



EVALUACIÓN DE LA CALIDAD EN ESTUDIOS SECUNDARIOS: UN ESTUDIO DE CASO

*José Ángel Ruiz Escalante**

*Raúl Antonio Aguilar Vera***

*Juan Francisco Garcilazo Ortiz****

*Antonio Armando Aguilera Güemez*****

Recibido: 06/10/2024 - Aceptado: 14/02/2025

<https://doi.org/10.22395/rium.v24n46a5>

RESUMEN

La necesidad de investigar la Ingeniería de Software surge desde su propia concepción, por lo que se utilizan diversas metodologías de investigación, como la Revisión Sistemática de la Literatura. Una parte importante del proceso de realización de este tipo de estudio es la evaluación de la calidad de los estudios primarios recopilados. El objetivo es ejemplificar cómo se realiza, modifica y adapta el proceso de evaluación. Se utiliza como estudio de caso la Revisión Sistemática de la Literatura titulada “Human Factors and their Impact on Technical Debt Management”.

En conclusión, la evaluación de la calidad es una de las tareas más importantes dentro del proceso de realización de una revisión; permite conocer si los estudios recopilados sobre un área específica son de calidad o no, y a su vez encontrar las áreas débiles u oportunidades de los estudios en base a los criterios de calidad para informar y ayudar en el área de investigación sobre dónde se debe prestar más atención a la hora de realizar un estudio primario para mejorar la calidad de los estudios y del área en general. Por último, permite determinar una calificación para cada estudio seleccionado y así crear una distinción entre ellos para priorizar la atención, la lectura y la confianza de sus conclusiones para ayudar en la síntesis de la revisión de la literatura.

Palabras clave: Ingeniería de Software, Ingeniería de Software Empírica, Criterios de Evaluación de la Calidad, Revisión Sistemática de la Literatura.

* Universidad Autónoma De Yucatán / Facultad De Matemáticas. Mérida, México. Correo electrónico: a19216272@alumnos.uady.com Orcid: <https://orcid.org/0009-0009-3019-5500>

** Universidad Autónoma De Yucatán / Facultad De Matemáticas. Mérida, México. Correo electrónico: avera@correo.uady.mx Orcid: <https://orcid.org/0000-0002-1711-7016>

*** Universidad Autónoma De Yucatán / Facultad De Matemáticas. Mérida, México. Correo electrónico: gortiz@correo.uady.mx Orcid: <https://orcid.org/0000-0002-3005-3432>

**** Universidad Autónoma De Yucatán / Facultad De Matemáticas. Mérida, México. Correo electrónico: aaguilet@correo.uady.mx Orcid: <https://orcid.org/0000-0001-5155-3543>

QUALITY ASSESSMENT IN SECONDARY STUDIES: A CASE STUDY

ABSTRACT

The need to investigate Software Engineering comes from its very conception, so various research methodologies are used, such as the Systematic Literature Review. An important part of the process of carrying out this type of study is the quality assessment of the primary studies that have been collected. The objective is to exemplify how the evaluation process is carried out, modified and adapted. The Systematic Literature Review entitled “Human Factors and their Impact on Technical Debt Management” is used as a case study. In conclusion, quality assessment is one of the most important tasks within the process of carrying out a review; it allows us to know if the studies collected on a specific area are of quality or not, and in turn to find the weak or opportunities areas of the studies based on the quality criteria to inform and help in the research area on where more attention should be paid when carrying out a primary study to improve the quality of the studies and the area in general; Finally, it allows determining a rating for each selected study and thus creating a distinction between them to prioritize the attention, reading and confidence of their findings to help in the synthesis of the literature review.

Keywords: Software Engineering, Empiric Software Engineering, Quality Assessment Criteria, Systematic Literature Review.

1. INTRODUCCIÓN

Con el paso de los años se han desarrollado disciplinas afines al *software*. Una de estas es la ingeniería de *software* (IS), que se puede entender como aquella disciplina que desarrolla el *software* utilizando diversas metodologías, procesos y herramientas definidas, para crear así productos de alta calidad [1], [2].

