Actas de las Jenui, vol. 9. 2024.
Páginas: 67-75

Análisis de la madurez de las rúbricas para la evaluación de los Trabajos Fin de Grado en Ingeniería Informática en el ámbito universitario español

Alfonso de la Vega, Juan María Rivas, Patricia López, Diego García, Julio Medina, Pablo Sánchez
Dpto. Ingeniería Informática y Electrónica
Universidad de Cantabria, Santander (Cantabria)
{alfonso.delavega, juanmaria.rivas, patricia.lopez, diego.garcia, julio.medina, p.sanchez}@unican.es

Resumen

La evaluación de los Trabajos de Fin de Grado presenta una serie de características que la hacen especialmente compleja. Entre otras cuestiones, en este proceso participan un amplio número de profesores, cada uno con su propia especialización, y cuyos criterios de evaluación pueden, y suelen, variar. Consecuentemente, resulta esencial unificar y clarificar estos criterios para garantizar una evaluación justa, predecible y formativa. Para alcanzar este objetivo, las universidades españolas han ido diseñando diferentes rúbricas para la evaluación de sus Trabajos Fin de Grado. Este artículo tiene como objetivo analizar dichas rúbricas en el contexto del Grado en Ingeniería Informática, para así poder conocer su grado de madurez actual y cómo han resuelto ciertos retos inherentes a la evaluación de estos trabajos. Tras este análisis, se podrá identificar con mayor precisión qué retos deberemos abordar en los próximos años para mejorar las rúbricas de evaluación de los Trabajos Fin de Grado en Ingeniería Informática. Para extraer dicha información, hemos realizado una búsqueda de rúbricas en las universidades públicas españolas, y analizado, mediante técnicas de análisis documental, un total de 38 rúbricas encontradas.

Abstract

The evaluation of Final Degree Projects presents a series of characteristics that make it especially complex. Among other issues, many teachers participate in this process, each with their own teaching and research focus, and whose evaluation criteria can, and often do, vary considerably. Consequently, it is essential to unify and clarify these criteria to ensure a fair, predictable, and formative evaluation. To achieve this goal, Spanish universities have been designing rubrics for the evaluation of their Final Degree Projects in recent years. This article aims to analyze these rubrics in the context of the Degree in Informatics Engineering, in order to understand their level of maturity, how hey have solved certain specific challenges inherent to the evaluation of Final Degree Projects and those aspects of these rubrics that still need improvement. With this analysis, we will be able to identify more precisely what challenges we will need to address in the coming years to ensure that the evaluation of these Final Degree Projects is more objective, authentic, and also promotes student learning. To extract this information, we conducted a search for rubrics in public Spanish universities; and analyzed, using document analysis techniques, the 38 rubrics found during this search.

Palabras clave

Rúbricas, Trabajo Fin de Grado, evaluación.

1. Introducción

La evaluación de los Trabajos Fin de Grado, por diversos motivos, resulta particularmente compleja, excediendo en su dificultad al resto de elementos de un grado [1, 3, 4, 5, 8, 13]. En un primer lugar, en la evaluación de estos trabajos interviene habitualmente un amplio número de profesores, cada uno de ellos con percepciones diferentes sobre qué deben contener y con distintos niveles de exigencia [3, 5]. Por ejemplo, la inexistencia de una fase de pruebas durante el desarrollo de un producto software es a veces considerado como un problema menor por algunos profesores, mientras que otros lo entienden como un fallo grave. Dentro de este segundo grupo, algunos entienden este fallo como grave pero no como un impedimento para superar el Trabajo Fin de Grado, mientras que para otros sí lo sería.

En segundo lugar, un Trabajo Fin de Grado suele englobar un alto número de competencias técnicas que el estudiante ha ido adquiriendo a lo largo de la titulación [8, 13, 14]. Los profesores que evalúan un Trabajo Fin de Grado desconocen, en muchas ocasiones, cómo se han evaluado estas competencias en cada asignatura, o a qué nivel de detalle se han trabajado. Por tanto, la evaluación de estas competencias durante un Trabajo Fin de Grado podría diferir de la que ha experimentado el estudiante durante sus estudios. Además, el propio profesorado podría carecer de algunas de estas competencias, debiendo en esos casos evaluar habilidades que ni siquiera posee. Por ejemplo, no es extraño encontrar profesores que deben evaluar el diseño de una red neuronal pero que nunca han diseñado una.

En tercer lugar, los Trabajos Fin de Grado suelen ser muy dispares entre sí, incluyendo temáticas tan variadas como el diseño de sistemas inteligentes, la evaluación de ciertas configuraciones hardware, el diseño de redes de comunicación, el desarrollo de aplicaciones de gestión o la integración de sistemas, entre otras muchas opciones [1, 13].

En cuarto lugar, dado el amplio número de elementos que se evalúan dentro de un Trabajo Fin de Grado, en ocasiones se dejan ciertos elementos sin evaluar simplemente por mero descuido, debido a la complejidad de manejar adecuadamente grandes volúmenes de información [1, 3, 5].

Todos estos factores hacen que la subjetividad en las evaluaciones y calificaciones de los Trabajos Fin de Grado sea mayor de lo deseable [3, 5]. Para solventar este problema, diversos trabajos han propuesto la utilización de rúbricas como mecanismo para tratar de establecer unos criterios comunes y objetivos para la evaluación de los Trabajos Fin de Grado [1, 5, 11, 13]. A lo largo de los últimos años, diferentes universidades españolas han ido adoptando rúbricas para este propósito [9].

