REVISTA PERSPECTIVAS

VOL. 8, N°1 / ENERO - JUNIO 2026 / e-ISSN: 266-6688
DOI: https://doi.org/10.47187/perspectivas.8.1.251

VERIFICACIÓN DE CALIDAD EN LAS HISTORIAS DE USUARIO: UN MAPEO SISTEMÁTICO

Quality Verification in User Stories: A Systematic Mapping

Leonardo Espinosa Arévalo	A24216469@alumnos.uady.mx
Antonio Armando Aguileta Güemez	aaguilet@correo.uady.mx
Raúl Antonio Aguilar Vera	avera@correo.uady.mx

[1] Facultad de Matemáticas, Universidad Autónoma de Yucatán Mérida, Yucatán, México.

RESUMEN

El desarrollo ágil ha consolidado el uso de historias de usuario como una práctica central para la recolección de requisitos, lo cual ha motivado un creciente interés por asegurar su calidad. El presente artículo reporta un estudio de mapeo sistemático cuyo objetivo es identificar los criterios de calidad utilizados en su verificación, analizando su frecuencia temporal, distribución geográfica y marcos conceptuales aplicados. Para la recopilación de estudios se siguieron metodológicas para el desarrollo de estudios de mapeo y revisiones sistemáticas. Se seleccionaron 58 estudios primarios publicados entre 2004 y 2024. Los marcos INVEST y QUS son los más referenciados, destacando la inambigüedad como el criterio más citado, aunque no el único que debe ser abordado. Además, se identificó una concentración de investigaciones en países tecnológicamente desarrollados y un incremento notable en 2024. El estudio confirma el crecimiento sostenido del interés académico en la verificación de la calidad de las historias de usuario.

Palabras Clave: Criterios de Calidad, Historias de Usuario, Ingeniería de Software, Mapeo Sistemático.

ABSTRACT

Agile development has established user stories as a fundamental practice for requirements elicitation, leading to increasing academic interest in assessing their quality. This paper presents a systematic mapping study aimed at identifying the quality criteria applied in user story verification, examining their temporal evolution, geographical distribution, and underlying conceptual frameworks. Established methodological guidelines for systematic mapping studies and systematic literature reviews were followed during the study selection process. A total of 58 primary studies published between 2004 and 2024 were analyzed. The INVEST and QUS frameworks emerged as the most frequently referenced, with unambiguity identified as the most commonly cited quality criterion, although it is not the sole aspect that should be considered. Furthermore, the results reveal a concentration of research efforts in technologically advanced countries and a significant increase in publications in 2024. These findings confirm a sustained growth in scholarly interest in the verification of user story quality.

Keywords: Quality Criteria, User Stories, Software Engineering, Systematic Mapping

Licencia: © 2026. Este artículo se publica bajo la licencia **Creative Commons Atribución 4.0 Internacional (CC BY 4.0)**.
https://creativecommons.org/licenses/by/4.0/

Fecha de Recepción: 26/11/2025. **Fecha de Aceptación:** 02/01/2026. **Fecha de Publicación:** 28/01/2026.

I. Introducción

El desarrollo de software implica un enfoque sistemático y disciplinado [1]. Esto quiere decir que los equipos de desarrollo se rigen por metodologías orientadas a la construcción de software de alta calidad. El modelo en cascada es considerado el enfoque clásico dentro del desarrollo de software [2]. La principal limitación de modelos como el modelo en cascada y otras metodologías tradicionales es su baja capacidad de respuesta ante cambios en los requisitos y en su implementación. La solución a esta problemática surgió con el desarrollo ágil de software, cuyo proceso se basa en la iteración y en las entregas continuas de valor.

El nacimiento de metodologías ágiles como SCRUM, Kanban o Extreme Programming (XP) prioriza aspectos distintos a los modelos tradicionales, como la iteración y las entregas continuas, el fortalecimiento de la comunicación tanto dentro del equipo como con el cliente, y el mantenimiento de una documentación menos exhaustiva [3]. Las metodologías ágiles son marcos de desarrollo que respaldan la filosofía propuesta en el Manifiesto Ágil [4]. Estas metodologías incorporan prácticas individuales que, con el tiempo, han evolucionado: algunas que en un principio pertenecían a un único modelo se han convertido en referentes dentro de la filosofía ágil [5]. Tal es el caso de las Historias de Usuario (HU).