Por lo anterior, el origen de la necesidad de investigar en el área de la IS viene desde su misma concepción en la década de 1960 [3], ya que, según [4], el simple concepto de “ingeniería” obliga a que los productos se desarrollen con un enfoque científico y sistemático. Por otro parte, el uso de métodos empíricos proporciona esa parte científica con evidencia sobre herramientas y técnicas que necesitamos. De acuerdo con [5], “Software engineering is fundamentally an empirical discipline” (La ingeniería de *software* es fundamentalmente una disciplina empírica). Por ende, fue creada la ingeniería de *software* empírica, que es el área de investigación que hace uso de métodos empíricos en la IS [4].

De acuerdo con [5], desde el año 2000, las metodologías de investigación (experimentación, encuesta, estudio de caso, entre otros) recibieron una atención considerable y, por lo tanto, el número de estudios empíricos primarios aumentó significativamente. Dicha tendencia provocó el surgimiento de estudios secundarios para poder recopilar y sintetizar los diferentes estudios primarios, y así fortalecer la evidencia sobre los fenómenos de *software* o para revelar la necesidad de más estudios primarios sobre el fenómeno de interés.

En este artículo se discute el proceso de evaluación de la calidad de los estudios primarios recopilados en una revisión sistemática de literatura (RSL), porque esta evaluación es importante para saber hasta qué punto los resultados del estudio son válidos y libres de sesgos, y aumentar el valor del estudio explicando las diferencias entre los resultados de los estudios seleccionados, ponderar la importancia de los estudios recopilados, interpretar los resultados de la revisión y orientar las recomendaciones para futuras investigaciones [6], [7]. Esta RSL fue realizada para un trabajo de tesis de la licenciatura de Ingeniería de Software en la Universidad Autónoma de Yucatán (UADY) [8].

2. EVALUACIÓN DE LA CALIDAD

Evaluar la calidad de un estudio es una tarea difícil, ya que no existe una definición acordada de “calidad” [6], [7]. Por lo tanto, esta debe definirse, y para ello se pueden utilizar y adaptar diferentes instrumentos de calidad, como las listas de verificación de factores que deben evaluarse a cada estudio primario [7]. A cada uno de los elementos

para evaluar la calidad se le pueden asignar escalas numéricas [7] y los revisores pueden optar por excluir del proceso de síntesis los estudios primarios de baja calidad o comprobar si la exclusión de estos tiene un efecto significativo en los resultados de la revisión [6]. Uno de estos instrumentos para evaluar la calidad es el establecido por [9]. Allí se define a través de cuatro criterios:

- **Calidad del informe:** Se relaciona con la justificación, los objetivos y el contexto de un estudio [9].
- **Rigor:** Se relaciona con los métodos de investigación empleados para establecer la validez de los instrumentos de recolección de datos y los métodos de análisis, y, por lo tanto, con la confiabilidad de los hallazgos [9].
- **Credibilidad:** Se relaciona con la evaluación de los métodos de estudio para asegurar que los hallazgos sean válidos y significativos [9].
- **Relevancia:** Se relaciona con la evaluación de la pertinencia del estudio para la industria del *software* en general y la comunidad investigadora [9].

Por último, para evaluar cada uno de los anteriores criterios, [9] define diferentes preguntas para cada uno (11 preguntas en total), que se responden de forma dicotómica, es decir, preguntas con respuestas de “Sí” o “No”. Además, a cada respuesta se le otorga un puntaje de 1 y 0, respectivamente.

3. ESTUDIO DE CASO

El estudio de caso que se utilizará para ejemplificar cómo se lleva a cabo, se modifica y se adecua. El proceso de evaluación de la calidad es la RSL titulada “Factores humanos y su incidencia en la gestión de deuda técnica”, que tiene como objetivo recopilar y sintetizar toda la información de los estudios primarios que se enfoquen o hablen sobre la relación entre deuda técnica (DT) y factores humanos (FH), para así obtener un estado del arte de este tema y conocer su situación actual [8]. Dicha RSL fue realizada según lo determinado por [7]. En ella existen tres fases principales: planeación, ejecución e informe de la revisión.

3.1. PLANIFICACIÓN

Como parte de la fase de planeación, se define todo lo necesario para realizar la evaluación de la calidad. En este caso, el análisis de los estudios se realiza con el objetivo de establecer una calificación para dar prioridad a ciertos estudios al momento de revisarlos y tomar en cuenta sus hallazgos sobre otros [8]. El instrumento de evaluación se basa

en [9] con modificaciones en las definiciones de los criterios, preguntas y calificación, tanto propias como tomadas de otros autores [10], [11], [12], [13].