Partiendo de estos trabajos, dentro de un proyecto de innovación educativa, nos marcamos como objetivo el diseño de una o varias rúbricas para la evaluación de los Trabajos Fin de Grado en Ingeniería Informática de nuestra universidad. Por tanto, antes de realizar el diseño de nuestra rúbrica, decidimos revisar las rúbricas actualmente existentes para dicho propósito dentro del marco universitario español. En una primera fase, que es donde se situaría este trabajo, analizaríamos cómo las rúbricas disponibles resuelven los problemas anteriormente mencionados. Por ejemplo, cómo responde cada rúbrica a la siguiente pregunta: ¿cómo afecta a la calificación del Trabajo Fin de Grado el hecho de que el estudiante haya ignorado por completo la fase de pruebas del desarrollo de un productos software?

A continuación, analizaríamos qué aspectos comunes tienen aquellas rúbricas que mejor resuelven los problemas previamente identificados y, en base a ellos, diseñaríamos nuestra propia rúbrica, aunque ambos pasos quedan fuera del ámbito de este trabajo.

Por tanto, el objetivo de este artículo es conocer, mediante técnicas de análisis documental [7], cuál es el estado de madurez actual de las rúbricas que se utilizan dentro del ámbito universitario español para la evaluación de los Trabajos Fin de Grado en Ingeniería Informática, para así poder saber con precisión en qué grado ayudan estas rúbricas a reducir la subjetividad de estos procesos de evaluación. Concretamente, se han buscado las rúbricas utilizadas por todas las universidades públicas españolas y se han analizado un total de 38 rúbricas.

Como resultado de este proceso, se pudo comprobar que un gran número de universidades utilizan rúbricas para la evaluación de los Trabajos Fin de Grado en Ingeniería Informática, pero que rara vez estas rúbricas permiten evaluar con un grado adecuado de objetividad y detalle sus aspectos técnicos. Además, muy pocas universidades han diseñado mecanismos para dar un soporte adecuado a la variabilidad inherente a estos trabajos.

El Trabajo Fin de Grado en Ingeniería Informática ha sido objeto de estudio por diversos trabajos anteriores. Por ejemplo, Luna et al. [6] analizaron cómo la evaluación del Trabajo Fin de Grado se alinea con las competencias asociadas al mismo, para lo cual revisaron las normativas, rúbricas e informes asociados a estos trabajos. Otros autores analizaron cómo evaluar adecuadamente los aspectos de sostenibilidad dentro de un Trabajo Fin de Grado, proponiendo una rúbrica concreta para ello [12]. No obstante, hasta donde alcanza nuestro conocimiento, este es el primer trabajo que analiza las características de las rúbricas utilizadas para la evaluación de estos Trabajos Fin de Grado.

En adelante, este artículo se estructura tal como sigue. El apartado 2 describe el método de investigación aplicado. El apartado 3 comenta los resultados obtenidos. Finalmente, el apartado 4 resume las conclusiones extraídas y detalla posibles trabajos futuros.

2. Método

Para poder conocer cuál es el estado actual de madurez de las rúbricas para la evaluación de los Trabajos Fin de Grado en Ingeniería Informática utilizadas por las universidades españolas, aplicamos la metodología de *análisis documental* [7]. El análisis documental es una técnica de investigación cualitativa que permite buscar, extraer y sintetizar información contenida en documentos. Esta técnica resultaba sencilla y rápida de aplicar, al no ser dependiente de terceros, como sería el caso en la realización de cuestionarios o entrevistas. Además, al ser las rúbricas documentos, los

resultados de esta técnica son comparables a los que se pueden obtener utilizando otras técnicas más complejas. La metodología de análisis documental implica la ejecución de las cinco fases que se describen en los siguientes subapartados.

2.1. Definición del problema

Nuestro objetivo final era saber si podíamos utilizar alguna de las rúbricas actualmente existentes dentro del ámbito universitario español para evaluar los Trabajos Fin de Grado en Ingeniería Informática de nuestros estudiantes. Concretamente, buscábamos rúbricas que guiasen el proceso de evaluación de un Trabajo Fin de Grado de manera muy precisa, reduciendo al máximo posible la subjetividad inherente a estos procesos. Por ejemplo, buscábamos rúbricas que, en caso de que hubiese un defecto en el diseño de una base de datos, indicasen con claridad cómo afectaba ese error a la calificación del Trabajo Fin de Grado. Por otra parte, estábamos interesados en rúbricas sencillas, cuya utilización no requiriese un esfuerzo considerable. Se debe tener en cuenta que la evaluación de los Trabajos Fin de Grado se realiza normalmente al término del curso académico, con el profesorado habitualmente exhausto y necesitado de un merecido descanso.

Para encontrar estas rúbricas, nuestra primera pregunta a resolver era cuántas universidades españolas estaban utilizando rúbricas para la evaluación de los Trabajo Fin de Grado y encontrar dichas rúbricas. A continuación, debíamos comprobar el grado de madurez de cada rúbrica con respecto a su *validez* y *fiabilidad*. La *validez* mide en qué grado una rúbrica sirve para realmente evaluar aquello que queremos que el estudiante aprenda. La *fiabilidad* de una rúbrica mide su capacidad para generar evaluaciones similares para un mismo elemento, tanto cuando son producidas por distintos evaluadores como por un mismo evaluador en diferentes instantes de tiempo.