En las fases iniciales del desarrollo ágil se lleva a cabo la elicitación de requisitos. Una historia de usuario (HU) es una representación concisa de una funcionalidad desde la perspectiva del rol que la solicita, estructurada y redactada en lenguaje natural [6]. Las HU se han incorporado de manera amplia en los flujos de trabajo de diversos marcos ágiles, convirtiéndose en una práctica que supera los límites de la metodología en la que surgieron. Actualmente, constituyen la técnica de elicitación de requisitos más utilizada en el desarrollo ágil de software [7].

Las HU cuentan, desde su popularización, con criterios establecidos para la evaluación de su calidad. Estos criterios se sintetizaron en el acrónimo INVEST [6]. Además de estos lineamientos, conceptos como la ambigüedad y la coherencia se han incorporado a la discusión en torno a la validación de la calidad de las HU. Asimismo, la necesidad de definir métricas y estándares formales se consolidó en lo que hoy se conoce como Software Quality Assurance (SQA), cuyas responsabilidades incluyen la revisión exhaustiva de los planes de desarrollo y de los procesos de calidad [8]. La ausencia de mecanismos claros para abordar errores críticos o de criterios específicos orientados a las HU motivó la creación de un marco de calidad complementario, Quality User Story (QUS) [9]; sin embargo, persiste la necesidad de determinar si los criterios propuestos son efectivamente los más adecuados.

El presente mapeo sistemático es un estudio secundario (ES) que busca identificar las tendencias de investigación relacionadas con los criterios empleados para la verificación de la calidad de las HU. En particular, pretende identificar si los criterios definidos por los marcos de calidad son utilizados en la literatura científica. Este análisis se realiza desde un enfoque cuantitativo.

La siguiente sección ofrece una visión de los trabajos previos y explica cómo se diferencia del presente estudio. Posteriormente, la sección 3 presenta una descripción breve de la metodología para el desarrollo de estudios secundarios utilizada como referencia. La sección 4 describe las tareas principales de la fase de planeación, incluyendo la formulación de preguntas de investigación, los criterios de selección considerados, la cadena de búsqueda elaborada y las bases de datos seleccionadas para la búsqueda automatizada. La sección 5 resume las actividades realizadas durante la fase de ejecución y detalla el número total de estudios primarios seleccionados. En la sección 6 se describen los resultados obtenidos, y finalmente, la sección 7 de presentan las conclusiones generales y las líneas de trabajo futuro.

Para comprender cómo se ha verificado la calidad de las historias de usuario (HU)—es decir, qué criterios se utilizan en su evaluación—

resulta necesario examinar también los errores comúnmente asociados a su formulación en esta práctica de elicitación propia del desarrollo ágil. Con el propósito de valorar la viabilidad de este estudio, se decidió analizar trabajos secundarios previos a fin de identificar hasta dónde se ha investigado este tema y cuáles son las conclusiones a las que se ha llegado.

En la literatura científica, algunos estudios se han enfocado en la problemática de la verificación de la calidad de las HU, empleando diversos enfoques y aproximaciones metodológicas:

Un ejemplo es un ES que analiza, con el objetivo de identificar, categorizar y examinar, la investigación empírica relacionada con la calidad de los requisitos en el ámbito de la ingeniería de software [10]. Si bien no aborda explícitamente las HU (una forma de elicitación de requisitos funcionales), sí examina aspectos afines, como el estudio de los requisitos funcionales y la búsqueda de soluciones a problemas recurrentes, entre ellos la falta de estandarización y la presencia de errores frecuentes como la ambigüedad o la incompletitud.

Otro estudio secundario relevante es el presentado en [11], el cual se centra en investigaciones específicamente relacionadas con las HU. Si bien proporciona una visión clara del incremento reciente de la investigación en este ámbito, su alcance permanece en un nivel descriptivo y no profundiza en aspectos específicos de los procesos de verificación o revisión de las HU. En consecuencia, su contenido se limita a responder preguntas generales, tales como: *"¿Qué áreas de investigación relacionadas con las historias de usuario existen?", "¿Qué tipos de problemas se han identificado?" y "¿Qué tipos de resultados se han reportado?"*, entre otras orientadas al tipo de investigación.