Además, se utiliza como apoyo para la evaluación de la calidad el enfoque de resúmenes estructurados de [14].

3.1.1. CRITERIOS DE CALIDAD

Tal y como lo menciona [9], se define la calidad a través de cuatro criterios: calidad del informe, rigor, credibilidad y relevancia, cada uno de los cuales tiene diferentes preguntas. En este caso se definieron 11 preguntas en total, que se podrán responder de tres maneras diferentes dependiendo de si cumple totalmente, parcialmente o no cumple. A cada respuesta se le asignó una determinada puntuación para poder evaluarlos cuantitativamente: 1, 0.5 o 0, respectivamente [12]. Cabe destacar que las preguntas fueron realizadas para encajar con cualquier tipo de estudio primario, es decir, experimento, estudio de caso, encuesta, entre otros.

Calidad del informe

Evalúa si el contexto, el objetivo, el enfoque y la manera de presentar un estudio se consideran significativos [9]. Es decir, qué tan completos son los antecedentes de la investigación [10].

A continuación se plantean las preguntas de evaluación (PE) y la forma de calificarlas para evaluar este criterio.

- **PE1.** ¿Existe una descripción clara de los objetivos de la investigación? [9], [12]
 - Los objetivos son presentados y descritos de forma explícita. (1)
 - Los objetivos son ambiguos o se pueden extraer de forma implícita. (0.5)
 - No hay objetivos definidos de ninguna forma. (0)
- **PE2.** ¿Está claramente descrito el contexto del estudio? [9], [11]
 - El sector y los participantes son presentados y descritos de forma explícita. (1)
 - El sector o los participantes son ambiguos o se pueden extraer de forma implícita. (0.5)
 - El sector y los participantes no son definidos de ninguna forma. (0)
- **PE3.** ¿El estudio se centra en deuda técnica (DT) y factores humanos (FH)? [10]

- El enfoque del estudio en su totalidad es sobre la relación entre DT y FH. (1)
- Si bien la relación entre DT y FH no es el tema principal, es uno de los aspectos analizados o se centra únicamente en uno de los dos. (0.5)
- El tema principal no es la DT ni los FH y solo trata de manera tangencial la temática. (0)
- **PE4.** ¿El uso de tablas y figuras es de ayuda para poder presentar una información o hallazgos a los lectores? [12], [13]
 - Todas las tablas y/o figuras son comentadas y complementan la información escrita. (1)
 - No todas las tablas y/o figuras son comentadas o en su caso su aportación es limitada. (0.5)
 - No hay tablas ni figuras o en caso de existir no son comentadas o no ofrecen aportación alguna. (0)

Hablando especialmente de la PE3, esta pregunta fue definida, ya que, en la ejecución, después de haber aplicado los criterios de inclusión y exclusión, se obtienen los estudios vinculados con el tema central [8], pero es interesante saber qué tanto y cuántos de los estudios seleccionados en verdad se enfocan sobre el tema de interés. Además de que esta pregunta encaja perfectamente con la definición calidad del informe.

Rigor

Evalúa si los métodos son correctamente utilizados para establecer la validez de los datos y están bien descritos [9], [10].

A continuación se plantean las PE y la forma de calificarlas para evaluar este criterio.

- **PE5.** ¿El/los métodos de investigación utilizados se corresponden con el propósito del estudio? [9], [12], [13]
 - El/los métodos y el propósito del estudio son descritos y relacionados de forma explícita, además tienen coherencia entre sí. Definiendo cuatro tipos de investigaciones [15]: (1)
 - *Exploratoria:* Su propósito es encontrar qué sucede y obtener ideas e hipótesis para nuevas investigaciones.