Por otro lado, queríamos poder evaluar la complejidad de las rúbricas encontradas para conocer así el esfuerzo que podría conllevar su utilización. Además, de cara a diseñar nuestra propia rúbrica, estábamos interesados en conocer si estaban aplicando algunas técnicas para gestionar la heterogeneidad existente dentro de los Trabajos Fin de Grado.

Todas estas cuestiones se pueden resumir en las siguientes preguntas de investigación:

PI.1 Pregunta de Investigación 1: ¿Cuántas universidades españolas están utilizando rúbricas para la evaluación de los Trabajos Fin de Grado en Ingeniería Informática, y cuáles son esas rúbricas?

PI.2 Pregunta de Investigación 2: ¿Qué grado de *validez* tienen estas rúbricas?

PI.3 Pregunta de Investigación 3: ¿Qué grado de *fiabilidad* tienen estas rúbricas?

PI.4 Pregunta de Investigación 4: ¿Qué complejidad de uso tienen estas rúbricas?

PI.5 Pregunta de Investigación 5: ¿Cómo se adaptan estas rúbricas a la heterogeneidad de los Trabajos Fin de Grado?

2.2. Selección de los documentos

Para dar respuesta a nuestras preguntas de investigación, utilizamos como documentos las rúbricas para la evaluación de los Trabajos Fin de Grado en Ingeniería Informática de las diferentes universidades públicas del ámbito universitario español. Restringimos la búsqueda a universidades públicas porque dichas universidades comparten una serie de objetivos comunes y se someten a unas inspecciones de calidad similares. Sin embargo, entre las universidades privadas existe una mayor diversidad de casos. Por tanto, para evitar malentendidos y dado que la muestra de las universidades públicas españolas resultaba suficientemente amplia, optamos por descartar inicialmente las universidades privadas de nuestra búsqueda. Obviamente, esta cuestión puede ser abordada por otros autores o en trabajos futuros.

Para encontrar la rúbrica utilizada por cada universidad, seguimos el siguiente procedimiento:

1. Localizábamos, mediante búsqueda directa o navegando a través de la web de cada universidad, la página web que proporcionaba información sobre los Trabajos Fin de Grado en Ingeniería Informática. A continuación, si la había, obteníamos la correspondiente rúbrica de dicha página.
2. En caso de no encontrar la rúbrica en esa página, examinábamos la normativa de los Trabajos Fin de Grado y su correspondiente guía docente, tanto para comprobar si mencionaban la utilización de alguna rúbrica como para ver si se encontraba incluida en algunos de estos documentos.
3. Por último, si las acciones anteriores resultaban infructuosas, escribíamos un correo a la persona de contacto de cada universidad para preguntar por la existencia de la rúbrica buscada.

Tras ejecutar este procedimiento, podíamos dar respuesta a la *pregunta de investigación 1*. Para atender al resto, analizamos cada rúbrica encontrada utilizando el criterio que se describe en el siguiente apartado.

2.3. Definición del criterio de análisis

Antes de definir el criterio de análisis para las rúbricas, describimos brevemente la terminología que vamos a utilizar, para asegurar que ésta sea conocida y entendida por el lector. Toda rúbrica contiene una serie

de *indicadores* o *dimensiones*, que definen los aspectos concretos que se deben evaluar de un determinado artefacto. Cada indicador tiene asociado unos *niveles de logro*, que muestran el desempeño del estudiante dentro del mismo. Por ejemplo, un indicador podría ser la fase de diseño de pruebas del desarrollo de un producto software, mientras que sus niveles de logro podrían ser *Excelente*, *Bueno*, *Aceptable* y *Necesita Mejorar*. Finalmente, los *descriptores de logro* indican qué propiedades concretas debe satisfacer un trabajo para alcanzar un nivel de logro determinado para una dimensión o indicador determinado.

Teniendo en cuenta estos elementos, para analizar la *validez* de cada rúbrica nos centramos en dos aspectos concretos: (1) el ámbito de aplicación de la rúbrica; y, (2) sus indicadores o dimensiones. En primer lugar, si una rúbrica se utiliza sólo para trabajos en Ingeniería Informática, o incluso para un subconjunto concreto de estos trabajos, es de suponer que sólo aborde competencias propias de esta titulación y, además, que pueda hacerlo de manera más o menos detallada. Por otro lado, si la rúbrica es compartida por varias titulaciones, los indicadores deberán ser entonces más generales y menos precisos y, además, puede que incluyan elementos que no son del todo pertinentes para el caso de Ingeniería Informática. Finalmente, analizamos si los indicadores utilizados servían para medir todas las competencias tanto técnicas como transversales que se espera que un estudiante adquiera durante un Trabajo Fin de Grado.

En función de todos esos aspectos, definimos las siguientes niveles de validez:

- Validez 0: No se descompone el Trabajo Fin de Grado en ninguna dimensión.
- Validez 1: Se descompone en ítems muy generales como contenido, memoria y presentación.
- Validez 2: Los aspectos técnicos del Trabajo Fin de Grado se descomponen en varios subelementos, pero hay competencias especificas, como el diseño de pruebas, que no están contempladas, o se incluyen competencias no propias de un Ingeniero Informático, como corrección de los cálculos del proyecto.
- Validez 3: Los indicadores de los aspectos técnicos incluyen de manera completa todas las competencias que un estudiante podría necesitar ejercitar para desarrollar correctamente un Trabajo Fin de Grado, y no se incluyen competencias no propias de un Ingeniero Informático.