Dentro de los estudios secundarios relacionados con las HU y su calidad, se encuentra también el estudio *Ambiguity in User Stories: A Systematic Literature Review* [12], que puede considerarse una continuación del trabajo previo desarrollado por Amna y Poels. Este estudio se centra en un problema recurrente en las HU: la ambigüedad. Sin embargo, este no es el único error que puede surgir durante la formulación de las HU, aunque sí constituye un hallazgo recurrente en los estudios secundarios existentes.

El último elemento que motiva este trabajo es el ES presentado en [13], el cual recopila, hasta 2023, los enfoques existentes sobre la calidad de las HU. Este estudio identifica diversos métodos, técnicas y tecnologías, y destaca la ambigüedad como uno de los principales factores de riesgo en la formulación de las HU.

En conjunto, estos estudios secundarios contribuyen a acotar el campo de investigación y proporcionan respuestas específicas a diversas preguntas relevantes. Asimismo, permiten identificar las bases de datos (BD) empleadas en cada uno de ellos, entre las que se incluyen **Scopus, ScienceDirect, SpringerLink, IEEE Xplore, ACM Digital Library y Google Scholar**.

El presente trabajo propone un nuevo ES motivado por la necesidad de identificar los criterios de calidad empleados en la investigación para evaluar las HU, así como comprender su relación con los errores más comunes reportados a lo largo de su desarrollo y uso. A diferencia de revisiones previas, este trabajo adopta un enfoque más amplio, evitando concentrarse únicamente en la ambigüedad como problema central [12]. La intención es ofrecer un panorama actualizado sobre los criterios de calidad y las tendencias de investigación, desde la concepción de las HU hasta los avances más recientes en sus técnicas de verificación. Como objetivo adicional, se busca analizar la evolución temporal de las publicaciones y los países en los que se concentran dichas investigaciones.

Finalmente, este mapeo se justifica también por la necesidad de ampliar los rangos temporales considerados: aunque varias de las revisiones existentes son relativamente recientes, la mayoría no incorpora el periodo posterior a la década de 2020, caracterizada por una acelerada transformación digital y avances significativos en inteligencia artificial. En este contexto, resulta

pertinente identificar qué criterios de calidad han cobrado mayor relevancia en la literatura contemporánea.

II. Metodología

El desarrollo del presente mapeo sistemático se basa en el marco metodológico propuesto en [14] y actualizado en [15]. Para la conducción de la investigación y la presentación del estudio, y considerando que un mapeo sistemático puede funcionar como un complemento de una revisión sistemática, se siguieron los pasos establecidos por Kitchenham [16]. El trabajo se estructuró en tres fases principales: una fase de planeación, una fase de ejecución de lo establecido y una fase final dedicada al reporte de los resultados y a la descripción de los hallazgos. En este estudio se amplía particularmente el contenido de la fase de reporte, la cual se integrará con mayor detalle en una versión extendida futura del mapeo.

A. **Planeación**

Esta fase comprende varias actividades, entre las que se encuentran: (i) la definición de las preguntas de investigación, basadas en el enfoque PICOC [15] e informadas por vacíos identificados en estudios previos; (ii) la selección de las bases de datos (BD) pertinentes al dominio de la ingeniería de software (IS); (iii) la definición de la cadena de búsqueda; y (iv) la formulación de los criterios empleados para incluir o excluir trabajos de investigación. A continuación, se describen los detalles de cada una de estas actividades.

1. Preguntas de investigación

El tema del presente ES se sitúa en la elicitación de requisitos dentro de las metodologías ágiles, con énfasis en las HU como mecanismo para la captura de requisitos funcionales. El estudio busca determinar los criterios de calidad empleados en la literatura y analizar cómo se verifican, así como examinar la distribución temporal y geográfica de las aportaciones científicas en este campo. Con este fin, se formularon las siguientes preguntas de investigación:

- *PI1. ¿Cuál es la distribución en el tiempo de las investigaciones publicadas sobre verificación de calidad de historias de usuario?*
- *PI2. ¿Cuáles son los países que tienen investigación sobre verificación de la calidad de historias de usuario?*
- *PI3. ¿Cuáles son los criterios de calidad reportados en la verificación de historias de usuario?*

Las dos primeras preguntas de investigación se orientan a analizar la distribución temporal y geográfica de la investigación sobre las HU, con el fin de identificar tendencias a lo largo del tiempo y determinar los países en los que se concentra esta línea de estudio. En este trabajo, la verificación se entiende como el proceso de evaluar si las HU cumplen con criterios de calidad establecidos.

