

**Vol. 24, nº2 (Julio, 2020)**

ISSN 1138-414X, ISSNe 1989-6395

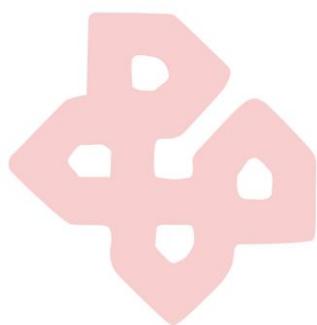
DOI: 10.30827/profesorado.v24i2.15151

Fecha de recepción: 09/04/2020

Fecha de aceptación: 30/05/2020

## **VALIDACIÓN DE UNA RÚBRICA PARA LA EVALUACIÓN DE TRABAJOS FIN DE MÁSTER**

*Validation of a rubric to assess the students' Final Master's Thesis*



*Ana García-Valcalver Muñoz-Repiso, Azucena Hernández Martín, Marta Martín del Pozo y Susana Olmos Migueláñez*

*Universidad de Salamanca*

*E-mail: [anagv@usal.es](mailto:anagv@usal.es); [azuher@usal.es](mailto:azuher@usal.es);*

*[mmdp@usal.es](mailto:mmdp@usal.es); [solmos@usal.es](mailto:solmos@usal.es)*

*ORCID ID: <https://orcid.org/0000-0003-0463-0192>;*

*<https://orcid.org/0000-0002-6731-7710>;*

*<https://orcid.org/0000-0002-1971-7904>;*

*<https://orcid.org/0000-0002-0816-4179>*

### **Resumen:**

Las rúbricas se han convertido en un instrumento de evaluación de competencias apropiado para valorar el proceso y los resultados de aprendizaje en el contexto de la educación superior, así como para orientar el trabajo de los estudiantes. En este trabajo se presenta una rúbrica para evaluar el Trabajo Fin de Máster (TFM) de los estudiantes y el proceso de validación de la misma. La rúbrica se compone de 4 dimensiones: aspectos formales, contenido y metodología, resultados y conclusiones, y relevancia de las aportaciones para la mejora educativa. Se ha aplicado a 34 TFM que han recibido un total de 136 valoraciones por parte de los docentes. Los resultados muestran un grado de fiabilidad excelente en base a la concordancia inter-jueces en la puntuación final de la rúbrica (CCI = 0,80), si bien la concordancia en las categorías cualitativas de evaluación es más baja (valores Kappa menores a 0,40). Se garantiza su validez de contenido a través del juicio de expertos y la validez criterial mediante correlación Pearson con la variable externa "nota media del alumnado en el máster exceptuando el TFM" que aporta un valor de 0,58. Así pues, la rúbrica que se propone resulta ser un instrumento de evaluación riguroso cuyo uso podría transferirse a otros contextos similares.

**Palabras claves:** *competencias; educación superior; evaluación; rúbrica; Trabajo Fin de Máster; validez.*

## Abstract:

Rubrics have become an appropriate competence assessment instrument to assess the learning process and outcomes in the context of higher education, as well as to guide students' work. This paper presents a rubric to assess the students' Final Master's Thesis (FMT) and its process of validation. The rubric is made up of 4 dimensions: formal aspects, content and methodology, results and conclusions, and relevance of the contributions for educational improvement. It has been applied to 34 FMTs that have been received 136 ratings from teachers. The results show an excellent degree of reliability based on the inter-judge agreement in the final score of the rubric (CCI = 0.80), although the agreement in the qualitative assessment categories is lower (Kappa values lower than 0.40). Its content validity is guaranteed through expert judgment and criterial validity through Pearson correlation with an external variable "students' average grade in the Master's Degree except the FMT", which provides a value of 0.58. Thus, the proposed rubric turns out to be a rigorous assessment instrument whose use could be transferred to other similar contexts.

*Key words:* competencias; Final Master's Thesis; higher education: assessment; rubric; validity.

## 1. Presentación y justificación del problema

Las universidades españolas han experimentado cambios notables como consecuencia del proceso de implantación del Espacio Europeo de Educación Superior, asumiendo el enfoque de enseñanza y aprendizaje por competencias, lo que ha supuesto el replanteamiento de las formas de enseñar de muchos docentes con la finalidad de que su alumnado pueda desarrollar todas las competencias propias del perfil profesional para el que se está formando.

Además, el replanteamiento de las metodologías didácticas obliga necesariamente también al profesorado universitario a pensar en nuevas formas de evaluación de los productos de aprendizaje, por considerar que el modo en que los docentes evalúan es una cuestión trascendental, con independencia del nivel educativo en el que éstos se sitúen. Junto a los demás elementos curriculares y en estrecha interacción con todos ellos, la evaluación es clave en cualquier nivel educativo, y en el ámbito universitario se han introducido diversas formas de evaluación continua y formativa, que apuestan por una nueva forma de entender la evaluación, de acuerdo con la cual:

El cambio no radica tanto en introducir nuevos elementos, aplicar metodologías diferentes, o utilizar instrumentos innovadores, sino en reorientar las herramientas, métodos y criterios de evaluación que tradicionalmente hemos venido utilizando, de tal modo que nos permitan alcanzar este objetivo: poder establecer en qué grado el estudiante posee la competencia o competencias que le habilitan para su ejercicio profesional (De la Mano y Moro, 2009, p.15).

El uso de las rúbricas como instrumentos de evaluación surge en el contexto de dicho cambio, en la medida en que permite valorar el logro alcanzado por los alumnos en las diferentes tareas que puede ir desarrollando. Como señalan diferentes especialistas en torno a este tema, si las rúbricas han sido construidas de modo que se

garantice su fiabilidad y validez para evaluar los niveles de logro de un estudiante, éstas pueden ser consideradas como herramientas con mucho potencial en el contexto de una evaluación cualitativa y cuantitativa de competencias en la Educación Superior, que es el ámbito en el que nos centramos en el presente estudio (Cebrián, 2014; Fraile, Pardo y Panadero 2017; Velasco y Tojar, 2018; Villa y Poblete, 2011).

La rúbrica, como instrumento de evaluación, viene a salvar la dificultad de valorar de forma objetiva y fiable, las tareas, ejercicios y trabajos realizados por los estudiantes cuando éstos tienen un contenido abierto, creativo o basado en la investigación. Aspecto especialmente relevante cuando estos resultados son la base de la calificación del estudiante en una asignatura y donde interviene más de un docente en el proceso de evaluación, como es el caso de los TFG y TFM. Situaciones en las cuales resulta especialmente necesario tener unos criterios comunes para emitir un juicio preciso sobre las evidencias presentadas, lo cual constituye el objetivo de diseñar y utilizar una rúbrica de evaluación.

En este trabajo se presenta el proceso de diseño y validación de una rúbrica para la evaluación de los Trabajos Fin de Máster (TFM), pertenecientes a un Máster que se imparte actualmente en la modalidad presencial en la Facultad de Educación de la Universidad de Salamanca. El profesorado que participaba en el mismo ya disponía desde los inicios de su impartición, de un conjunto de indicadores, acordados entre todos, para la valoración de los TFM. En el contexto de la Educación Superior se cuenta con variados ejemplos de rúbricas para valorar tareas y asignaturas de distintas titulaciones universitarias (Chica, 2011; Díaz-Barriga y De la Cruz, 2011; Expósito, Nicolau, Tomas Miquel 2017; Madariaga y Lekue, 2019), pero al estimar que no todas se adaptan a los rasgos específicos y competencias del trabajo solicitado en el TFM del citado Máster, se consideró preciso la elaboración de una rúbrica concreta que aumentase el rigor y objetividad de la valoración de una asignatura, el TFM, que posee un número significativo de créditos en la titulación (9 de 60 ECTS). Esta rúbrica, además, podría orientar a los estudiantes en su elaboración y servir, atendiendo al proceso de validación de la misma, como marco de referencia para la valoración de los TFM en otras titulaciones universitarias.

### **1.1. Caracterización de la rúbrica como una herramienta para la evaluación de competencias**

Por su naturaleza, toda competencia implica diferentes niveles de adquisición o de logro, por lo que, como ya se ha señalado, las rúbricas son herramientas que, por su idiosincrasia, caracterizada por la propia construcción de las mismas, permiten la valoración sistemática de esos niveles en un contexto de evaluación formativa; tanto para la autoevaluación, la evaluación entre pares, como para el trabajo colaborativo e interdisciplinar (Cebrián y Monedero, 2014). Las rúbricas, como señala García Sanz (2014), están conformadas por una serie de criterios que definen la calidad de las competencias objeto de evaluación, en términos de diferentes niveles de desempeño, posibilitando, en consecuencia, determinar la calidad y cantidad del aprendizaje realizado.

Las experiencias, tanto nacionales como internacionales, en el empleo de las rúbricas, han demostrado sus beneficios en los procesos de evaluación (Cano, 2015; Cebrián, 2007; García Sanz, 2014; Rochele, Tractenberg, Kevin y FitzGerald, 2012; Velasco y Tójar, 2018), entre los cuales se podrían resaltar:

- La posibilidad de reflexión de los estudiantes, a partir de los resultados obtenidos con su empleo, por lo que el valor formativo de las rúbricas es un elemento clave en la caracterización de las mismas, y permite al alumnado conocer en qué medida han alcanzado determinados objetivos o en qué deben incidir de modo más específico.
- El valor de construirla y consumirla conjuntamente por los estudiantes y el docente, seleccionando aquellos indicadores que realmente permitan valorar en qué medida se consiguen los objetivos de aprendizaje previstos. En consecuencia, los alumnos no sólo conocen dichos indicadores sino que, como señala Cano (2015), pueden hacerlos suyos al haber participado en su elaboración, y considerarlos como referencia para mejorar su rendimiento en una determinada tarea o materia.
- Relacionado con el aspecto anterior, se puede referir además el potencial de las mismas para estimular un aprendizaje profundo y de calidad por parte de los estudiantes, al conocer con anterioridad a la realización de la actividad o trabajo cuáles son las expectativas concretas del profesor sobre lo que se espera del producto realizado, a partir de la rúbrica presentada por éste, o confeccionada conjuntamente con los estudiantes.