Merece la pena destacar que estos niveles de validez no intentan en modo alguno fomentar el uso de *rúbricas analíticas*, aquellas que compartimentan la evaluación en apartados estancos, frente a *rúbricas holísticas*, aquellas que consideran el trabajo como un todo, ya que tal como muestran ciertas evidencias [2], una buena rúbrica debe combinar aspectos analíticos con aspectos holísticos. Por ejemplo, un Trabajo Fin de Grado que presente claras deficiencias técnicas, con errores conceptuales graves, debería tener una calificación de suspenso, con independencia de lo bien escrito o lo bien presentado que haya podido estar. No obstante, consideramos importante que una rúbrica detalle con claridad todos los aspectos que deben evaluarse de un Trabajo Fin de Grado, de manera que, por ejemplo, un determinado tribunal no pueda olvidarse de evaluar la fase de pruebas del desarrollo de un producto software, con independencia de que estos se califiquen luego de manera holística o analítica.

Con respecto a la *fiabilidad*, y siguiendo las recomendaciones de Jonsson y Svingby [10], analizamos si las dimensiones se descomponían en un número adecuado de niveles de logro, y si las descripciones de estos niveles de logro proporcionaban instrucciones precisas para su interpretación, incluyendo incluso algunos ejemplos concretos de aplicación. En base a estos argumentos, definimos los siguientes niveles:

- Fiabilidad 0: En las dimensiones, no hay descriptores de qué se considera un trabajo excelente o correcto. Por ejemplo, se indica que hay que evaluar la presentación de un Trabajo Fin de Grado, pero no se define qué es una buena presentación.
- Fiabilidad 1: Hay indicaciones de qué se considera un trabajo correcto para cada dimensión, pero no se establecen niveles de logro. Por tanto, cuando una dimensión no es 100 % correcta, no se sabe con qué calificación concreta debe evaluarse.
- Fiabilidad 2: Se dan indicaciones generales que pueden ser interpretadas de manera diferente por distintos evaluadores. Por ejemplo, la descripción simplemente indica que el diseño de la base de datos es *generalmente* correcto. En este caso, un determinado defecto podría ser considerado como un fallo menor por un evaluador, mientras que otro evaluador podría consideralo como un fallo grave y le asignaría otro nivel de logro.
- Fiabilidad 3: Los descriptores de logro proporcionan instrucciones claras y precisas, acompañadas de ejemplos, sobre cómo alcanzar un determinado nivel de logro.

Para medir la complejidad de cada rúbrica, utilizaremos el número total de indicadores y niveles de logro de cada rúbrica. Cuanto más altos sean estos números, más información tendremos que considerar y mayor será el trabajo de evaluación, aunque la claridad de las instrucciones también afecta claramente al tiempo que debemos dedicar a una tarea de evaluación. A mayor claridad, menos dudas nos surgirán, y menos tiempo tardaremos.

Por último, para dar respuesta a la pregunta de investigación 5, anotaremos para cada rúbrica si utiliza o no mecanismos para dar soporte a la heterogeneidad de los Trabajos Fin de Grado, junto con una breve descripción de cada mecanismo.

2.4. Revisión de los documentos

Para buscar, revisar y analizar las rúbricas, dividimos las universidades públicas españolas entre los autores de este trabajo, de manera que cada universidad fuese procesada por un único autor. El trabajo individual de cada autor fue revisado por otro autor. Cuando surgió alguna duda o discrepancia, se puso en conocimiento del resto de autores, que procedieron a debatirla para poder adoptar una decisión consensuada.

Tras extraer los datos requeridos de los documentos, procedimos a su síntesis y análisis. El resultado se describe en el siguiente apartado.

3. Resultados

En este apartado tratamos de dar respuesta a cada una de las preguntas de investigación planteadas utilizando los datos recopilados por nuestro proceso de búsqueda. Estos datos pueden consultarse en un repositorio externo.[1]

3.1. PI.1 - Grado de utilización

Los datos recabados indican que la gran mayoría de universidades españolas utilizan algún tipo de rúbrica para evaluar los Trabajos Fin de Grado. Tal como se puede ver en la Figura 1, de las 38 universidades de 48 que imparten el Grado en Ingeniería Informática, un 79.17 % utilizan rúbricas, 6 no las utilizan, y para 4 no pudimos obtener información.

Además, pudimos comprobar que estas rúbricas suelen estar públicamente accesibles en sus páginas web, ya que sólo en un único caso no lo estaba y la rúbrica nos fue proporcionada tras preguntar por ella mediante un correo electrónico.

Para las universidades de las que no tenemos información, no encontramos ninguna rúbrica en sus páginas web ni obtuvimos respuesta a los correos electrónicos enviados. No obstante, de acuerdo con nuestra experiencia, si estas rúbricas no están públicamente disponibles, lo más probable es que no existan.

Como dato curioso, merece la pena destacar que una universidad respondió a nuestro correo electrónico indicando que habían utilizado una rúbrica en años anteriores, pero no les había resultado efectiva y la habían dejado de utilizar.