En relación con los criterios de calidad, se propone clasificarlos según los modelos y estándares existentes, y analizar de manera individual los atributos de calidad para examinar la distribución de criterios fuera de los marcos formales y determinar cuáles se emplean con mayor frecuencia en la literatura reciente.

El enfoque PICOC orientó la formulación de las preguntas de investigación, así como el diseño de la cadena de búsqueda y de los criterios de inclusión y exclusión. En particular, para delimitar el alcance temático del estudio se emplearon las dimensiones P (**P**opulation), en este caso las historias de usuario; e I (**I**ntervention), procesos y herramientas de verificación.

2. Bases de datos

Para llevar a cabo un estudio secundario completo en el ámbito de la Ingeniería de Software (IS), resulta fundamental seleccionar fuentes de información que aseguren la disponibilidad de investigaciones pertinentes. En consecuencia, se eligieron BD bibliográficas como IEEE Xplore, ACM Digital Library, SpringerLink y ScienceDirect, debido a su reconocida producción científica en el campo de las ciencias de la computación. Asimismo, se incorporó el uso complementario del repositorio

Google Scholar, el cual permite recuperar trabajos que, aun cuando no se encuentren indexados en las bases de datos anteriores, conservan la relevancia científica requerida. Estas bases y repositorio fueron seleccionados por su uso recurrente en estudios secundarios previos en áreas relacionadas con la computación y la ingeniería de software, tal como se ha documentado en trabajos previos [11,12].

Estas BD contienen artículos revisados por pares, actas de congresos y revistas especializadas que abarcan áreas fundamentales de la IS, desde metodologías de desarrollo hasta tecnologías emergentes. Tal es el caso de IEEE Xplore, plataforma que desde hace décadas alberga e indexa una amplia variedad de investigaciones en ingeniería de software, incluidos documentos normativos ampliamente citados, como el *IEEE Standard Glossary of Software Engineering Terminology (1990)* y el estándar *IEEE 830-1998*. De manera similar, ACM Digital Library reúne conferencias y artículos publicados en revistas de referencia en el área, entre ellas *Foundations of Software Engineering (FSE)*. Algo comparable ocurre con SpringerLink, plataforma de *Springer Nature* que publica libros, volúmenes de conferencias y revistas especializadas en ingeniería de software, inteligencia artificial aplicada al desarrollo de software y metodologías de verificación, como *Software Quality Journal*. Adicionalmente, se integró Google Scholar como repositorio complementario con el fin de ampliar el alcance de la búsqueda mediante la indexación de literatura proveniente de diversas fuentes académicas y el uso de herramientas avanzadas para el filtrado de información.

3. Cadena de búsqueda

La cadena de búsqueda genérica utilizada para la identificación de los estudios primarios fue la siguiente:

("User Story" OR "User Stories") AND ("Agile" OR "Scrum" OR "Extreme Programming" OR "XP") AND ("Validation" OR "Verification" OR "Quality Assurance" OR "Evaluation").

Las HU constituyen el primer bloque de la cadena, dado que su presencia explícita en los estudios recuperados es un requisito fundamental. No obstante, el término HU puede presentar ambigüedades, ya que, pese a referirse a una práctica ampliamente reconocida dentro de las metodologías ágiles, también puede arrojar resultados ajenos al ámbito de la IS. El segundo bloque delimita el alcance temático hacia los marcos de desarrollo ágil, incorporando metodologías ampliamente implementadas y, adicionalmente, el término "*Agile*" con el propósito de identificar trabajos que no se encuentren asociados a los marcos más conocidos. El tercer bloque integra los conceptos vinculados con la verificación y la calidad de las HU. Es importante señalar que la cadena de búsqueda se configuró de acuerdo con la sintaxis requerida en cada una de las fuentes de información consultadas.

4. Criterios de inclusión y exclusión

Los criterios de inclusión (CI) y exclusión (CE) establecen las condiciones que determinan qué trabajos serán considerados en el análisis, asegurando su alineación con los objetivos de la investigación. En este estudio, dichos criterios se diseñaron para identificar trabajos relevantes en la verificación de la calidad de las HU, atendiendo a aspectos como el periodo de publicación, la pertinencia temática y el tipo de estudio. De este modo, se conforma un corpus de estudios representativo, a la vez que se excluyen aquellos trabajos cuya falta de relación temática podría afectar la validez de los resultados. Se definieron cuatro criterios de inclusión y dos criterios de exclusión:

- CI1. Ventana de tiempo de 2004 a 2024: Este intervalo se seleccionó tomando como punto de partida el año 2004, correspondiente a la publicación del primer libro que aborda de manera sistemática las historias de usuario [6].
- CI2. Textos en inglés: Se seleccionaron únicamente estudios escritos en inglés con el fin de evitar ambigüedades derivadas del proceso de traducción.
- CI3. Relevancia respecto a las preguntas de investigación: Se consideraron únicamente

aquellos estudios que contribuyan directamente a responder las preguntas de investigación planteadas.