- *Descriptiva*: Su propósito es detallar el estado actual de una situación o fenómeno.
 - *Explicativa*: Su propósito es encontrar la causa o razón de una situación o problema.
 - *Mejora*: Su propósito es mejorar algún aspecto del fenómeno estudiado.
 - El/los métodos y el propósito del estudio tienen coherencia entre sí, pero alguno de ellos o su relación es ambiguo o se pueden extraer de forma implícita. (0.5)
 - El/los métodos y el propósito del estudio no tienen coherencia entre sí o alguno de ellos y su relación no son descritos de alguna forma. (0)
- **PE6.** ¿El/los procesos de recopilación de datos son acordes con la metodología utilizada? [9], [13]
 - El/los procesos de recopilación utilizados son descritos y se apegan a la metodología. (1)
 - El/los procesos de recopilación utilizados son acordes con la metodología, pero son ambiguos o se pueden extraer de forma implícita, o son descritos, pero no son del todo correctos. (0.5)
 - El/los procesos de recopilación utilizados no son descritos de alguna forma o no son acordes con la metodología. (0)
 - **PE7.** ¿El/los procesos de análisis de datos son acordes con la metodología utilizada? [9]
 - El/los procesos de análisis utilizados son descritos y se apegan a la metodología. (1)
 - El/los procesos de análisis utilizados son acordes a la metodología, pero son ambiguos o se pueden extraer de forma implícita, o son descritos, pero no son del todo correctos. (0.5)
 - El/los procesos de análisis utilizados no son descritos de alguna forma o no son acordes con la metodología. (0)

Credibilidad

Evalúa que los hallazgos y conclusiones sean válidos, efectivos y significativos [9], [10]. Es decir, que los hallazgos sean transmitidos correctamente, en especial a otros investigadores [10].

A continuación se plantean las PE y la forma de calificarlas para evaluar este criterio.

- **PE8.** ¿Existe una descripción clara de los hallazgos? [9], [12]
 - Los hallazgos son presentados, descritos de forma explícita y justificados por los resultados. (1)
 - Los hallazgos son ambiguos, se pueden extraer de forma implícita o no son justificados por los resultados. (0.5)
 - Los hallazgos o resultados no son presentados de ninguna forma. (0)
- **PE9.** ¿Discuten los investigadores las limitaciones, amenazas y/o sesgos con la validez de los hallazgos? [10], [11]
 - Las limitaciones, amenazas y/o sesgos con la validez de los hallazgos son presentados y descritos de forma explícita. (1)
 - Las limitaciones, amenazas y/o sesgos con la validez de los hallazgos son ambiguos o se pueden extraer de forma implícita. (0.5)
 - Las limitaciones, amenazas y/o sesgos con la validez de los hallazgos no son presentados de ninguna forma. (0)

Relevancia

Evaluar qué tanta importancia y repercusión tienen los hallazgos en la comunidad académica y en el entorno industrial [9], [11].

A continuación se plantean las PE y la forma de calificarlas para evaluar este criterio:

- **PE10.** ¿El estudio menciona la utilidad de los hallazgos para la industria y la comunidad investigadora? [10]
 - El impacto del estudio en la industria y comunidad investigadora son presentados, descritos y justificados de forma explícita. (1)
 - El impacto del estudio en la industria o comunidad investigadora son ambiguos, no justificados o se pueden extraer de forma implícita. (0.5)
 - El impacto del estudio en la industria o comunidad investigadora no son presentados de ninguna forma. (0)

- **PE11.** ¿El estudio identifica nuevas áreas en las que es necesaria la investigación? [12]
 - Las áreas o trabajo para continuar la investigación en un futuro son presentadas, descritas y justificadas de forma explícita. (1)
 - Las áreas o trabajo para continuar la investigación en un futuro son ambiguas, no justificadas o se pueden extraer de forma implícita. (0.5)
 - Las áreas o trabajo para continuar la investigación en un futuro no son presentadas de ninguna forma. (0)

3.1.2. CALIFICACIÓN DE LOS ESTUDIOS

Con base en lo realizado por [11] y [12], con la puntuación obtenida de cada PE se obtendrá una calificación por cada estudio y criterio.

Calificación por estudio

La calificación por estudio es útil para poder definir qué estudio es más relevante y otorgarle prioridad en sus hallazgos y tiempo de lectura sobre otros. Se sumarán los valores otorgados a cada una de las 11 preguntas por cada estudio para así obtener su calificación, que se consigue con la siguiente fórmula:

$$\text{Calificación del estudio} = \sum_{i=1}^n v_i$$

$$n = \text{Número total de preguntas}$$

$$v_i = \text{Valor otorgado en la pregunta}$$

Dada la calificación obtenida, el estudio es clasificado de acuerdo con la tabla 1.