[1] https://doi.org/10.6084/m9.figshare.25594980

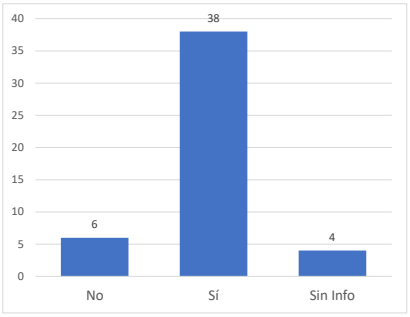

Figura 1: Utilización de rúbricas

3.2. PI.2 - Validez de las rúbricas

Tal como muestra la Figura 2, la mayoría de rúbricas se mueven entre los niveles de validez 1 y 2. Un 52.63 % de las rúbricas analizadas, es decir, 20 de ellas, presentan un nivel de validez 1. Estas rúbricas indican simplemente que, para cada trabajo, se deben evaluar por separado su contenido, memoria y su presentación. Esta división es absolutamente genérica y valdría incluso para un trabajo de una asignatura que tuviese asociada una presentación oral. Por tanto, esta división no garantiza que se evalúe correctamente que el estudiante haya adquirido las competencias concretas asociadas al Trabajo Fin de Grado.

Otro 42.10 % de rúbricas, es decir, 16 de ellas, alcanzan el nivel de validez 2, al refinar contenido, memoria y presentación en una serie de subelementos. No obstante, estos subelementos, en particular en lo concerniente al contenido, suelen ser muy genéricos. Por lo general, estas rúbricas evalúan aspectos como la correcta selección de herramientas, la corrección técnica del proyecto como un único elemento, la metodología y la planificación seguidas o la solidez del producto creado. Al no descomponerse la corrección técnica en más niveles, diferentes tribunales podrían valorar ciertos aspectos técnicos de un Trabajo Fin de Grado de manera distinta, o incluso obviarlos. Por ejemplo, un determinado tribunal podría dar mucha más importancia a la fase de diseño arquitectónico que otro. De igual forma, un tribunal podría considerar la ausencia de una adecuada fase de pruebas como un error muy grave, mientras que otro tribunal, por despiste, podría olvidarse incluso de que tiene que evaluar dicha fase de pruebas. Por tanto, consideramos conveniente descomponer estos aspectos de corrección técnica en más subniveles que actúen como lista de comprobación de lo que ha de evaluar cada tribunal.

Sólo una rúbrica, la de la Universidad de Vigo, posee un grado de validez 3. En ella, la evaluación del contenido está alineada con las fases de desarrollo de un proyecto de Ingeniería Informática, por lo que la rúbrica es, en cierto modo, una lista de comprobación de

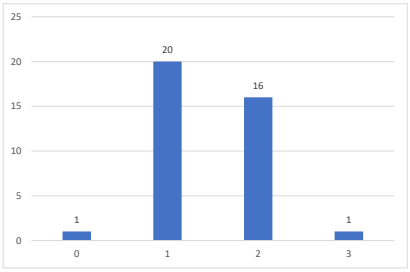

Figura 2: Niveles de validez

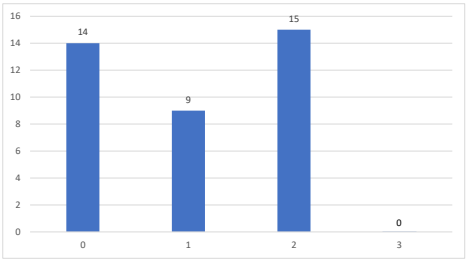

Figura 3: Niveles de fiabilidad

tareas a completar durante el desarrollo de un Trabajo Fin de Grado. No obstante, esta rúbrica sólo permite evaluar el desarrollo de proyectos de ingeniería. Otros tipo de trabajos, como los de evaluación y análisis de algoritmos, necesitarían de otro tipo de rúbrica.

Por otro lado, el 65.79 % de universidades define rúbricas a nivel de titulación, mientras que en 34.21 % restante están definidas a nivel de centro, normalmente dentro de una escuela de ingeniería. En estos últimos casos se comparte titulación con ingenierías de todo tipo, incluyendo desde telecomunicaciones hasta ingeniería civil o mecánica. Esta cuestión no parece afectar al nivel de validez de las rúbricas, pues un 61.54 % de las rúbricas definidas a nivel de centro alcanzan el nivel de validez 2, mientras que a nivel de titulación sólo el 44 % llega hasta esa validez o más. No obstante, entendemos que pasar de un nivel de validez 2 a un nivel 3 resultará siempre más fácil para las rúbricas definidas para una titulación concreta, ya que podrán centrarse en aspectos específicos de dicha titulación, cosa que resulta más compleja con rúbricas definidas para varias titulaciones levemente relacionadas.

Por lo general, las rúbricas no suelen contener elementos que no sean de aplicación al Grado de Ingeniería Informática, salvo en el caso de un par de rúbricas definidas a nivel de centro que incluyen aspectos de documentación técnica o cálculos aplicables exclusivamente a otras ingenierías.

3.3. PI.3 - Fiabilidad de las rúbricas

Como se puede observar en la Figura 3, ninguna universidad alcanza el nivel de fiabilidad 3. Es decir, ninguna universidad establece unos criterios claros que puedan ser interpretados sin ambigüedades por distintos evaluadores. Este hecho era esperado, ya que es muy complicado proporcionar estas instrucciones para los aspectos técnicos de los Trabajos Fin de Grado, tanto por la envergadura de estos trabajos como por la variabilidad de sus contenidos.