En suma, las rúbricas constituyen una herramienta de evaluación de una gran variedad de competencias que el estudiante puede demostrar en el contexto de una tarea, más allá de la realización de pruebas tradicionales como los exámenes de carácter memorístico. Pero además, el hecho de que estas herramientas estén constituidas por descriptores concretos que indican diferentes niveles de logro, favorece una evaluación más objetiva de aspectos que son complejos, no siempre precisos y, en ocasiones subjetivos (Hernández, 2016, p.6); ayudando, como señala Gairín (2009), a lograr una mayor consistencia entre las expectativas y los resultados. Como hemos indicado en el estudio ya referido (Hernández, 2006), en el caso de la valoración de los Trabajos de Fin de Máster, las rúbricas se pueden emplear para:

(...) mejorar el proceso de realización, aportando orientaciones precisas que mejoren el rendimiento de los estudiantes al elaborar dichos productos. Y finalmente pueden contribuir a sistematizar el proceso de evaluación y valorar el grado de aprendizaje adquirido por el estudiante de manera más rigurosa, proporcionando información en torno a aquellos aspectos que pueden ser susceptibles de mejora (p. 7).

Aunque son numerosas las ventajas enumeradas, diferentes especialistas ponen en evidencia algunas de sus limitaciones. Bailey (2014) alude, por ejemplo, al peligro de valorar en exceso los resultados que se obtienen con su empleo. Y los propios estudiantes perciben en ocasiones que las rúbricas generan estrés y una mayor exigencia de trabajo (García, Belmonte y Nicolásac 2017; Panadero y Romero, 2014).

Junto a estas dificultades, es importante referirse también a la falta de exigencia que en ocasiones se percibe en el diseño de estas herramientas, en relación con la fiabilidad y validez de contenido, constructo y criterio; así como a las limitaciones que conlleva su empleo cuando los indicadores y criterios no son comprendidos correctamente por los evaluadores, por lo que es muy compleja su interiorización para utilizar la rúbrica con la finalidad con la que ésta fue planificada (Hernández, 2006).

### 1.2. Diseño de una rúbrica para evaluar un TFM

La utilidad de las rúbricas radica, entre otras cuestiones que ya se han indicado, en su capacidad para valorar aquellos trabajos que impliquen una participación activa e investigadora del alumnado. Por ello, si estos instrumentos no están bien diseñados, tanto en términos técnicos como pedagógicos, no serán herramientas válidas para la valoración de este tipo de actividades. De acuerdo con ello, autores como Popham (1997), García, Terrón y Blanco (2009), y Zazueta y Herrera (2008), entre otros, plantean algunas indicaciones para diseñar rúbricas rigurosas, entre las que destacamos las siguientes:

- Analizar las características y contenidos que definen el trabajo que se va a efectuar.
- Seleccionar criterios con un nivel adecuado de especificidad, de modo que la valoración de los mismos no entrañe dificultad por su excesiva generalidad o por una exhaustividad extrema, por ejemplo.
- Detallar cuidadosamente y con claridad las finalidades, niveles de logro y competencias que deben quedar reflejados en los trabajos efectuados por los estudiantes; así como también qué criterios de evaluación se van a considerar y cuál será la calificación que se otorgará finalmente a cada uno de ellos. En este sentido, el establecimiento de una escala graduada de niveles para los criterios considerados es fundamental, junto a la descripción detallada de lo que implica cada nivel de logro.

Una vez diseñada, la rúbrica se aplicará en el proceso de evaluación y se analizarán los resultados, prestando especial atención a dos aspectos que, como evidencian estudios ya clásicos en torno a este tema (Mokal y Leydens, 2000), resultan relevantes para determinar la calidad del instrumento: la fiabilidad y la validez de la rúbrica, aspectos a los que ya hemos hecho referencia y se tratarán de forma más detenida en los siguientes epígrafes.

## 2. Método

El objetivo de este estudio es analizar la fiabilidad y validez de una rúbrica diseñada para evaluar las competencias demostradas, a través de la presentación escrita del denominado “Trabajo Fin de Máster”, por los estudiantes de un Master universitario del ámbito de la educación. Los objetivos específicos, por lo tanto, se

podrían concretar en: analizar la fiabilidad de la rúbrica a través de la variabilidad en las puntuaciones inter-jueces y determinar su validez mediante dos estrategias: racional (mediante juicios de expertos) y empírica (con referencia a un criterio externo).

La construcción de un instrumento de medida de competencias desarrolladas en un Máster universitario, en este caso en forma de rúbrica, es un proceso complejo, que persigue obtener una herramienta útil al profesorado para determinar el nivel de competencias adquiridas por los estudiantes. Este estudio se presenta describiendo, en una primera fase, el proceso de diseño y elaboración de la rúbrica y, en una segunda, los resultados sobre validez y fiabilidad del instrumento. En la segunda fase, se han utilizado a su vez varias pruebas de evidencia: el juicio de expertos de carácter cualitativo, la validación empírica a partir de un grupo piloto de estudiantes con el objetivo de comprobar y adecuar el instrumento al perfil de los estudiantes y al contenido del TFM objeto de evaluación; y la validación a partir de un criterio externo, de carácter cuantitativo (validez criterial). En este caso, se trata de un diseño cuantitativo no experimental-correlacional que pretende evaluar la validez y fiabilidad de un instrumento de evaluación de las competencias exigidas en un TFM, dentro de un nivel 3 de MECES<sup>1</sup> (nivel de Master universitario) (Carrió, Soler y Aymerich, 2015; Mateo y Martínez, 2008; Muñiz y Fonseca-Pedrero, 2019).

### 2.1. Participantes y procedimiento de recogida de información

La construcción del instrumento estuvo avalada por la Comisión Académica del Máster y comenzaron a trabajar en su diseño 5 docentes del propio máster, profesores universitarios doctores de diferentes áreas de conocimiento y con más de 10 años de experiencia docente en la titulación. Así se elaboró un primer borrador o versión de la propuesta de rúbrica.

En un segundo momento, la elaboración de un instrumento de evaluación requiere contar con expertos que posibiliten validar su contenido. En este sentido, a partir de la primera versión de la rúbrica elaborada se procedió al análisis de la validez de contenido mediante el método de juicio de expertos (Urrutia, Barrios, Gutiérrez y Mayorga, 2014). Para ello se invitó a un total de 19 profesores, el criterio de inclusión adoptado es que fuesen expertos en la materia e implicados directamente con el Máster, con la intención de obtener sugerencias de mejora, tanto en las dimensiones como en los descriptores de la rúbrica. Los jueces realizaron una revisión cualitativa de las dimensiones y niveles de logro, atendiendo a cuestiones de redacción, organización del instrumento, comprensión, etc. Con estas sugerencias se procedió a la elaboración de una segunda versión de la rúbrica, más depurada. En tercer lugar, se contó con la aportación de una egresada del máster quien realizó sugerencias desde el punto de vista de una persona que participó y realizó el TFM en ediciones anteriores, lo que supuso un aporte enriquecedor que derivó en la tercera versión de la rúbrica. Esta se envió a los 25 estudiantes matriculados durante el curso 2016/17, que en esos momentos estaban desarrollando su trabajo y tenían criterio para valorar su

<sup>1</sup> Recuperado de <https://www.boe.es/buscar/pdf/2014/BOE-A-2014-12098-consolidado.pdf>

pertinencia y adecuación al objeto de evaluación. Se obtuvieron sugerencias de mejora de 12 alumnos/as. La totalidad de sugerencias sirvió para dar forma a la versión definitiva de la rúbrica, la cual fue aprobada en la Comisión Académica del máster.

El instrumento se aplicó a 16 TFM en el curso 2017-18 y a 18 TFM en el curso 2018-19, lo que suma un total de 34 TFM evaluados por 4 profesores cada uno. Los autores de los trabajos eran estudiantes del máster, con edades comprendidas entre 22 y 46 años, y de procedencia variada, incluyendo un porcentaje de extranjeros de un 37%. Los TFM se distribuyeron de forma equitativa entre los profesores del máster, de modo que casi todos los docentes asumieran la tutorización de algún TFM. Al finalizar el proceso, los tutores utilizaron la rúbrica para evaluarlo. Además, cada trabajo fue evaluado también por la Comisión de evaluación de TFM correspondiente, formada por 3 profesores del máster diferentes al director del trabajo. En resumen, los análisis se corresponden con las evaluaciones realizadas por los 34 tutores de los estudiantes en la asignatura Trabajo Fin de Máster, y por los miembros que configuraban las comisiones de evaluación, lo que suma un total de 136 valoraciones.

## 2.2. Análisis de fiabilidad y validez

La fiabilidad es un requisito previo para determinar la validez de un instrumento de evaluación. Por ello se estudiará la fiabilidad de la rúbrica en base a la concordancia o grado de acuerdo inter-jueces, entendida como la consistencia entre las puntuaciones otorgadas a cada trabajo.

La validez de un instrumento indica el grado de veracidad con el que mide el constructo teórico que pretende medir y si, por tanto, se puede utilizar con el fin previsto (medida de las competencias del máster). El trabajo conjunto de diferentes especialistas conocedores de las características del máster en el diseño de la rúbrica supone una primera aproximación para asegurar la validez de contenido del instrumento, desde una perspectiva holística y cualitativa.

Posteriormente, el juicio de expertos y los análisis realizados a partir de su aplicación están orientados a comprobar la fiabilidad y validez de la rúbrica. Para ello se analizará, tanto la validez de contenido, grado en que el instrumento presenta una muestra adecuada de los contenidos a los que se refiere, sin omisiones y sin desequilibrios de contenido, realizada en base al juicio de expertos; como la validez criterial, referida al grado en que los resultados producidos por la aplicación del instrumento correlacionan con los resultados obtenidos por otras pruebas que miden lo mismo que el primero, pero ajenas al test (criterios). Lo esperable es que exista entre el test y el criterio una correlación alta y positiva. En nuestro caso, se ha tomado como criterio la calificación global del estudiante en el conjunto de asignaturas del máster, exceptuando el TFM.