- CI4. Artículos provenientes de revistas de Investigación y Memorias de Conferencia.
- CE1. Artículos duplicados: Al existir un repositorio, pueden encontrarse artículos que también están en la base de datos específica.
- CE2. Resultados parciales o versiones preliminares: Se excluyen las versiones preliminares o parciales de un mismo estudio que presenten resultados no definitivos. En aquellos casos en los que un conjunto de autores haya publicado inicialmente un artículo con resultados parciales y, posteriormente, una versión más completa con análisis finales y conclusiones consolidadas, se seleccionará únicamente la versión final.

Estos criterios permiten seleccionar los estudios que abordan de manera directa las historias de usuario (la población) en el contexto de las metodologías ágiles, garantizando que el corpus resultante sea pertinente para los objetivos de la investigación. Asimismo, se excluyen aquellos trabajos cuya falta de relevancia o adecuación metodológica podría introducir sesgos o afectar la validez de los resultados que se obtengan del corpus.

B. Ejecución

El siguiente paso comprende la ejecución de la cadena de búsqueda en las distintas BD y en el repositorio consultado, seguida de la selección de los estudios primarios conforme a los criterios de inclusión y exclusión establecidos. Posteriormente, se realiza la evaluación de su relevancia para la investigación y, finalmente, el análisis y la síntesis de los estudios.

Tabla. I. *Configuración de la cadena de búsqueda en las distintas bases de datos.*

Bases de datos	Cadena de búsqueda
IEEE Xplorer, ACM Digital Library y SpringerLink	(“User story” OR “User stories”) AND ("Agile" OR "Scrum" OR "Extreme programming" OR "XP") AND (“Validation” OR “Verification” OR “Quality Assurance” OR "Evaluation")
ScienceDirect	(“User story” OR “User stories”) AND ("Agile" OR "Scrum" OR "XP") AND (“Validation” OR “Verification” OR “Quality Assurance” OR "Evaluation")
Google Scholar	(“User story” OR “User stories”) ("Agile" OR "Scrum" OR "Extreme programming" OR "XP") (“Validation” OR “Verification” OR “Quality Assurance” OR "Evaluation")

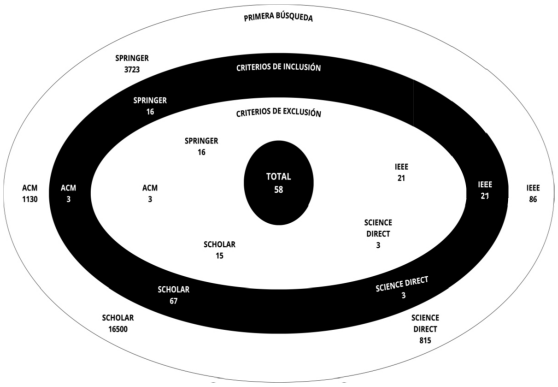


Fig. 1: *Resumen del proceso de selección (diseño de los autores).*

1. Ejecución de la cadena de búsqueda

La cadena de búsqueda se aplicó en las distintas BD y en el repositorio correspondiente, de acuerdo con la sintaxis y reglas de búsqueda establecidas por cada plataforma, los resultados se muestran en la Tabla I.

Dentro de la ejecución y configuración de la cadena de búsqueda, tres de las bases de datos utilizaron exactamente la misma configuración.

2. Ejecución de criterios de inclusión y exclusión

Para cumplir los criterios de inclusión, algunas bases de datos ofrecieron diversos filtros. Sin embargo, para el criterio CI3 fue necesario realizar una lectura breve de los resúmenes y, en caso de duda, una primera revisión de los trabajos completos. En cuanto a los criterios de exclusión, se requirió una lectura detallada de los artículos para identificar si se trataban de algún duplicado o de una versión más extensa de otro.