El resto de opciones aparecen equilibradas. Un 36.84 % de las rúbricas analizadas, es decir, un total de 14 rúbricas, tienen una fiabilidad de nivel cero, donde ni siquiera existe una indicación de qué se considera exactamente correcto para una dimensión determinada. Por tanto, la evaluación queda completamente a criterio de los evaluadores, que podrían interpretar cada dimensión de formas muy variadas. Otro 23.68 %, 9 rúbricas, define qué se espera de cada dimensión, pero no establece niveles de logro para ellas. Por tanto, los evaluadores conocen qué elementos se deben evaluar en cada dimensión, pero podrían evaluarlos de manera diferente en función de sus criterios y niveles de exigencia personales.

Finalmente, 15 rúbricas, un 39.47 %, definen niveles de logro, aunque las descripciones de los niveles de logro podrían ser interpretadas de manera diferente por distintos evaluadores. Por ejemplo, es habitual especificar que el nivel de excelencia se alcanza cuando algo es completamente correcto, el siguiente nivel cuando existen errores leves y el siguiente cuando hay errores graves, pero en ningún momento se define qué es un error leve y qué es un error grave. Por tanto, esta diferenciación queda a criterio personal de los evaluadores.

También merece la pena destacar que prácticamente ninguna de las rúbricas analizadas proporciona ejemplos, tal como recomiendan Jonsson y Svingby [10], que permitan interpretar sus niveles de logro con mayor exactitud. Destaca a este respecto la rúbrica de la Universidad de Málaga, la cual proporciona algunos ejemplos de qué se considera como error leve o grave, aunque no lo hace para todos los elementos a evaluar.

Por último, no parece existir relación entre el ámbito para el cual se define la rúbrica y su fiabilidad. Para las rúbricas definidas a nivel de escuela de ingeniería, un 53.85 % alcanzan un nivel de fiabilidad 2, mientras que sólo un 36 % de las definidas a nivel de Grado en Ingeniería Informática alcanzan ese nivel.

3.4. PI.4 - Complejidad de las rúbricas

La Figura 4 indica que hay un número destacado de universidades que sólo definen tres elementos a analizar en sus rúbricas, y que suelen ser contenido, memoria y presentación. El resto de rúbricas, el 81.58 %, define un número de elementos a evaluar que varía más o menos uniformente entre 4 y 33.

Tal como hemos podido constatar, resulta difícil ex-

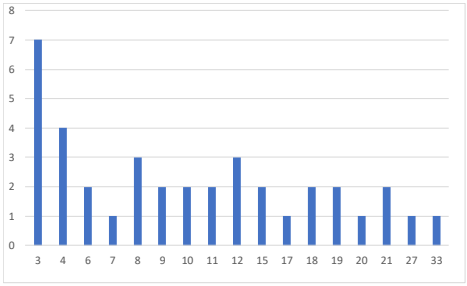

Figura 4: Número de elementos

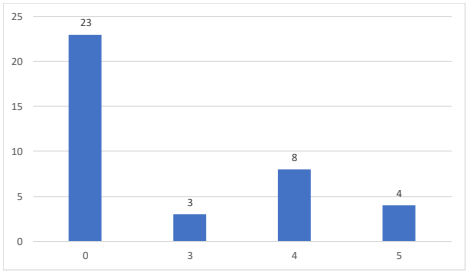

Figura 5: Número de niveles de logro

traer conclusiones sobre la complejidad de las rúbricas analizando exclusivamente sus documentos y su número de elementos a evaluar, ya que estos elementos tienen una granularidad muy diversa. En ocasiones un único elemento hace referencia a todo el contenido técnico del Trabajo Fin de Grado, mientras que en otros casos un elemento se refiere únicamente a un aspecto muy concreto del mismo, como la estabilidad del producto final. Por tanto, estos elementos pueden requerir tiempos de evaluación muy diversos. Consecuentemente, la complejidad de las rúbricas tendrá que ser analizada en el futuro por medio de otro tipo de técnicas de investigación educativa.

La Figura 5 especifica cuántos niveles de logro tienen las rúbricas analizadas. Obviamente, las rúbricas con niveles de fiabilidad inferiores a 1 ó 2, que son 23, no tienen niveles de logro. Para las 16 rúbricas con nivel de fiabilidad 2, los niveles de logro oscilan entre 3 y 5, siendo 4 el valor más común. Estos cuatro niveles suelen ser similares a *deficiente*, *apto*, *notable* y *excelente*. En el caso de cinco niveles, el nivel deficiente se descompone habitualmente en *muy deficiente* y *deficiente*, mientras que en el caso de rúbricas con 3 niveles se fusionan los niveles *apto* y *notable* en un único nivel.

3.5. PI.4 - Gestión de la variabilidad

Por último, tal como se observa en la Figura 6, sólo 3 universidades disponen de mecanismos para adaptar las rúbricas a la variabilidad de los Trabajos Fin de Grado. En un caso, la variabilidad se soporta haciendo

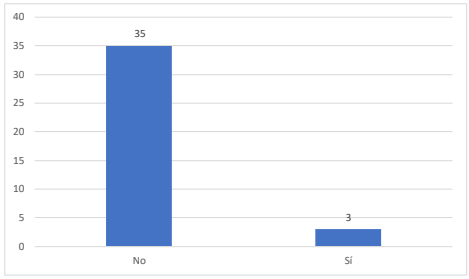

Figura 6: Soporte a la heterogeneidad

que ciertas dimensiones de la rúbrica sean opcionales.

En los otros casos, la variabilidad se gestiona mediante la definición primero de distintos tipos de trabajos. A continuación, para cada tipo de trabajo se define su propia rúbrica.