## 2.3. Diseño y presentación de la rúbrica para evaluar el TFM

En primera instancia, para la construcción de la rúbrica se tuvieron en cuenta las competencias que, de acuerdo a lo estipulado en la memoria del título, deben adquirir los estudiantes al realizar el Trabajo Fin de Máster (<https://www.usal.es/las->

tic-en-educacion-analisis-y-diseno-de-procesos-recursos-y-practicas-formativas-informacion).

Una vez seleccionadas las competencias se definieron los criterios de evaluación a tener en cuenta en la valoración de los trabajos mediante la rúbrica. Los criterios que se establecieron se corresponden con:

- Adecuación de los aspectos formales (estructura/presentación, referencias y normas de citación, redacción).
- Pertinencia del contenido y metodología empleada (objetivos, contenidos y metodología).
- Idoneidad y coherencia de los resultados, conclusiones y grado de adquisición de los conocimientos.
- Relevancia de las aportaciones para la mejora educativa y prospectiva.

Posteriormente se acordaron cuatro niveles de logro o ejecución y para cada nivel un intervalo de valoración: deficiente (0-8 puntos), aceptable (de 9-13 puntos), notable-bueno (de 14-17 puntos) y excelente (de 18-20 a puntos). Incorporando en cada caso la descripción cualitativa que contribuye a facilitar la valoración de cada uno de los elementos. En este sentido, los aspectos formales constituyen un 20% del total de la calificación, el contenido y metodología empleada un 30% los resultados, conclusiones y grado de adquisición de conocimientos un 30% y las aportaciones para la mejora educativa y prospectiva un 20%.

El proceso de diseño y validación de la rúbrica da lugar al instrumento que se presenta a continuación en las tablas 1 a 3, donde se puede observar la descripción de los criterios y niveles de consecución (indicadores).

Tabla 1. Descripción de los niveles de logro para la Dimensión 1: Aspectos formales

RÚBRICA PARA EVALUAR LOS TRABAJOS DE FIN DE MÁSTER (TFM)			
Niveles de logro en Dimensión 1: Aspectos formales			
DEFICIENTE	ACEPTABLE	NOTABLE	EXCELENTE
<p><b><u>Estructura/presentación</u></b></p> <p>La presentación del trabajo es muy deficiente (no hay un índice o está mal organizado, interlineado arbitrario, falta de justificación de márgenes, de encabezados, ausencia de paginación, incoherencia en el formato de epígrafes, presentación incorrecta</p>	<p><b><u>Estructura/presentación</u></b></p> <p>La presentación del trabajo se ajusta a la normativa exigida, pero es muy básica (no se aprecia un buen manejo y ejecución del procesador de textos, índices auto-paginados, saltos de página, uso de estilos del título, empleo apropiado de las sangrías, inclusión de tablas en el propio</p>	<p><b><u>Estructura/presentación</u></b></p> <p>La presentación está bastante cuidada y facilita la lectura. Se emplean con frecuencia elementos icónicos que ilustran y/o ayudan a clarificar los distintos contenidos.</p> <p><b><u>Referencias y normas de citación</u></b></p> <p>Emplea las normas de</p>	<p><b><u>Estructura/presentación</u></b></p> <p>La presentación del trabajo está muy cuidada (presenta un índice bien organizado, interlineado justificado, márgenes y encabezados uniformes, formato correcto de epígrafes, presentación correcta de figuras, tablas e imágenes).</p>

<p>de figuras, tablas, imágenes...).</p> <p><b>Referencias y normas de citación</b></p> <p>Emplea arbitrariamente las normas de citación; en unas ocasiones las referencias bibliográficas siguen las normas APA, y en otras no.</p> <p>No incluye una variedad de referencias bibliográficas ni están actualizadas.</p> <p>La bibliografía no recoge todas las referencias que se han citado en el texto.</p> <p><b>Redacción</b></p> <p>La expresión escrita de las ideas es deficiente (hay frecuentes errores gramaticales y no se emplean correctamente los signos de puntuación). Además, no se aprecia análisis o síntesis de la información, así como tampoco el empleo de un vocabulario específico.</p> <p>Hay reiteradas faltas de ortografía.</p> <p>Hay numerosos errores tipográficos que denotan una falta de revisión del trabajo.</p>	<p>formato del texto, etc.), y, en algunas ocasiones, incorrecta.</p> <p>No se utilizan, o en muy pocas ocasiones, elementos que ilustren y clarifiquen ciertas informaciones (gráficos, imágenes, tablas...)</p> <p><b>Referencias y normas de citación</b></p> <p>Bastantes referencias bibliográficas no están actualizadas, son muy limitadas, incluyéndose únicamente las de ámbito nacional. No hay referencias a revistas científicas de carácter internacional.</p> <p><b>Redacción</b></p> <p>Se presentan las ideas con una moderada claridad expositiva y se muestra una capacidad de análisis y síntesis suficiente.</p> <p>Se aprecia poco dominio de vocabulario específico sobre el tema.</p> <p>Se observan algunos errores tipográficos y ortográficos que indican una revisión parcial del trabajo.</p>	<p>citación APA, aunque existen algunas incorrecciones.</p> <p>La mayor parte de las referencias bibliográficas son actuales y variadas (nacionales e internacionales). Se incluyen bastantes referencias de artículos de revistas científicas.</p> <p><b>Redacción</b></p> <p>Se exponen las ideas con bastante precisión, claridad, mostrando análisis y síntesis de las mismas, según los casos. Se utiliza un vocabulario específico bastante preciso.</p> <p>No hay errores tipográficos ni ortográficos, o éstos son muy puntuales y poco relevantes.</p>	<p>Los elementos gráficos empleados se han elaborado personalmente, e ilustran y/o enriquecen los contenidos.</p> <p><b>Normas de citación</b></p> <p>Utiliza correctamente las normas de citación APA.</p> <p>Las referencias bibliográficas son actuales, relevantes, con un equilibrio adecuado entre las procedentes del ámbito nacional e internacional.</p> <p><b>Redacción</b></p> <p>Existe un alto nivel de precisión y claridad en la exposición de las ideas, con una elevada capacidad de análisis y síntesis de las mismas.</p> <p>El vocabulario empleado es específico, preciso y rico.</p> <p>No hay errores tipográficos ni ortográficos.</p>
<p>Puntuación: 0 a 8</p>	<p>Puntuación: 9-13</p>	<p>Puntuación: 14-17</p>	<p>Puntuación: 18-20</p>

Fuente: Elaboración propia.

Tabla 2. Descripción de los niveles de logro para la Dimensión 2: Contenido y metodología

Niveles de logro en Dimensión 2: Contenido y metodología			
DEFICIENTE	ACEPTABLE	NOTABLE	EXCELENTE
<p><b>Objetivos</b></p> <p>No se plantean objetivos de trabajo o éstos son muy confusos, no están bien definidos y son escasamente relevantes para el ámbito educativo.</p> <p><b>Contenidos</b></p>	<p><b>Objetivos</b></p> <p>Se plantean unos objetivos mínimos, pero algo confusos por estar poco definidos. No son muy relevantes en relación con el ámbito educativo.</p> <p><b>Contenidos</b></p> <p>La organización y</p>	<p><b>Objetivos</b></p> <p>Se definen correctamente los objetivos que van a guiar el desarrollo del trabajo, siendo éstos interesantes y de relevancia educativa.</p> <p><b>Contenidos</b></p> <p>La organización y</p>	<p><b>Objetivos</b></p> <p>Los objetivos formulados resultan interesantes, novedosos y tienen gran relevancia educativa.</p> <p><b>Contenidos</b></p> <p>La estructura y organización de los contenidos es muy</p>

<p>La organización y estructura de los contenidos es inadecuada por su arbitrariedad y falta de rigor científico. Los contenidos son muy superficiales y de escasa calidad, con explicaciones poco coherentes. La relevancia, originalidad y grado de innovación de los contenidos son nulos. Los contenidos no están bien fundamentados en base a referencias bibliográficas.</p> <p><b>Metodología</b></p> <p>La metodología no es adecuada ni pertinente para alcanzar los objetivos fijados, faltando el rigor mínimo exigible. Se cometen errores importantes relacionados con el diseño metodológico (expresado de modo confuso), los análisis realizados (insuficientes) y los resultados obtenidos (parciales y escuetamente expresados).</p>	<p>estructura de los contenidos es aceptable, si bien no se observan unos criterios claros de presentación de la información. Existe un exceso de citas directas. En ocasiones, realiza juicios de valor sin citar fuentes o evidencias en la que fundamentarlos. Los contenidos no profundizan suficientemente en la temática estudiada. Se ajustan escasamente a los objetivos formulados.</p> <p><b>Metodología</b></p> <p>La metodología planteada no es la más adecuada para la consecución de los objetivos. Se cometen algunos errores relacionados, bien con el diseño metodológico, los análisis realizados y/o los resultados obtenidos.</p>	<p>estructura de los contenidos es lógica, clara y pertinente. Los contenidos se abordan de modo riguroso y fundamentado. Plantean un alto grado de interés, originalidad e innovación. Las afirmaciones, juicios de valor y evidencias presentadas se fundamentan en datos empíricos y en el estudio del estado de la cuestión.</p> <p><b>Metodología</b></p> <p>La metodología planteada es adecuada para alcanzar los objetivos, realizándose un proceso de investigación riguroso o una propuesta de intervención interesante, detallada y bien fundamentada.</p>	<p>adecuada para comprender el tema. La presentación y desarrollo de los contenidos están muy bien fundamentados. Los contenidos son originales e innovadores, mostrando ideas y hechos relevantes.</p> <p><b>Metodología</b></p> <p>La metodología planteada es muy adecuada para conseguir los objetivos previstos (intervención novedosa y bien fundamentada; datos de investigación con rigor metodológico, diseño de algún recurso de interés educativo, etc.).</p>
<p><b>Puntuación: 0-9</b></p>	<p><b>Puntuación: 10-17</b></p>	<p><b>Puntuación: 18-25</b></p>	<p><b>Puntuación: 26-30</b></p>

Fuente: Elaboración propia.