Tabla 1. Calificación por estudio

Calidad	Rango
Excelente	$x \geq 9$
Buena	$7 \leq x < 9$
Satisfactoria	$5 \leq x < 7$
No satisfactoria	$x < 5$

Cabe resaltar que un estudio de calidad “No satisfactoria” no se descartará o no se tomará en cuenta. Simplemente, como se explicó anteriormente, se dará prioridad a los de mejor calidad, ya que la información de estos puede ser relevante a pesar de su calidad.

Calificación por criterio

La calificación por criterio es útil para poder saber la calidad en general de los estudios en el área estudiada y qué criterios pueden o no necesitar más atención cuando se realicen estudios empíricos. Se sumarán todos los valores dados a cada pregunta relacionada con el criterio a evaluar por cada estudio primario, para así obtener la calificación de dicho criterio. Se ha definido la siguiente fórmula para obtener la calificación:

$$\text{Calificación del estudio} = \left(\frac{\sum_1^n v_p + \dots + \sum_1^n v_p}{n * x} \right) * 100$$

n = Número total de estudios

v_p = Valor otorgado en la pregunta ligada al criterio. Donde $p = 1, 2, \dots, 11$

x = Número de preguntas ligadas al criterio

Dada la calificación obtenida, el criterio es clasificado de acuerdo con la tabla 2.

Tabla 2. Calificación por criterio

Calidad	Rango
Excelente	$x \geq 90 \%$
Pertinente	$70 \% \leq x < 90 \%$
Aceptable	$50 \% \leq x < 70 \%$
Mejorable	$x < 50 \%$

3.1.3. RESÚMENES ESTRUCTURADOS

Se decidió realizar resúmenes estructurados de cada uno de los estudios seleccionados utilizando como guía lo realizado en [14] porque usualmente la información de los resúmenes en cada estudio se encuentra incompleta, lo que obliga a leer otras partes del artículo para saber si es relevante o no [14]. Además, esto servirá de apoyo para poder contestar a las PE definidas anteriormente y a los lectores para poder comprender las características de cada uno de los estudios seleccionados sin necesidad de tener que ir a la fuente original. En dichos resúmenes estructurados se sintetizan aspectos clave del estudio utilizando títulos, tales como: Contexto, Objetivo, Método, Resultados y Conclusiones [14]. Además, se decidió incluir un identificador (EP# = estudio primario #), título y cita.

3.2. EJECUCIÓN

Después de ejecutar la búsqueda y aplicar los criterios de inclusión y exclusión se obtuvieron 72 estudios primarios. Se aplicó el instrumento de evaluación explicado

previamente a cada uno de los estudios primarios seleccionados con ayuda de los resúmenes estructurados para responder a cada una de las PE. A continuación se muestra a modo de ejemplo el resumen estructurado del primer estudio primario seleccionado:

EP01 – An Exploratory Study on the Influence of Developers in Technical Debt – [16]

Antecedentes: Los sistemas de *software* suelen ser realizados por muchos desarrolladores que tienen una gama variable de habilidades y hábitos. Estos tienen un gran impacto en la calidad del *software*. **Objetivo:** Explorar cómo los diferentes desarrolladores y las características de estos influyen en el sistema de *software* con respecto a su calidad y mantenibilidad utilizando DT como medida. **Método:** Se realiza un análisis a gran escala de 19,088 confirmaciones de 38 sistemas Java Apache y múltiples pruebas de análisis estadístico. **Resultados:** Los desarrolladores aumentan y disminuyen DT de manera desigual. La antigüedad del desarrollador en un sistema de *software* y la frecuencia de las confirmaciones se correlacionan negativamente con la DT agregada por este. Un intervalo de confirmación del programador tiene una correlación positiva con la DT agregada. **Conclusiones:** Como limitaciones se encuentran la precisión de la medición de la DT, métricas adecuadas, generalización de otros lenguajes de programación, sesgo hacia sistemas bien conocidos y sesgo hacia sistemas de código abierto. En el futuro se ampliará el estudio para comprender las características de los desarrolladores y cómo aumenta y disminuye la DT, e incluir un mayor número de sistemas de **código abierto y cerrado, más lenguajes de programación y más dominios**.