Finalmente, se obtuvieron un total de 109 trabajos al aplicar únicamente los criterios de inclusión y 58 trabajos en total una vez aplicados todos los criterios (ver Fig. 1).

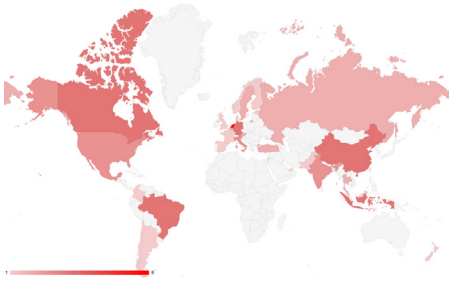


Fig. 2: *Mapa geográfico de trabajos.*

Los artículos seleccionados pueden consultarse en el enlace correspondiente:

https://docs.google.com/spreadsheets/d/1KOXqN89dlOT17RkEJ7Gm2kfWoEpdHAW5/edit?gid=288974753#gid=288974753

III. Results

Una vez analizados los 58 estudios seleccionados, se procedió a responder cada una de las tres preguntas de investigación. A continuación, se presenta una síntesis de los hallazgos.

P1. ¿Cuál es la distribución en el tiempo de las investigaciones publicadas sobre verificación de calidad de historias de usuario?

La ventana de tiempo se estableció desde el año más lejano posible, 2004. Sin embargo, la frecuencia de publicaciones comienza a partir de 2009. Es por ello por lo que en la Fig. 2 se presenta una ventana de tiempo de 2009 a 2024. Se observa que, si bien fue hasta 2009 cuando se encontró al menos un trabajo relacionado con la verificación de las HU, ha habido un aumento progresivo en este tema, lo que refleja que se trata de un tópico de interés en la actualidad, especialmente en el año más reciente, 2024. Se puede inferir que, aunque los criterios INVEST para las historias de usuario ya existían desde 2004 [6], la gráfica sugiere que los equipos actuales requieren comprobar las buenas prácticas en esta técnica de elicitación de requisitos dentro de metodologías ágiles.

Comparando con la literatura, los trabajos de verificación comienzan poco antes de la primera guía de Scrum [18], la cual es uno de los métodos con mayor implementación y uso de las HU [19]. Se observa también que, a partir de 2014, se empiezan a realizar más de dos estudios por año, con excepción de 2015. Las gráficas reportadas en [7] revelan que, para ese entonces (2014), las historias de usuario ya eran, con gran diferencia, el documento más usado en la lista de requisitos o product backlog de la metodología Scrum. El último año marcado, 2024, es el de mayor tendencia. Esto vuelve relevante esta investigación, ya que el límite de los anteriores ES llega hasta 2023, previo al pico más alto. La evolución del interés por garantizar la calidad en las HU y, en consecuencia, mejorar la calidad de los productos de software, es clara y muestra una tendencia ascendente.

P2. ¿Cuáles son los países que tienen investigación sobre verificación de la calidad de historias de usuario?

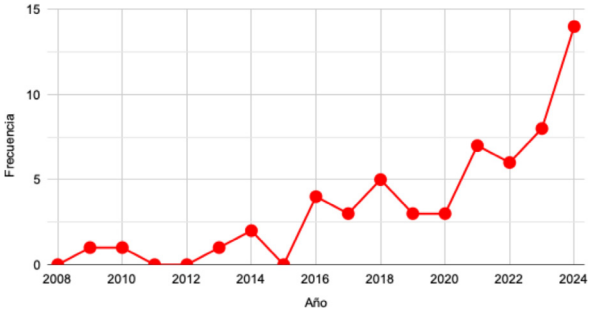


Fig. 3: *Frecuencia de trabajos publicados por año.*

La información referida a los aspectos geográficos (véase Fig. 3) muestra que los países con mayor desarrollo científico son también los principales exponentes. El país con más publicaciones es Países Bajos, con un total de 8 artículos. La tendencia que muestra este país puede verse influenciada por las investigaciones desarrolladas en el territorio, desde la propuesta en 2016 del marco QUS (Quality User Story) [9], el cual posteriormente sirvió de base para numerosos trabajos al establecer un marco de calidad ampliamente adoptado.