Tabla 3. Descripción de los niveles de logro para la Dimensión 3: Resultados, conclusiones y grado de adquisición de conocimientos; y 4) Aportaciones para la mejora educativa y futuros estudios.

Niveles de logro en Dimensión 3: Resultados, conclusiones y grado de adquisición de conocimientos			
DEFICIENTE	ACEPTABLE	NOTABLE	EXCELENTE
<p><b>Resultados</b></p> <p>No se analizan ni interpretan los resultados obtenidos. No aportan información relevante. Los análisis efectuados y los resultados</p>	<p><b>Resultados</b></p> <p>El análisis e interpretación de los resultados es incompleto (ausencia de análisis de datos relevantes, de gráficos o tablas que clarifiquen la información,</p>	<p><b>Resultados</b></p> <p>Se presentan de manera suficiente los resultados, aunque se podría haber profundizado más a través de otros análisis complementarios.</p>	<p><b>Resultados</b></p> <p>Los resultados del trabajo se presentan de manera muy rigurosa y completa.</p> <p><b>Conclusiones</b></p> <p>La formulación de las conclusiones y</p>

<p>obtenidos son incorrectos.</p> <p><b>Conclusiones</b></p> <p>No hay conclusiones, o éstas no se relacionan con el trabajo, por lo que no son significativas.</p> <p>No enlaza las conclusiones con los objetivos del trabajo. Reitera datos ya comentados en otros apartados.</p> <p><b>Adquisición de conocimientos</b></p> <p>No se evidencia una suficiente adquisición de conocimientos sobre el tema tratado.</p>	<p>incorrecciones en la interpretación).</p> <p><b>Conclusiones</b></p> <p>Las conclusiones que se extraen son muy generales, no aprovechándose otras posibles en función de la fundamentación teórica y de los resultados obtenidos.</p> <p><b>Adquisición de conocimientos</b></p> <p>Se evidencia, a partir de alguna conclusión, la adquisición de ciertos conocimientos sobre el tema tratado.</p>	<p><b>Conclusiones</b></p> <p>Se aprecia una correcta presentación y defensa de las principales conclusiones, pero sin la suficiente creatividad y madurez en su desarrollo.</p> <p><b>Adquisición de conocimientos</b></p> <p>Las conclusiones evidencian que se ha profundizado en algunos aspectos y se han realizado aportaciones sobre el tema tratado.</p>	<p>planteamientos presentados refleja madurez y creatividad, relacionando los resultados con el marco teórico.</p> <p><b>Adquisición de conocimientos</b></p> <p>Las conclusiones y reflexiones que se realizan evidencian un dominio importante del tema abordado y la aportación de datos relevantes.</p>
<b>Puntuación: 0-9</b>	<b>Puntuación: 10-17</b>	<b>Puntuación: 18-25</b>	<b>Puntuación: 26-30</b>
<b>Niveles de logro en Dimensión 4: Aportaciones para la mejora educativa y futuros estudios</b>			
<p><b>Aportaciones</b></p> <p>El trabajo no efectúa ninguna aportación a la temática que desarrolla a través de la fundamentación teórica, los resultados obtenidos y las conclusiones.</p> <p>No es posible su transferencia al campo educativo, en ninguna de sus vertientes.</p> <p><b>Prospectiva</b></p> <p>No hay explícita ninguna prospectiva que invite a seguir investigando sobre la temática.</p>	<p><b>Aportaciones</b></p> <p>Las aportaciones del trabajo son poco significativas. Básicamente son reiteraciones de ideas presentadas en otros estudios.</p> <p>Es posible su transferencia a algún ámbito del campo educativo, pero con muchas limitaciones.</p> <p><b>Prospectiva</b></p> <p>Se presentan algunas aportaciones para futuras investigaciones pero de forma muy general.</p>	<p><b>Aportaciones</b></p> <p>Presenta importantes aportaciones personales a la temática trabajada, por lo que su transferencia es posible, con pequeñas adaptaciones.</p> <p><b>Prospectiva</b></p> <p>Se evidencia una clara prospectiva que da continuidad al trabajo, mediante el planteamiento explícito de posibles estudios futuros.</p> <p>El trabajo puede ser publicable en alguna revista de difusión del entorno educativo.</p>	<p><b>Aportaciones</b></p> <p>Presenta aportaciones de gran relevancia educativa.</p> <p>Las aportaciones permiten su transferencia al contexto educativo.</p> <p><b>Prospectiva</b></p> <p>Se presenta una prospectiva muy clara y amplia para diseñar y desarrollar futuras investigaciones sobre el tema.</p> <p>El trabajo puede ser publicable en una revista científica de prestigio en el ámbito educativo.</p>
<b>Puntuación: 0-8</b>	<b>Puntuación: 9-13</b>	<b>Puntuación: 14-17</b>	<b>Puntuación: 18-20</b>

Fuente: Elaboración propia.

### 3. Resultados

Se analizan, en primer lugar, las puntuaciones finales asignadas por los evaluadores a los TFM para comprobar la proximidad de las mismas en la calificación del trabajo. Los profesores, tanto el tutor, como los miembros de la comisión, evaluaron los trabajos, desde un punto de vista analítico y holístico. Lo que nos permite comparar, en el primer caso, el nivel de acuerdo en las dimensiones entre el tutor y

los profesores de la comisión, y entre todos éstos; y en el segundo caso, el nivel de acuerdo entre las valoraciones finales (puntuaciones globales) otorgadas en cada caso. Para ello se utiliza el coeficiente de correlación de Pearson entre las valoraciones finales y por dimensiones.

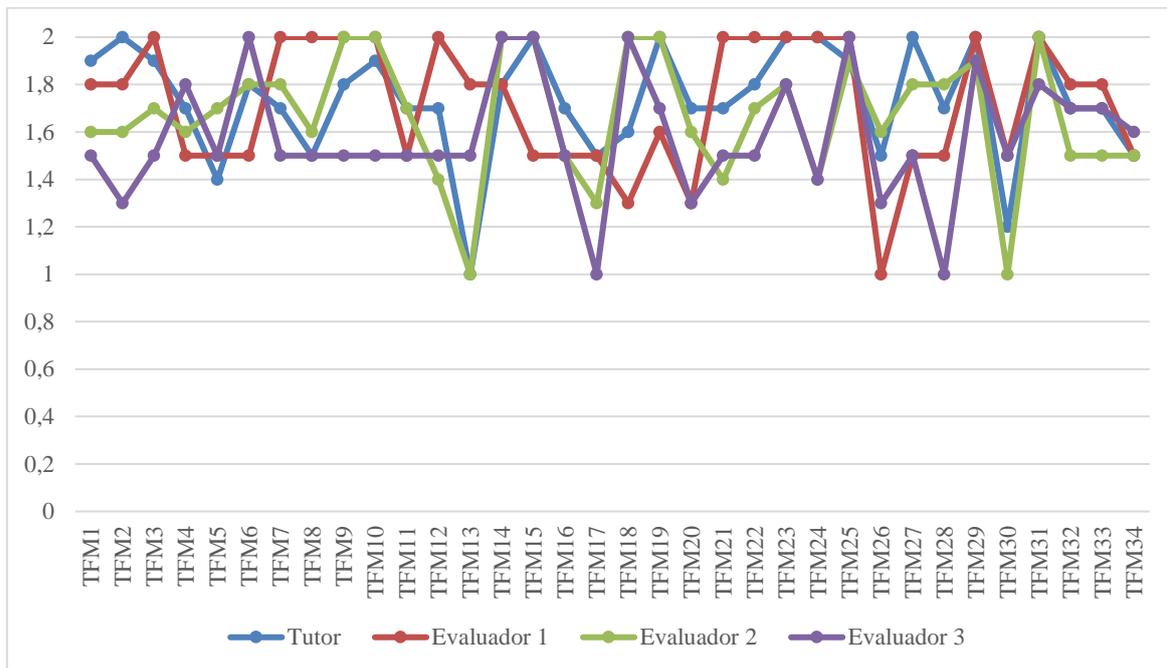
En segundo lugar, se estudia la fiabilidad de la rúbrica en base a la concordancia o grado de acuerdo inter-jueces, entendida como la consistencia entre las puntuaciones otorgadas por el docente-tutor y los miembros de la comisión de evaluación al trabajo escrito presentado por los estudiantes. El procedimiento estadístico para determinar la concordancia entre distintos evaluadores es la correlación de medidas de los distintos jueces sobre un mismo objeto evaluado. Así pues, se muestran los resultados de los análisis de concordancia (variabilidad inter-jueces) entre los evaluadores a través del coeficiente Kappa, considerando, por un lado, los distintos criterios o dimensiones (aspectos formales, contenido-metodología, resultados-conclusiones y aportaciones para la mejora educativa-futuros estudios), y por otro, las categorías de evaluación (deficiente, aceptable, notable y excelente).

En tercer lugar, con objeto de estudiar la validez criterial (con referencia a criterio), se ha calculado la correlación entre la nota derivada de la rúbrica aplicada, es decir, la puntuación final del trabajo escrito, considerando la puntuación media de los 4 evaluadores, con la calificación global del estudiante en el conjunto de asignaturas del Máster, excluyendo la calificación del trabajo escrito del TFM.