Si bien este es un primer paso para gestionar esta variabilidad, podría aún resultar insuficiente, ya que habitualmente sólo se distinguen dos tipos distintos de trabajos. Por ejemplo, para una universidad que distingue entre trabajos científicos y tecnológicos, dentro de los trabajos tecnológicos se integrarían proyectos de desarrollo software, de modernización de sistemas o de diseño de redes neuronales, entre muchas otras posibilidades. Cada uno de estos trabajos tiene sus propias peculiaridades que deberían ser consideradas individualmente.

3.6. Amenazas a la validez

Como en todo trabajo de investigación, existen potenciales amenazas a la validez de los resultados mostrados, que comentamos a continuación.

La primera amenaza proviene del propio método de investigación utilizado. Al analizar las rúbricas como documentos, sin considerar en ningún momento la opinión de los usuarios de dichas rúbricas, estamos obteniendo una visión sesgada de las mismas. Puede ser, por ejemplo, que algún elemento que hayamos considerado como ambiguamente especificado, no resulte ambiguo en un contexto determinado. Por ejemplo, dentro de un centro podría haberse debatido en profundidad cuál es la complejidad adecuada para un Trabajo Fin de Grado y, como consecuencia de dichos debates, este conocimiento podría haber sido interiorizado por sus profesores. No obstante, hay que aclarar que en estos casos las rúbricas no cumplirían con la propiedad de ser instrumentos de evaluación que puedan ser utilizados por cualquier profesor, y este conocimiento contextual implícito podría perderse o difuminarse con el tiempo.

Como segunda amenaza, hemos extraído las rúbricas de las páginas web de cada universidad, sin corroborar si dichas rúbricas se están realmente utilizando o

si está previsto su reemplazo en breve. Por tanto, podría darse el caso de que parte de la información analizada esté obsoleta o en desuso.

Como tercera amenaza, el proceso de clasificación de las rúbricas en diferentes niveles de validez y fiabilidad tiene un cierto grado de subjetividad. Por tanto, dos investigadores podrían clasificar una misma rúbrica de manera diferente. Para mitigar este problema, la clasificación de cada rúbrica fue revisada por un autor diferente al que propuso la clasificación inicial. Cuando se detectaron discrepancias, éstas se debatieron entre todos los autores.

4. Conclusiones

En este artículo hemos analizado, mediante técnicas de análisis documental, la madurez de las rúbricas utilizadas actualmente para la evaluación de los Trabajos de Fin de Grado en Ingeniería Informática por las universidades públicas españolas.

Tras revisar 50 universidades públicas españolas, se ha podido constatar que la mayoría de ellas, un 79.17 %, utilizan algún tipo de rúbrica para la evaluación de los Trabajos de Fin de Grado en Ingeniería Informática. Además, estas rúbricas, por lo general, se encuentran públicamente accesibles a través de la página web de cada universidad.

Con respecto a la validez, aproximadamente la mitad de estas rúbricas no van más allá de la simple indicación de que se debe evaluar contenido, memoria y presentación por separado; mientras que la otra mitad sí define qué elementos concretos hay que considerar dentro del contenido, memoria y presentación de un Trabajo Fin de Grado, No obstante, en estos casos, la división de estos tres elementos suele estar poco alineada con las competencias técnicas que se deben ejercitar durante el desarrollo de estos trabajos. Sólo una rúbrica descompone el contenido conforme a las etapas habituales del proceso de desarrollo de un proyecto de Ingeniería Informática.

El 65.79 % de las rúbricas se definen para el Grado de Ingeniería Informática, mientras que el resto se hace a nivel de Escuela de Ingeniería, siendo en estos casos compartida por ingenierías muy diversas. No obstante, sólo en dos rúbricas definidas a nivel de Escuela se incluyen dimensiones que no son relevantes para la Ingeniería Informática.

En lo concerniente a la fiabilidad, un número considerable de rúbricas, aproximadamente el 34.21 %, no proporciona instrucciones acerca de cómo evaluar sus dimensiones. Un 23.68 % sí indica cómo evaluar cada una de sus dimensiones, pero no establece diferentes niveles de logros, por lo que queda a criterio del tribunal cómo calificar exactamente cada dimensión. Finalmente, el restante 42.11 % de las rúbricas analizadas sí establece diferentes niveles de logro, pero las descripciones de estos niveles contienen ciertas ambigüedades. Prácticamente ninguna rúbrica contiene ejemplos de cómo interpretar adecuadamente sus niveles de logro. Por tanto, la fiabilidad de estas rúbricas es un aspecto claro donde se debe seguir trabajando.

Respecto a su complejidad, las rúbricas analizadas contienen un número muy variable de elementos a evaluar, que van de los 3 a los 33 elementos. En base a nuestras observaciones, no se puede concluir que un mayor número de elementos implique una mayor complejidad, ya que el tamaño y complejidad de estos elementos es muy variable. La mayoría de las rúbricas utilizan los cuatro niveles de logros clásicos (*deficiente*, *apto*, *notable* y *excelente*), aunque algunas universidades distinguen entre dos niveles de *deficiente* y otras fusionan los niveles *apto* y *notable* en un único nivel.

Por último, a pesar de la gran variabilidad existente entre los Trabajos Fin de Grado, sólo cuatro rúbricas, aproximadamente un 10.53 % de las analizadas, ofrecen mecanismos para dar soporte a dicha variabilidad. Estos mecanismos consisten en la utilización de rúbricas diferentes para cada tipo de proyecto o la especificación de ciertos elementos a evaluar como opcionales.