El resto de los resúmenes estructurados se encuentra en el siguiente enlace:

Resúmenes Estructurados de los Estudios Seleccionados

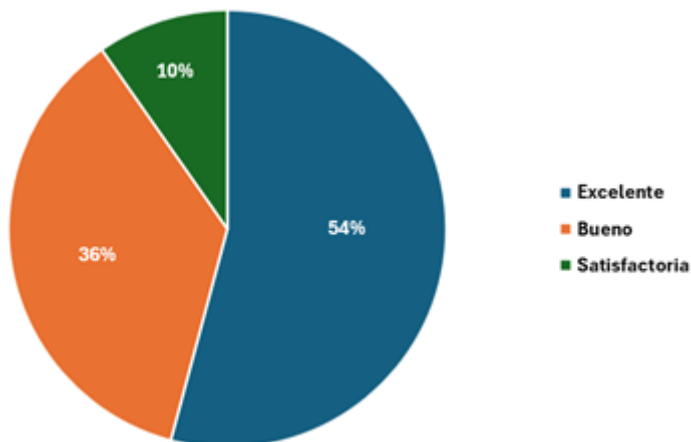
Ahora, para cada uno de los 72 estudios seleccionados se obtuvieron sus puntuaciones, que se pueden ver en:

Calificaciones de los Estudios Seleccionados

3.2.1. EVALUACIÓN POR ESTUDIO

A cada estudio seleccionado se le otorgó una calificación según la tabla 1. En la figura 1 se puede notar que la mayoría de los estudios primarios seleccionados tienen una calidad excelente, lo que representa un poco más de la mitad (39); Por último, las categorías “buena” y “satisfactoria” tienen 26 y 7 estudios, respectivamente.

Figura 1. Distribución de las calificaciones por estudio primario



Estos resultados nos muestran que todos los estudios tienen por lo menos una calidad mínima, ya que no hay ninguno de calidad “no satisfactoria”. Estos son buenos resultados, sobre todo porque predominan por mucho los de calidad “excelente” y “buena” sobre los de “satisfactoria”.

3.2.2. EVALUACIÓN POR CRITERIO

Calidad del informe

Se reemplazan los valores en la fórmula de la siguiente manera:

$$\text{Calidad del informe} = \left(\frac{\sum_1^{72} v_1 + \sum_1^{72} v_2 + \sum_1^{72} v_3 + \sum_1^{72} v_4}{72 * 4} \right) * 100$$

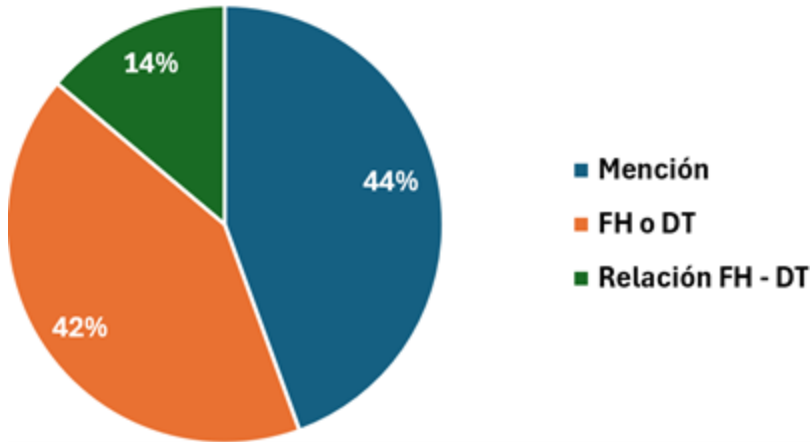
$$\text{Calidad del informe} = \left(\frac{60.5 + 68.5 + 25 + 63}{72 * 4} \right) * 100$$

$$\text{Calidad del informe} = 75.347 \%$$

Según la tabla 2, este criterio tiene una calidad pertinente. Este resultado se puede deber a los valores obtenidos en la mayoría de los estudios para la PE3. En esta se obtuvo una puntuación baja de 25 puntos de 72, lo que en promedio es 0.347. Esto quiere decir que la mayoría de los estudios no se enfocan en su totalidad en la relación entre FH y DT. Los estudios en su mayoría cumplen con un objetivo, contexto, y tablas y/o figuras, claras y bien descritas.