### 3.1. Descriptivos de las puntuaciones finales del TFM - trabajo escrito.

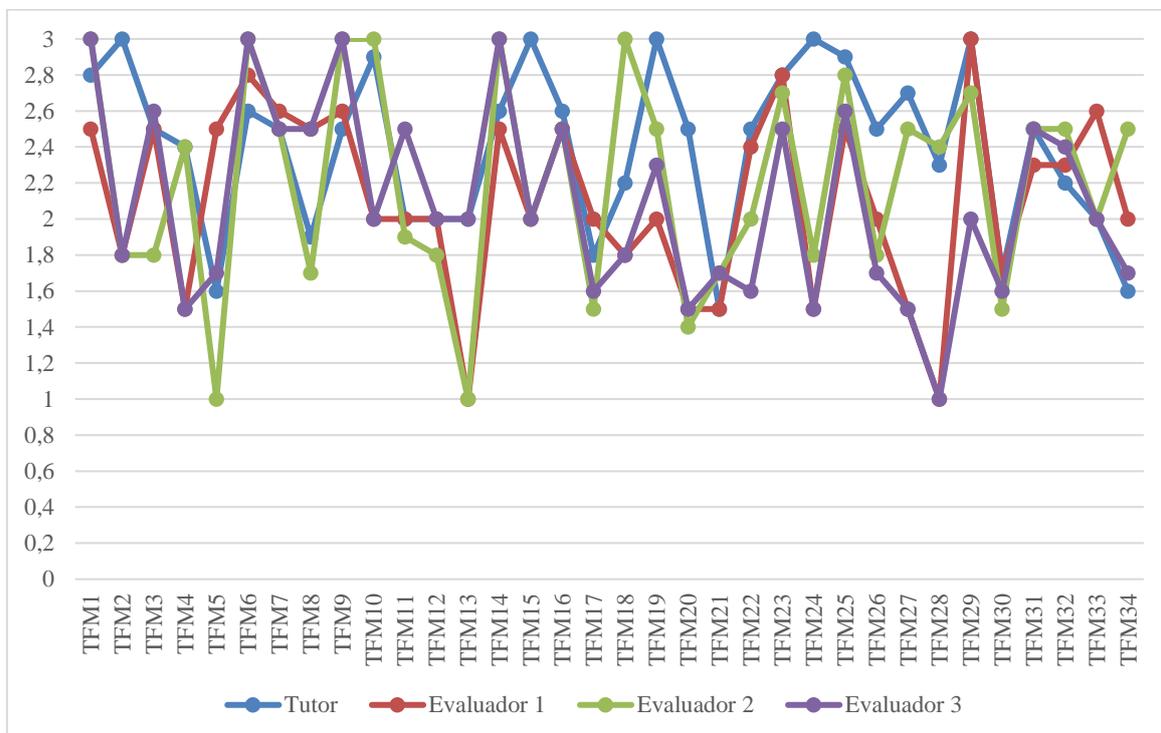
Las puntuaciones asignadas por los cuatro evaluadores en cada uno de los 34 TFM se presentan gráficamente en las figuras 1 a 5, pudiéndose observar que predominan las puntuaciones medio-altas en las escalas utilizadas. Las gráficas reflejan la variabilidad inter-jueces de las puntuaciones en la rúbrica, tanto en lo que respecta a las 4 dimensiones de evaluación (D1: Aspectos formales; D2: Contenido y metodología empleada; D3: Resultados, conclusiones y grado de adquisición de conocimientos; D4: Aportaciones para la mejora educativa y futuros estudios) como a la puntuación final obtenida en el trabajo escrito (figura 5). Como se puede observar, en muchos casos los 4 evaluadores otorgan puntuaciones bastante similares o cercanas, tanto en las 4 dimensiones como en la puntuación final del trabajo escrito, pero sobre todo en esta última. En el siguiente apartado profundizaremos en estos datos a nivel estadístico.

Figura 1. Dimensión 1 - Aspectos formales.



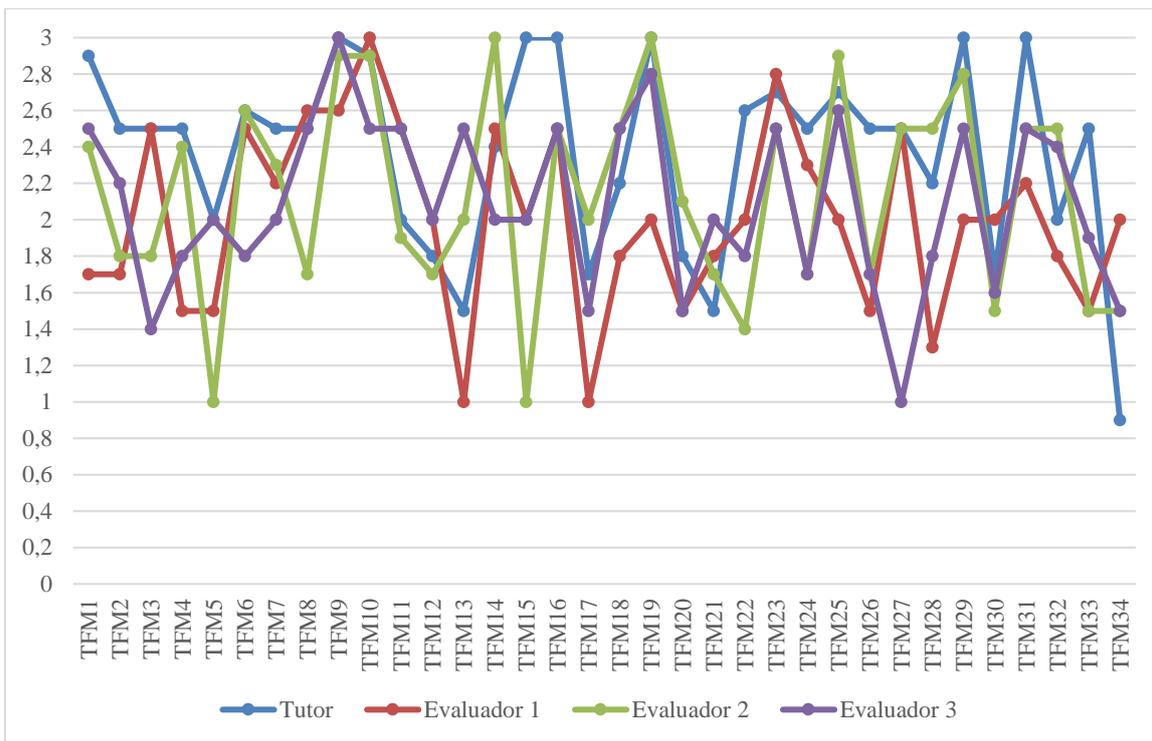
Fuente: Elaboración propia.

Figura 2. Dimensión 2 - Contenido y metodología empleada.



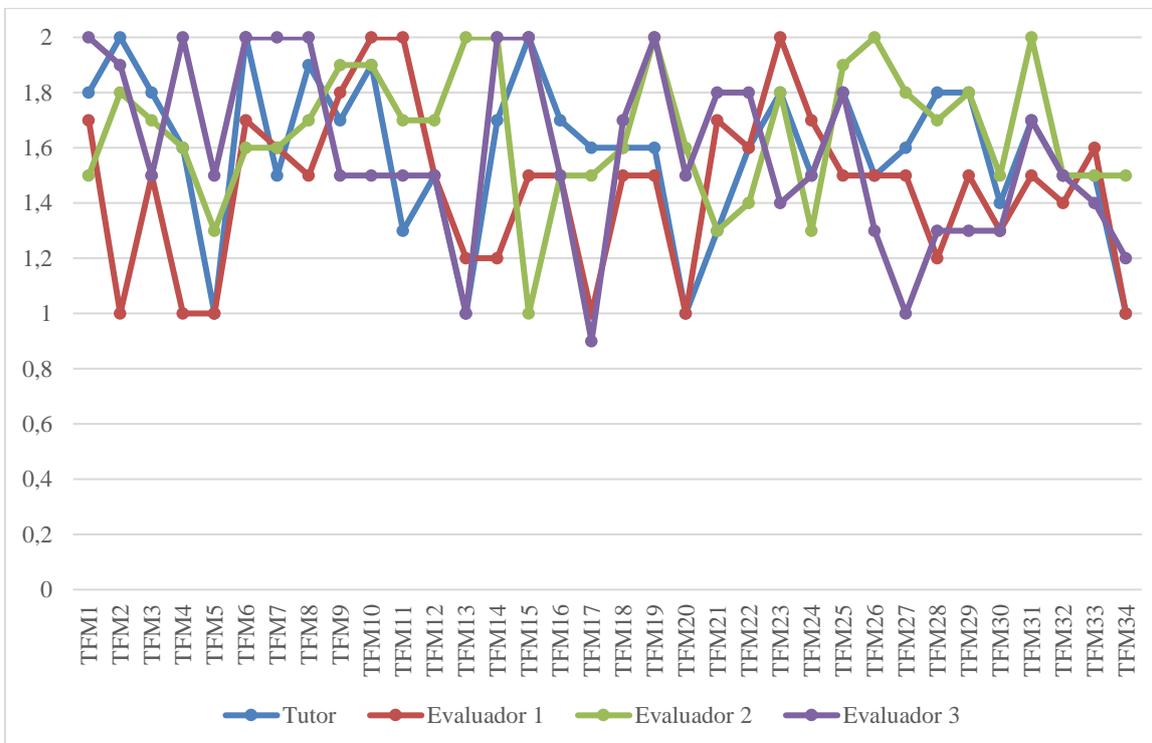
Fuente: Elaboración propia.

Figura 3. Dimensión 3 - Resultados, conclusiones y grado de adquisición de conocimientos.