Como resultado de este análisis podemos concluir que no existe actualmente ninguna rúbrica que cumpla con los requisitos que buscábamos, que eran reducir lo máximo posible la subjetividad existente dentro de la evaluación de los Trabajos Fin de Grado. Consecuentemente, deberemos trabajar en el diseño de nuestra propia rúbrica, tomando como base las ya existentes. Para ello, la rúbrica de la Universidad de Oviedo propociona una excelente base en cuanto a validez, mientras que la de la Universidad de Málaga sería el camino a seguir en cuanto a fiabilidad.

Como trabajo futuro, tenemos interés en extender este análisis a un conjunto adecuadamente seleccionado de universidades privadas y extranjeras. Además, nos gustaría complementar los resultados obtenidos en este trabajo con otras técnicas de investigación cualitativa, como la realización de cuestionarios o entrevistas, para tratar de obtener información más detallada sobre los problemas y bondades de ciertas rúbricas. Por ejemplo, sería interesante determinar, mediante entrevistas o cuestionarios, cuáles son las ventajas e inconvenientes de tener tres niveles de logro en lugar de cuatro, o cuánto esfuerzo requiere realmente la evaluación de un Trabajo Fin de Grado utilizando una rúbrica con un alto número de elementos, una vez que los evaluadores están familiarizados con la misma.

Agradecimientos

Financiado por la VI Convocatoria de Innovación Docente de la Universidad de Cantabria.

Referencias

[1] Ehsan Ahmad, Bilal Raza y Robert Feldt. Assessment and support for software capstone projects at the undergraduate level: A survey and rubrics. En *Proceedings of 9th Internal Conference on Frontiers of Information Technology (FIT)*, pp. 25–32, Islamabad (Pakistan), diciembre 2011.

[2] Josette Bettany-Saltikov, Stephanie Kilinc y Karen Stow. Bones, boys, bombs and booze: an exploratory study of the reliability of marking dissertations across disciplines. *Assessment & Evaluation in Higher Education*, 34(6):621–639, diciembre 2009.

[3] César Domínguez, Arturo Jaime, Francisco J. García-Izquierdo y Juan J. Olarte. Factors considered in the assessment of computer science engineering capstone projects and their influence on discrepancies between assessors. *ACM Transactions on Computing Education*, 20(2):14:1–14:23, marzo 2020.

[4] Robert F. Dugan. A survey of computer science capstone course literature. *Computer Science Education*, 21(3):201–267, septiembre 2011.

[5] Vivienne Farrell, Gilbert Ravalli, Graham Farrell, Paul Kindler y David Hall. Capstone project: fair, just and accountable assessment. En *Proceedings of the 17th Annual Conference on Innovation and Technology in Computer Science Education (ITiCSE)*, pp. 168–173, Haifa (Israel), julio 2012.

[6] Juan Manuel Fernández Luna, Eugenio Martínez Cámara, Rocío Romero Zaliz, Pablo García Sánchez, Alberto Guillén, Manuel Noguera y María José Rodríguez Fórtiz. Qué y cómo se evalúa en el TFG del Grado en Ingeniería Informática en España. En *Actas de las XXIX Jornadas sobre Enseñanza Universitaria de la Informática (JENUI)*, volumen 8, pp. 307–314, Granada (Andalucía, España), julio 2023.

[7] Tanya Fitzgerald. Documents and documentary analysis. En Ann R. J. Briggs, Marianne Coleman y Marlene Morrison, editores, *Research Methods in Educational Leadership & Management*, pp. 296–308. SAGE, 3 edition, 2012.

[8] Nicole Herbert. Reflections on 17 years of ict capstone project coordination: Effective strategies for managing clients, teams and assessment. En *Proceedings of the 49th ACM Technical Symposium on Computer Science Education (SIGCSE)*, pp. 215–220, Baltimore (Maryland, USA), febrero 2018.

[9] Antoni Jaume-i Capó, Carlos Guerrero, Joe Miró y Antonio Egea. Elaboración de una rúbrica para la evaluación TFG y TFM de informática en la Universitat de les Illes Balears. En *Actas del Simposio-Taller de las XVII Jornadas sobre Enseñanza Universitaria de la Informática (JENUI)*, pp. 17–24, Ciudad Real (Castilla La Mancha, España), julio 2012.

[10] Anders Jonsson y Gunilla Svingby. The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2(2):130–144, enero 2007.

[11] María del Carmen Pegalajar. La rúbrica como instrumento para la evaluación de trabajos fin de grado. *Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación (REICE)*, 19(3):67–81, junio 2021.

[12] Fermín Sánchez, Jordi García, Eva Vidal, David López, José Cabré, Helena García y Marc Alier. Guía y evaluación de la sostenibilidad en los Trabajos de Fin de Grado. En *Actas de las XIX Jornadas sobre Enseñanza Universitaria de la Informática (JENUI)*, pp. 34–41, Andorra La Vella (Andorra), julio 2015.

[13] Saara Tenhunen, Tomi Männistö, Matti Luukkainen y Petri Ihantola. A systematic literature review of capstone courses in software engineering. *Information and Software Technology*, 159:107191, julio 2023.

[14] Brian R. von Konsky y Jim Ivins. Assessing the capability and maturity of capstone software engineering projects. En *Proceedings of the 10th Conference on Australasian Computing Education (ACE)*, pp. 171–180, Wollongong (New South Wales, Australia), enero 2008.