Además, lo dicho anteriormente sobre la PE3 se puede observar más claramente en la figura 2. Los estudios que se enfocan específicamente en la relación entre FH y DT son el grupo más pequeño (10). Sin embargo, los que solo lo mencionan son la mayoría (32). Por último, los que se enfocan ya sea solo en FH o DT cuentan con 30 estudios. Esto es importante y se hace énfasis porque se concluye que hacen falta más esfuerzos en cuanto al número de estudios que investiguen únicamente esta relación y es una buena área de oportunidad para trabajar sobre ella.

Figura 2. Distribución de la PE3



Rigor

Se reemplazan los valores en la fórmula de la siguiente manera:

$$Rigor = \left(\frac{\sum_1^{72} v_5 + \sum_1^{72} v_6 + \sum_1^{72} v_7}{72 * 4} \right) * 100$$

$$Rigor = \left(\frac{60 + 65 + 61.5}{72 * 3} \right) * 100$$

$$Rigor = 86.342 \%$$

Según la tabla 2, este criterio tiene una calidad pertinente. Se obtuvieron buenas puntuaciones para cada pregunta que evalúa este criterio. Esto significa que en general los estudios seleccionados eligieron correctamente su método de investigación de acuerdo con su objetivo, y los métodos de recolección y análisis de datos también fueron elegidos correctamente según el método de investigación. Sin embargo, algunos fueron descritos ambiguamente o no fueron suficientemente descriptivos.

Credibilidad

Se reemplazan los valores en la fórmula de la siguiente manera:

$$Credibilidad = \left(\frac{\sum_1^{72} v_8 + \sum_1^{72} v_9}{72 * 2} \right) * 100$$

$$Credibilidad = \left(\frac{60 + 66}{72 * 2} \right) * 100$$

$$Credibilidad = 87.5 \%$$

Según la tabla 2, este criterio tiene una calidad pertinente. Se obtuvieron buenas puntuaciones para las preguntas que responden este criterio. Esto quiere decir que en general presentan los hallazgos, al igual que las limitaciones, amenazas y/o sesgos a la validez.

Relevancia

Se reemplazan los valores en la fórmula de la siguiente manera:

$$Relevancia = \left(\frac{\sum_1^{72} v_{10} + \sum_1^{72} v_{11}}{72 * 2} \right) * 100$$

$$Relevancia = \left(\frac{36 + 52.5}{72 * 2} \right) * 100$$

$$Relevancia = 61.458 \%$$

Según la tabla 2, este criterio tiene una calidad aceptable. Se tiene un valor bajo para la PE10 que evalúa si los estudios mencionan de manera correcta la utilidad de los hallazgos en la industria y la comunidad investigadora, ya que ella obtuvo en total un promedio de 0.5. Esto significa que la mitad de los estudios presenta de manera ambigua este aspecto, no lo describe y/o no lo justifica de manera pertinente. Por otra parte, en general, ellos mencionan de alguna manera las áreas o trabajo futuro en las que se puede seguir investigando o puede que no esté del todo bien justificada.

3.2.3. SÍNTESIS DE LOS RESULTADOS

Como se mencionó anteriormente en la planeación, al momento de leer a fondo los estudios primarios seleccionados, se les proporcionó más atención y tiempo de lectura a los estudios que obtuvieron mejores calificaciones en la calidad, que a los que no.

Por otro lado, se encontraron estudios que contradecían los hallazgos de otros, por lo que fue necesario utilizar lo obtenido en la evaluación de la calidad para poder dar una aproximación sobre qué hallazgos son más confiables. A continuación se muestra un ejemplo:

En EP03 se menciona que la reputación, es decir, los Incentivos, no está tan relacionada con la DT. Sin embargo, este estudio tiene una calificación de 7.5 (Bueno) en la calidad, mientras que tres estudios dicen lo contrario. Además, dos de los tres estudios, es decir, EP57 y EP68, que están a favor de que este factor incide en la aparición de DT tienen una calidad Excelente, por lo que es más confiable esperar que este factor esté relacionado. ([8], p. 133)

4. CONCLUSIONES

Se presenta de inicio a fin cómo se implementó la evaluación de la calidad de los estudios primarios seleccionados en una revisión sistemática de literatura en la que se muestra el instrumento utilizado, y todas las adecuaciones y modificaciones que se le realizaron a este para ejecutarlo de la mejor manera y pudiera encajar con los objetivos de la revisión.