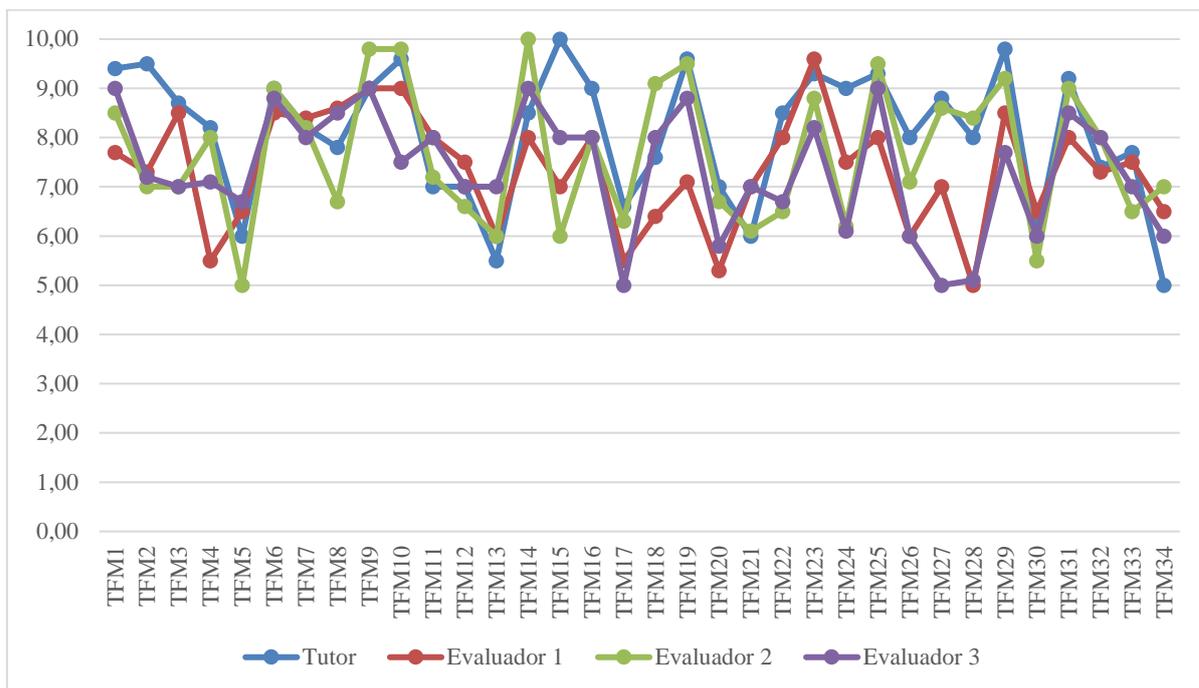
Fuente: Elaboración propia.

Figura 4. Dimensión 4 - Aportaciones para la mejora educativa y futuros estudios.



Fuente: Elaboración propia.

Figura 5. Puntuación final del trabajo escrito.



Fuente: Elaboración propia.

### 3.2. Análisis de la fiabilidad en base a la concordancia entre los evaluadores

La variabilidad entre observadores o evaluadores puede llevar a errores de medida, por lo que estimar el grado de dicha variabilidad es importante para estudiar la fiabilidad de un instrumento. Por lo tanto, en el estudio de la fiabilidad se debería analizar, por una parte, el sesgo entre observadores -por ejemplo, la tendencia de un evaluador a dar puntuaciones mayores que otro- y, por otra, la concordancia entre observadores -coincidencia de los observadores en su medición-. Considerando este segundo aspecto, habrá que tener en cuenta la naturaleza de los datos: si éstos son de tipo continuo es habitual la utilización de estimadores del coeficiente de correlación intraclass, mientras que cuando se trata de datos de tipo categórico el estadístico más empleado es el índice Kappa (Villasís-Keever, Márquez-González, Zurita-Cruz, Miranda-Novales, Escamilla-Núñez, 2018).

#### 3.2.1. Análisis de la concordancia entre evaluadores en las puntuaciones cuantitativas.

Procedemos a valorar la concordancia entre las puntuaciones cuantitativas otorgadas por el tutor y los tres profesores del tribunal en cada una de las dimensiones y en la puntuación final otorgada al trabajo escrito. Para ello, se ha utilizado el coeficiente de correlación intraclass, que permite valorar la concordancia entre medidas continuas realizadas sobre una serie de sujetos. De forma particular, se ha utilizado el coeficiente de correlación intraclass siguiendo el modelo aleatorio bidireccional y el tipo de acuerdo absoluto. Los resultados se muestran en la tabla 4.

Siguiendo lo señalado por Cicchetti (1994), podemos indicar que el coeficiente de correlación intraclase de la puntuación final del trabajo escrito, es decir, la concordancia entre las puntuaciones de los evaluadores, se encuentra a un nivel “excelente” (CCI=0,805; IC 95% 0,668-0,893), seguido de la Dimensión 2 (CCI=0,737; IC 95% 0,559-0,855), la Dimensión 3 (CCI=0,691; IC 95% 0,485-0,830) y la Dimensión 1 (CCI=0,647; CI 95% 0,413-0,805), que se encuentran a un nivel “bueno”. Mientras que el resultado más bajo se encuentra en la Dimensión 4 (CCI=0,458; IC 0,107-0,699), lo que implica un nivel de concordancia entre los evaluadores “suficiente”. Esto nos hace pensar que algunas de las dimensiones están mejor descritas y precisadas que otras, necesitando la D4 una nueva revisión, ya sea de la dimensión general o de sus descriptores específicos.

Tabla 4

*Concordancia entre los evaluadores en las puntuaciones cuantitativas.*

	Coeficiente de correlación intraclase	Prueba F con valor verdadero 0			
		Valor	gl1	gl2	p
D1	0,647 (0,413-0,805)	2,975	33	99	0,000
D2	0,737 (0,559-0,855)	4,081	33	99	0,000
D3	0,691 (0,485-0,830)	3,505	33	99	0,000
D4	0,458 (0,107-0,699)	1,901	33	99	0,008
Puntuación final	0,805 (0,668-0,893)	5,674	33	99	0,000

Fuente: Elaboración propia.

### 3.2.2. Análisis de la concordancia entre los evaluadores en las categorías de evaluación cualitativas.

Como se ha señalado previamente, la rúbrica agrupaba las puntuaciones cuantitativas en categorías cualitativas (Categoría 1= Deficiente; 2= Aceptable; 3= Notable; 4= Excelente). En este sentido, para responder al objetivo señalado referido a la validación de la rúbrica, en particular, a las categorías creadas, queremos conocer la concordancia entre evaluadores en estas categorías. Para ello, se ha utilizado la prueba Kappa de Fleiss. En las tablas 5 a 8, se presentan los resultados para las 4 dimensiones.

En todos los casos, los valores están por debajo de 0,40 lo que representa un nivel de concordancia bajo, siguiendo lo señalado por Fleiss, Levin y Paik (2003). Además, se observa que al realizar el análisis de la concordancia sólo entre los 3 evaluadores del tribunal, eliminando la valoración del tutor, los resultados son muy similares.

Esto nos hace pensar en la necesidad de reestructurar las categorías, ya sea modificando el rango de puntos que está incluido en cada categoría (ya que se ha visto que a nivel cuantitativo la concordancia es, en general, aceptable) o modificando la redacción de cada categoría en cada dimensión, de modo que sea más claro para los evaluadores. En relación con esto, podemos señalar que alguna de las categorías de

alguna de las dimensiones obtiene mejores resultados, como lo es la categoría 2 (Aceptable) de la D2 ( $Kappa=0,401$ ), lo que estaría en un nivel “suficiente”.

Tabla 5

Concordancia entre los evaluadores (tutor y 3 profesores del tribunal) en las categorías de puntuación de la dimensión D1.

	Kappa	Límite del Intervalo de Confianza asintótico por debajo del 95%	Límite del intervalo de confianza asintótico por encima del 95%	Z	p
D1 - General	0,206	0,095	0,317	3,646	0,000
Categoría 2	0,178	0,041	0,315	2,540	0,011
Categoría 3	0,145	0,008	0,283	2,077	0,038
Categoría 4	0,279	0,141	0,416	3,978	0,000

Fuente: Elaboración propia.

Tabla 6

Concordancia entre los evaluadores (tutor y 3 profesores del tribunal) en las categorías de puntuación de la dimensión D2.

	Kappa	Límite del Intervalo de Confianza asintótico por debajo del 95%	Límite del intervalo de confianza asintótico por encima del 95%	Z	p
D2 - General	0,272	0,173	0,372	5,368	0,000
Categoría 2	0,401	0,264	0,538	5,729	0,000
Categoría 3	0,145	0,008	0,283	2,077	0,038
Categoría 4	0,320	0,183	0,457	4,574	0,000

Fuente: Elaboración propia.

Tabla 7

Concordancia entre los evaluadores (tutor y 3 profesores del tribunal) en las categorías de puntuación de la dimensión D3.

	Kappa	Límite del Intervalo de Confianza asintótico por debajo del 95%	Límite del intervalo de confianza asintótico por encima del 95%	Z	p
D3 - General	0,123	0,024	0,222	2,428	0,015
Categoría 1	-0,007	-0,145	0,130	-0,106	0,916
Categoría 2	0,106	-0,031	0,244	1,520	0,129
Categoría 3	0,006	-0,131	0,143	0,084	0,933
Categoría 4	0,334	0,197	0,471	4,774	0,000

Fuente: Elaboración propia.

Tabla 8

Concordancia entre los evaluadores (tutor y 3 profesores del tribunal) en las categorías de puntuación de la dimensión D4.

	Kappa	Límite del Intervalo de Confianza asintótico por debajo del 95%	Límite del intervalo de confianza asintótico por encima del 95%	Z	p
D4 - General	0,104	0,004	0,203	2,035	0,042
Categoría 2	0,160	0,023	0,298	2,292	0,022
Categoría 3	0,088	-0,049	0,225	1,257	0,209
Categoría 4	0,077	-0,060	0,215	1,104	0,270

Fuente: Elaboración propia.

### 3.3. Análisis de la validez con referencia a un criterio externo

A la hora de analizar la validez con referencia a un criterio externo, se ha estimado oportuno elegir como variable criterio, la calificación global del estudiante en el Máster, considerando la nota media del conjunto de todas las asignaturas cursadas, exceptuando la nota del TFM. Subyace la creencia de que, si un estudiante ha obtenido una nota media global en el máster alta o muy alta, sería previsible que también la obtuviera en el trabajo presentado. Por tanto, hemos procedido a realizar la correlación entre la nota derivada de la rúbrica aplicada por los 4 evaluadores y la variable criterio.

En primer lugar, se presentan los estadísticos descriptivos para ambas variables en la tabla 9, donde se puede ver que la puntuación final del trabajo escrito en el TFM es algo inferior (7,62) a la nota media del alumnado en el Máster sin considerar el TFM (8,59).

Tabla 9

Estadísticos descriptivos sobre nota del TFM-trabajo escrito y nota media en el Máster sin considerar el TFM.

