university environment, were trained to develop this instrument. Different evaluation tools and essays were analyzed for its elaboration. Subsequently, the reliability of the resulting instrument was studied by applying the intraclass correlation coefficient (ICC) to twenty dissertations by six independent experts. **RESULTS.** The result is an instrument that combines the best properties of the rubric and the checklist, with a concrete and precise breakdown of all the performance indicators that make up the written part of a dissertation, and that assesses elements considered central in a final manuscript: the contents and their quality, the aspects of format and expression. The intraclass correlation analysis offered excellent results in the measured scales, all of which were $>.90$, which shows their high consistency. **DISCUSSION.** These results allow confidence in the reliability of the tool. The instrument can be a clear support for dissertation supervisors and for the members of the dissertation panel due to its clarity, objectivity and ease of use. Likewise, it can be a useful tool for students' learning and favour their self-regulation and self-evaluation since it includes in detail all the indicators by which they will be evaluated.

Keywords: *Scoring rubric, Dissertation, Undergraduate study, Early Childhood Education, Primary Education.*

Résumé

Conception et validation d'un instrument d'évaluation du travail de fin d'études

INTRODUCTION. Le travail de fin d'études (TFE) constitue l'aboutissement d'études universitaires où l'étudiant démontre la maîtrise et l'intégration des contenus et des compétences développées. Par conséquent, son évaluation nécessite d'instruments complets et complexes. L'objectif de cette recherche est de concevoir et de valider un outil d'évaluation du TFE en éducation afin de corriger les lacunes et les problèmes observés dans les rubriques en usage. **MÉTHODE.** Un groupe de sept experts, issus du milieu scolaire et universitaire, a été constitué pour développer cet instrument. Différents outils et travaux d'évaluation ont été ainsi analysés. Par la suite, la fiabilité de l'instrument résultant a été étudiée en appliquant le coefficient de corrélation intraclasse (ICC) à vingt articles de six experts indépendants. **RÉSULTATS.** Le résultat est un instrument qui combine les meilleures propriétés d'une rubrique-checklist, avec le détail de tous les indicateurs qui composent la partie écrite d'un travail de fin d'études. Il évalue les éléments considérés comme centraux : le contenu et sa qualité ainsi que les aspects de format et d'expression. L'analyse de corrélation intraclasse a offert d'excellents résultats dans les échelles mesurées, qui étaient toutes supérieures à $.90$, ce qui montre sa grande cohérence. **DISCUSSION.** Ces résultats démontrent la fiabilité de l'outil. L'instrument, en raison de sa clarté, de son

objectivité et de sa facilité d'utilisation, permet aux directeurs de TFE et aux membres du jury de disposer d'un support clair. De même, il peut être un outil d'apprentissage utile pour les étudiants en favorisant leur autorégulation et leur autoévaluation, puisqu'il comprend en détail tous les indicateurs à partir desquels ils seront évalués.

Mots-clés: *Rubrique, Travail de fin d'études, Diplôme, Éducation maternelle, Éducation primaire.*

Perfil profesional de las autoras

Consuelo García Tamarit (autora de contacto)

Directora académica del Máster Universitario en Innovación Educativa de la Universidad Internacional de La Rioja (UNIR). Doctora en Educación y TIC por la Universitat Oberta de Catalunya (UOC) y licenciada en Psicología por la Universidad de Valencia. Es miembro del grupo de investigación sobre Competencias Académicas Universitarias de la UNIR y del grupo SINTE de la UAB.

Correo electrónico de contacto: consuelo.garcia@unir.net

Dirección para la correspondencia: Facultad de Educación. Universidad Internacional de La Rioja (UNIR). Av. de la Paz, 137, 26006 Logroño, La Rioja (España).

Paola Perochena González

Profesora universitaria desde 2009. Doctora y Máster en TIC para Educación. Máster en Humanidades y Licenciada en Psicología. Profesora Contratada Doctora que ha participado en proyectos de investigación I+D con financiación pública y privada. Ha coordinado diversos departamentos en varias instituciones relacionadas con la formación. Sus líneas de investigación: estilos de enseñanza y de aprendizaje, formación docente y evaluación.

Correo electrónico de contacto: paola.perochena@edu.uned.es

Lara Orcos Palma

Profesora universitaria desde 2015. Doctora en Didáctica de las Ciencias. Máster en Formación de Profesorado. Profesora contratada doctora que ha participado en proyectos de investigación I+D con financiación pública y privada. Ha sido profesora en la etapa de educación secundaria durante ocho años. Sus líneas de investigación son: didáctica de las matemáticas y las ciencias experimentales.

Correo electrónico de contacto: lara.orcos@unir.net