Finalmente, como se muestra a lo largo de este escrito, la evaluación de la calidad es una tarea muy importante; permite conocer en general si los estudios recopilados sobre un área en específico son de calidad o no, y a su vez encontrar las áreas débiles o de oportunidad de los estudios, basándose en los criterios de calidad. Esto, a su vez, ayuda a informar y orientar el área de investigación sobre a qué aspectos se debe prestar más atención al realizar un estudio primario, para así mejorar tanto la calidad de los estudios como el área en general. Por último, permite asignar una calificación a cada estudio seleccionado, lo que facilita la distinción entre los escogidos y ayuda a priorizar la atención, lectura y confianza en sus hallazgos, con lo que se contribuye a la **síntesis de la revisión de literatura**.

REFERENCIAS

- [1] R. Pressman, Ingeniería del software: Un enfoque práctico, McGraw-Hill Interamericana Editores, S.A. DE C.V., 2010.
- [2] I. Sommerville, Ingeniería del software, Madrid: Pearson Educación, S.A., 2005.
- [3] H. v. Vliet, Software engineering: principles and practice, J. Wiley & Sons, Ltd, 2008.
- [4] R. Malhotra, Empirical research in software engineering. Concepts, analysis, and applications, Nueva York: CRC Press, 2016.

- [5] M. Felderer, & G. H. Travassos, *Contemporary Empirical Methods in Software Engineering*, Heidelberg: Springer Cham, 2020.
- [6] B. A. Kitchenham, D. Budgen, & P. Brereton, *Evidence-Based Software Engineering And Systematic Reviews*, CRC Press, 2016.
- [7] B. Kitchenham, & S. Charters, *Guidelines for performing Systematic Literature Reviews in Software Engineering*, 2007.
- [8] J. Á. Ruiz Escalante, *Factores Humanos y su Incidencia en la Gestión de Deuda Técnica*, Mérida: Universidad Autónoma de Yucatán, 2024.
- [9] T. Dybå, & T. Dingsøyr, «Strength of evidence in systematic reviews in software engineering,» de *Proceedings of the Second ACM-IEEE International Symposium on Empirical Software Engineering and Measurement*, 2008.
- [10] L. Yang, H. Zhang, H. Shen, X. Huang, X. Zhou, G. Rong, & D. Shao, «Quality Assessment in Systematic Literature Reviews: A Software Engineering Perspective,» *Information and Software Technology*, 2021.
- [11] R. A. Santos Charruf, *Revisión Sistemática de Literatura sobre Factores sociales que inciden en la Priorización de Requisitos*, Mérida, Yucatán: Universidad Autónoma de Yucatán (UADY), 2022.
- [12] J. d. I. R. Baeza Pérez, & R. Aguilar Vera, «Quality Assessment for Selected Primary Studies in a Systematic Literature Review: A Case Study,» de *2022 IEEE Mexican International Conference on Computer Science (ENC)*, Xalapa, 2022.
- [13] N. García Sánchez, *Diseño de Bases de Datos en la Nube: Una Revisión Sistemática de Literatura*, Mérida, Yucatán: Universidad Autónoma de Yucatán (UADY), 2022.
- [14] D. Budgen, B. Kitchenham, S. Charters, M. Turner, P. Brereton, & S. Linkman, «Presenting software engineering results using structured abstracts: a randomised experiment,» *Empirical Software Engineering*, p. 435–468, 2008.
- [15] M. Genero Bocco, J. Cruz-Lemus y M. Piattini Velthuis, *Métodos de investigación en ingeniería del software*, Ediciones de la U, 2015.
- [16] R. Alfayez, P. Behnamghader, K. Srisopha, & B. W. Boehm, «An exploratory study on the influence of developers in technical debt,» de *TechDebt '18: Proceedings of the 2018 International Conference on Technical Debt*, 2018.



Open Access: This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which allows non-commercial use, distribution, and reproduction with proper credit to the author and source, without modifications. Third-party material is included under the same license unless otherwise stated. For uses beyond this license, permission must be obtained from the copyright holder.

© The Author(s) 2025 - <http://creativecommons.org/licenses/by-nc-nd/4.0/>