	N	Mínimo	Máximo	Media	Desviación estándar
Puntuación final en la rúbrica del TFM	34	5,85	9,20	7,6169	1,03492
Nota media global en el Máster excluyendo el TFM	34	7,21	9,40	8,5885	,52737

Fuente: Elaboración propia.

Para hallar la correlación entre la puntuación final en la rúbrica del TFM y la variable criterio, se ha utilizado el coeficiente Pearson entre ambas variables, obteniendo un resultado de  $r = 0,581$ , con una significación bilateral de  $p = 0,000$ . De este modo, podemos decir que existe una correlación positiva moderada, lo que permite afirmar que la rúbrica en su conjunto tiene un coeficiente de validez externa de 0,58, el cual puede considerarse moderadamente positivo, para el número de casos del estudio.

#### 4. Discusión

En cuanto a las características del instrumento de evaluación del TFM diseñado, se ha puesto de manifiesto que la rúbrica permite cumplir el requisito fundamental de la evaluación de competencias, como es la pertinencia del criterio de desempeño, es decir, que sirve para inferir el estado de competencia de una persona.

Los resultados obtenidos, en los que se ha verificado la “excelente” concordancia entre las puntuaciones de los evaluadores en el total de la rúbrica son altamente positivos para garantizar la fiabilidad del instrumento de evaluación diseñado, así como la correlación positiva con un criterio externo permite atribuir suficiente validez externa y calidad a la rúbrica. Desde este punto de vista se valora positivamente el trabajo realizado y se ofrece este instrumento de evaluación a la comunidad universitaria, destacando su utilidad para otros contextos educativos similares.

En cuanto al análisis de las dimensiones del instrumento, si bien es cierto que obtiene buenos resultados, la dimensión 4, relacionada con las aportaciones para la mejora educativa y futuros trabajos, resulta la más confusa o desafiante para los evaluadores, que muestran menor acuerdo al evaluar esta cuestión. Se puede argumentar que resulta difícil, incluso para investigadores acreditados, valorar de forma unánime si las aportaciones de un trabajo académico resultan relevantes, son transferibles y tienen utilidad para la mejora educativa, dado el carácter genérico de esta cuestión y las diferentes percepciones que pueden tener los evaluadores a la hora de juzgar esta dimensión. En cualquier caso, no por ello deja de ser un criterio de relevancia al juzgar un TFM, al igual que es aplicado en el ámbito profesional cuando se evalúa, por ejemplo, la calidad de un artículo científico para su posible publicación en una revista de prestigio (Aguaded, 2020). En cualquier caso, habría que revisar los indicadores descritos en las categorías de cada dimensión para tratar de precisarlos mejor, en aras de conseguir un mayor grado de consenso. En esta línea, los resultados sobre el pobre nivel de concordancia en las categorías de evaluación, con la mayoría de los valores inferiores a 0,40, reclama reflexión sobre la estructuración y descripción de las categorías, con rangos quizá demasiado amplios y una redacción que podría simplificarse para adquirir mayor precisión.

En conjunto se puede decir que los datos obtenidos muestran índices de valoración coherentes pero no tan objetivos como las expectativas puestas en esta prueba de evaluación, los profesores no coinciden un 100% en sus valoraciones de las evidencias presentadas, a pesar de la explicitación de los criterios de evaluación y sus niveles, lo que se podría explicar en gran medida porque aplican diferentes modelos de referencia surgidos de sus ideas personales de lo que es un buen trabajo (modelo ideal), su experiencia en situaciones similares previas, etc. Esto lleva a sugerir que además de una rúbrica, podría complementarse su aplicación con la presentación de modelos-tipo que representen diferentes niveles de consecución competencial, aunque esta idea puede ser muy difícil de aplicar en la práctica, según la variedad de casos, posibles temáticas a abordar, metodologías de trabajo, etc.

Por otra parte, hay que considerar que la realización del TFM está más centrado en habilidades académicas de redacción de un trabajo científico que en evidencias que determinen el grado alcanzado en las competencias en comparación con un estándar establecido, lo que sería básico para la evaluación de la competencia profesional y más claro para la definición de los indicadores. Además, en la compleja tarea de desarrollar un trabajo de investigación y realizar el informe escrito del mismo, lo que supone presentar el TFM, se requieren tanto competencias técnicas como sociales, estando las primeras relacionadas con las tareas y recursos de un puesto de trabajo y las segundas con el contexto social, aspectos personales y comunicativos (Del Pozo, 2012). Ambas entretejen el objeto de evaluación que supone el TFM y añaden complejidad a la tarea evaluativa. Claro está que cuando se evalúa un TFM no se realiza una observación directa sobre la ejecución de una competencia sino que se lleva a cabo una interpretación y valoración de la misma, y esto hace que los testimonios de los evaluadores no sean tan objetivos y veraces como si se tratara de una observación directa de una actividad.

A pesar de esta dificultad, no se debe perder de vista que la evaluación de las competencias debe basarse en evidencias observables que permitan inferir que se ha logrado el desempeño esperado (del Pozo, 2012) y en principios como la eficacia (agilidad en su aplicación, tiempo razonable para obtener evidencias, rentabilidad económica...), flexibilidad (aplicación de diferentes métodos para obtener las evidencias) y validez (capaz de obtener evidencias suficientes y fiables).

El valor de la rúbrica, además de ofrecer unas pautas de evaluación a los docentes para conseguir un mayor grado de objetividad y equidad, se dirigen al aprendiz, quien utiliza ésta como un andamiaje que le proporciona una guía en el logro de la competencia, posibilitando un auto-monitoreo del logro (Vázquez-Cano, Martín-Monje y Fernández-Álvarez, 2014) y la autocorrección en los aprendizajes que se van consiguiendo y en su forma de mostrarlos, por ejemplo, en el TFM. Esto nos lleva a afianzar su aplicación en el proceso de enseñanza-aprendizaje, viendo mejores resultados en las calificaciones del alumnado implicado, así como en la dinámica de evaluación del profesorado, coincidiendo con los resultados de trabajos similares (Velasco et al., 2018).

En cuanto a las limitaciones del estudio, se señalan como tales que la nota del TFM incluye una parte de exposición que no aparece en la rúbrica, por lo tanto, no se ha hecho mención a la evaluación completa del TFM. En segundo lugar, en cuanto al procedimiento de validez de contenido basado en expertos, se apunta la falta de revisión más sistemática sobre pertinencia, adecuación, redacción, etc., de manera cuantitativa, lo que hubiese permitido calcular algún coeficiente de validez de contenido más preciso.

Asumiendo las limitaciones, el estudio realizado se enmarca en una perspectiva de la educación superior que plantea la necesidad de centrar la enseñanza en los propios estudiantes y en su capacidad para autorregular sus aprendizajes, tanto de forma personal como en colaboración con otros (Raposo y Sarceda, 2010), siendo necesario disponer de entornos de formación flexibles, pero suficientemente guiados

con criterios de evaluación precisos y servicios de acompañamiento y tutorización por parte del docente. En relación al uso de la rúbrica como recurso de evaluación en la educación superior se aporta un instrumento válido para la finalidad de evaluar un TFM, coincidiendo con Cano (2015) cuando asocia la rúbrica con una herramienta de evaluación acorde a una visión de competencias, señalando su valor formativo versus sumativo, su capacidad de guiar de forma efectiva un proceso, así como el valor agregado de construirla versus consumirla. Si bien se apunta la necesidad de seguir investigando sobre el tema y validando los instrumentos que se emplean para valorar los trabajos de los estudiantes, máxime cuando éstos tienen una importante repercusión en las calificaciones de un título universitario.

### Referencias bibliográficas

- Aguaded, I. (2020). *Los procesos de selección académica del manuscrito por una revista científica*. Recuperado de <https://comunicarautores.com/2020/02/17/los-procesos-de-seleccion-academica-del-manuscrito-por-una-revista-cientifica/>
- Bailey, J.R. (2014). Crossing the Rubric-Con. *Journal of Management Education*, 38(3), 313-318. doi: <https://doi.org/10.1177/1052562913515657>
- Blanco, A. (2008). Las rúbricas: un instrumento útil para la evaluación de competencias. En L. Prieto (Coord.) *La enseñanza universitaria centrada en el aprendizaje*, (pp. 171-188). Octaedro.
- Cano, E. (2015). Las rúbricas como instrumento de evaluación de competencias en educación superior: ¿Uso o abuso? *Profesorado. Revista de Currículum y Formación del Profesorado*, 19(2), 265-280. Recuperado de <https://recyt.fecyt.es/index.php/profesorado/article/view/40925/23289>
- Carrió, C., Soler, M. y Aymerich, M. (2015). Análisis de la Validez de Contenido de un Cuestionario de Evaluación del Aprendizaje Basado en Problemas: Un Enfoque Cualitativo. *Formación universitaria*, 8(1), 13-22. doi: <https://doi.org/10.4067/S0718-50062015000100003>
- Cebrián, M. (2007). Buenas prácticas en el uso de e-portafolios y e-rúbrica. En A. Cid, M. Raposo, M. y A. Pérez (Coords.), *El prácticum: buenas prácticas en el Espacio Europeo de Educación Superior*, (pp. 67-87). Tórculo Edicións.
- Cebrián, M. (2014). Evaluación formativa con e-rúbrica: aproximación al estado del arte. *Revista de Docencia Universitaria*, 12(1), 15-2. Recuperado de [http://red-u.net/redu/documentos/vol12\\_n1\\_completo.pdf](http://red-u.net/redu/documentos/vol12_n1_completo.pdf)
- Cebrián, M. & Monedero, J.J. (2014). Evolución en el diseño y funcionalidad de las rúbricas: desde las rúbricas “cuadradas” a las erúbricas federadas. *Revista de Docencia Universitaria*, 12(1), 81-98. Recuperado de

<https://riunet.upv.es/bitstream/handle/10251/140001/Cebri%C3%A1n?sequence=1>

- Cicchetti, D. V. (1994). Guidelines, Criteria, and Rules of Thumb for Evaluating Normed and Standardized Assessment Instruments in Psychology. *Psychological Assessment*, 6(4), 284-290. doi: <https://doi.org/10.1037/1040-3590.6.4.284>
- Chica, E. (2011). Una propuesta de evaluación para el trabajo en grupo mediante rúbrica. *Escuela Abierta*, 14, 67-81. Recuperado de [http://www.ceuandalucia.es/escuelaabierta/pdf/articulos\\_ea14pdf/ea14\\_chica.pdf](http://www.ceuandalucia.es/escuelaabierta/pdf/articulos_ea14pdf/ea14_chica.pdf)
- Dawson, P. (2017). Assessment rubrics: towards clearer and more replicable design, research and practice. *Assessment & Evaluation in Higher Education*, 42(3), 347-360. doi: <https://doi.org/10.1080/02602938.2015.1111294>
- Del Pozo, J. Á. (2012). *Competencias profesionales: Herramientas de evaluación: el portafolios, la rúbrica y las pruebas situacionales* (Vol. 9). Narcea Ediciones.
- De la Mano, M. & Moro Cabero, M. (2009). La evaluación por competencias: propuesta de un sistema de medida para el grado en Información y Documentación. *Bid: Textos Univervisitaris de Biblioteconomía i Documentació*, 23, 1-20. Recuperado de <http://eprints.rclis.org/14993/>
- Díaz-Barriga, F. & De la Cruz, G. (2011). Rúbricas en la evaluación de competencias y aprendizaje complejo. Alcances y restricciones en educación superior. En K. Bujan, I. Rekelde, y P. Aramendi, (Coords.), *La evaluación de competencias en la Educación Superior. Las rúbricas como instrumento de evaluación* (pp. 9-12). MAD.
- Expósito, M., Nicolau, D. & Tomas, J.V. (2017). La evaluación por competencias en los grados. Desarrollo de una experiencia en el Grado en Administración y Dirección de Empresas mediante el uso y validación de rúbrica. *Revista Complutense de Educación*. 28(4), 1155-1171. doi: <https://doi.org/10.5209/RCED.51618>
- Fleiss, J. L., Levin, B. & Paik, M. C. (2003). *Statistical Methods for Rates and Proportions*. John Wiley & Sons.
- Fraile, J., Pardo, R. & Panadero, E. (2017). ¿Cómo emplear las rúbricas para implementar una verdadera evaluación formativa? *Revista Complutense de Educación* 28(4), 1321-1334. doi: <https://doi.org/10.5209/RCED.51915>
- Gairín, I. (2009). *Guía para la evaluación de competencias en el Area de Ciencias Sociales*. Agència per a la Qualitat del Sistema Universitari de Catalunya.
- García, M., Belmonte, B. & Nicolásac, G. (2017). Opinión del alumnado sobre el empleo de rúbricas en la universidad. *Estudios Pedagógicos*, 2, 93-113. Recuperado de <http://hdl.handle.net/10230/21151>

- García, M. J., Terrón, M<sup>a</sup> J. & Blanco, Y. (2009). *Desarrollo de recursos docentes para la evaluación de competencias genéricas*. XV Jornadas de Enseñanza Universitaria de la Informática (JENUI). UPC. Recuperado de <http://hdl.handle.net/2099/7902>
- García Sanz, M.P. (2014). La evaluación de competencias en Educación Superior mediante rúbricas: un caso práctico. *Revista electrónica interuniversitaria de formación del profesorado*, 17(1), 87-106. doi: <https://doi.org/10.6018/reifop.17.1.198861>
- Hernández Martín, A. (2006). *MEMORIA DEL PROYECTO DE INNOVACIÓN. Elaboración y validación de una rúbrica para la evaluación de los Trabajos de Fin de Máster (TFM) pertenecientes al Máster "Las TIC en educación. Análisis y diseño de procesos, recursos y prácticas formativas*. Recuperado de [https://gredos.usal.es/bitstream/handle/10366/135453/MID\\_16\\_113.pdf?sequence=1&isAllowed=y](https://gredos.usal.es/bitstream/handle/10366/135453/MID_16_113.pdf?sequence=1&isAllowed=y)
- Kohn, A. (2006). The trouble with rubrics. *The English Journal*, 95(4), 12-15. Recuperado de <https://www.alfiekohn.org/teaching/rubrics.pdf>
- Madariaga, I. & Lekue, P. (2019) Evaluación mediante una rúbrica de las guías docentes de másteres universitarios de arte de la Universidad del País Vasco (UPV/EHU). La valoración de expertos internos y externos. *Observar*, 13, 26-38. Recuperado de <https://www.observar.eu/ojs/index.php/Observar/article/view/101>
- Mateo, J. & Martínez, F. (2008). *Mediación y evaluación educativa*. La Muralla.
- Moskal, B. & Leydens, J.A. (2000). Scoring Rubric Development: Validity and Reliability, *Practical Assessment, Research, and Evaluation*, 7(10). doi: <https://doi.org/10.7275/q7rm-gg74>
- Muñiz, J., & Fonseca-Pedrero, E. (2019). Diez pasos para la construcción de un test, *Psicothema*, 31(1), 7-16 doi: <https://doi.org/10.7334/psicothema2018.291>
- Panadero, E., & Romero, M. (2014). To rubric or not to rubric? The effects of self-assessment on self-regulation, performance and self-efficacy. *Assessment in Education: Principles, Policy & Practice*, 21(2), 133-148. doi: <https://doi.org/10.1016/j.stueduc.2013.04.001>
- Popham, W. (1997). *What's wrong - and what's right - with rubrics?* Recuperado de <http://www.ascd.org/publications/educational-leadership/oct97/vol55/num02/What%27s-Wrong%E2%80%94and-What%27s-Right%E2%80%94with-Rubrics.aspx>
- Raposo, M. & Sarceda, M.C. (2010). El trabajo en las aulas con perspectiva europea: medios y recursos para el aprendizaje autónomo. *Enseñanza & Teaching: Revista interuniversitaria de didáctica*, 28(2), 45-60. Recuperado de <https://revistas.usal.es/index.php/0212-5374/article/view/7889>

- Reddy, M. & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, 35(4), 435-448. doi: <https://doi.org/10.1080/02602930902862859>
- Rekalde, I., & Buján, K. (2014). Las eRúbricas ante la evaluación de competencias transversales en Educación Superior. *Revista Complutense de Educación*, 25(2), 355-374. doi: [https://doi.org/10.5209/rev\\_RCED.2014.v25.n2.41594](https://doi.org/10.5209/rev_RCED.2014.v25.n2.41594)
- Rochelle, E., Tractenberg, R.E., Kevin, T, & FitGerarld, K.T. (2012). A Mastery Rubric for the design and evaluation of an institucional curriculum in the responsible conduct of research. *Assessment and Evaluation In Higher Education*, 37(8), 1003-10021. doi: <https://doi.org/10.1080/02602938.2011.596923>
- Sadler, P. M. (2009). Indeterminacy in the use of present criteria for assessment and grading in higher education. *Assessment & Evaluation in Higher Education*, 34,159-179. doi: <https://doi.org/10.1080/02602930801956059>
- Urrutia, M., Barrios, S., Gutiérrez, M. y Mayorga, M. (2014). Optimal method for content validity. *Revista cubana de educación médica superior*, 3, 547-558. Recuperado de [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S0864-21412014000300014](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-21412014000300014)
- Vázquez-Cano, E., Martín-Monje, E. & Fernández-Álvarez, M. (2014). El rol de las e-rúbricas en la evaluación de materiales digitales para la enseñanza de lenguas en entornos virtuales de aprendizaje. *Revista de Docencia Universitaria*, 12(1), 135-157. doi: <https://doi.org/10.4995/redu.2014.6414>
- Velasco, L.C. & Tójar, J.C. (2018). Uso de rúbricas en educación superior y evaluación de competencias. *Profesorado. Revista de Currículum y Formación del Profesorado*, 22(3), 183-208. Recuperado de <https://revistaseug.ugr.es/index.php/profesorado/article/view/7998/6946>
- Velasco, R., Pérez, M., Díaz, A.J., Jordano, M.A. Márquez, C., Soria, E. Sanchidrián, J.L., Bonilla, R., Romero, D., Muñoz, I., Salas, L., Sánchez, B., Mata, J., Carrillo, A. & García Ríos, J.M. (2018). La rúbrica como elemento de evaluación: aplicación y análisis en disciplinas humanísticas. *Revista de innovación y buenas prácticas docentes*, 18-22. Recuperado de <https://www.uco.es/ucopress/ojs/index.php/ripadoc/article/view/11665/10643>
- Villa, A. & Poblete, M. (2011). Evaluación de competencias genéricas: principios, oportunidades y limitaciones. *Bordón*, 63, 147-170. Recuperado de <https://recyt.fecyt.es/index.php/BORDON/article/view/28910>
- Villasís-Keever M. A., Márquez-González, H., Zurita-Cruz, J.N., Miranda-Novales, G., & Escamilla-Núñez, A. (2018). El protocolo de investigación VII. Validez y confiabilidad de las mediciones. *Revista Alergia México*, 65(4), 414-421.

Recuperado de <http://www.scielo.org.mx/pdf/ram/v65n4/2448-9190-ram-65-04-414.pdf>

Zazueta, M.A. & Herrera, L.F. (2008). *Rúbrica o matriz de valoración, herramienta de evaluación formativa y sumativa*. Recuperado de [http://www.quaderns digitals.net/index.php?accionMenu=hemeroteca.VisualizaArticuloIU.visualiza&articulo\\_id=10816](http://www.quaderns digitals.net/index.php?accionMenu=hemeroteca.VisualizaArticuloIU.visualiza&articulo_id=10816)

### **Cómo citar este artículo:**

García-Valcárcel Muñoz-Repiso, A., Hernández Martín, A., Martín del Pozo, M & Olmos Migueláñez, S. (2020). Validación de una rúbrica para la evaluación de trabajos fin de máster. *Profesorado. Revista de Currículum y Formación de Profesorado*, 24(2), p-p. DOI: 10.30827/profesorado.v24i2.15151