La mayor parte de la investigación proviene de países desarrollados, especialmente de Europa, América del Norte y Asia Oriental. Entre los territorios más representados se encuentran Canadá, Indonesia, China, Alemania y Brasil. Países Bajos es particularmente activo en este

ámbito, lo que puede influir en los enfoques metodológicos o temáticos predominantes, como se observa en sus aportaciones. Otros países como México, Estados Unidos, Italia e India presentan un número menor de publicaciones, aunque mantienen aportes relevantes que refuerzan los puntos clave en la literatura. Finalmente, regiones de Latinoamérica, con excepción de Brasil, como Colombia, Argentina y Chile, así como zonas de Europa y Asia como Rusia, España, Francia e Israel, cuentan con entre una y dos publicaciones al respecto. Por el contrario, zonas como África, Medio Oriente y regiones del Sudeste Asiático no aparecen como origen de estudios.

P3. ¿Cuáles son los criterios de calidad reportados en la verificación de historias de usuario?

Gran parte de los estudios realizados no se ajusta a un marco formal establecido y opta por proponer su propia lista de criterios a emplear. Sin embargo, aunque los criterios se presenten de manera individual o descriptiva, la mayoría de estos se encuentra dentro de los lineamientos establecidos en los marcos de mayor reconocimiento, como INVEST [6] y QUS [9]. Esto se refleja en la Fig. 4. Entre estos, el criterio con mayor relevancia en los estudios es la inambigüedad, ya que se busca que las HU sean claras y no estén sujetas a múltiples interpretaciones, como se menciona en estudios previos. Finalmente, todos los criterios que forman parte de estos dos marcos son los únicos cuya frecuencia supera los diez estudios que los referencian.

Estos criterios de calidad destacan por su utilidad e implementación adecuada ante problemas ligados a equipos de desarrollo ágil, donde las HU se aplican en las primeras etapas de desarrollo para la redacción de requisitos, en donde se vinculan con las secciones mencionadas en el trabajo de Lucassen [9], basadas en el proceso de semiótica para clasificar sus criterios: sintácticos, referidos a reglas para la correcta combinación de palabras; semánticos, relacionados con el significado de las palabras y oraciones; y pragmáticos, orientados al análisis de la interpretación a través del contexto [20].

Tabla. II. *Marco, categoría semió y agrupación por cada criterio*

Criterios	Marco de calidad	Semiótica	Individual o en set
Inambigüo	QUS	Semántica	Individual
Sentencia completa	QUS	Pragmática	Individual
Estimable	INVEST / QUS	Pragmática	Individual
Independiente	INVEST / QUS	Pragmática	En set
Bien formado	QUS	Sintáctica	Individual
Atómico	QUS	Sintáctica	Individual
Único	QUS	Pragmática	En set
Comprobable (Testeable)	INVEST	Pragmática	Individual
Libre de conflicto	QUS	Semántica	En set
Mínima	QUS	Sintáctica	Individual
Uniforme	QUS	Pragmática	En set
Completo	QUS	Pragmática	En set
Conceptualmente sólido	QUS	Pragmática	Individual
Orientado al problema	QUS	Semántica	Individual
Negociable	INVEST	Pragmática	Individual
Pequeño	INVEST	Sintáctica	Individual
Valuable	INVEST	Semántica	Individual

A partir de los datos obtenidos en la Fig. 4, se elaboró una tabla con los criterios más mencionados, en la cual se identificó a qué marco corresponden y se asignó la categoría semiótica a la que pertenecen, clasificando así los criterios del marco INVEST que no contaban con una categoría definida. Asimismo, se realizó la clasificación correspondiente para distinguir entre criterios de carácter individual y criterios en set (es decir, aquellos que requieren una serie de HU para poder evaluarse), concentrando toda esta información en la Tabla II.

El análisis de dicha tabla muestra una alta tendencia hacia criterios de tipo pragmático; en los equipos de desarrollo se valora especialmente la interpretación del signo (HU) en el contexto en el cual es aplicado, lo que indica que las HU pueden utilizarse en múltiples escenarios con contextos enriquecidos. Los criterios sintácticos y semánticos presentan una frecuencia de cuatro estudios cada uno; sin embargo, es relevante que el criterio más mencionado pertenece a una categoría semántica, lo cual sugiere que se prioriza de manera significativa el entendimiento de las oraciones, en particular las narraciones que siguen la estructura de las HU.

La clasificación por tipo de agrupamiento muestra, además, una marcada tendencia hacia criterios de carácter individual, dado que la mayoría de los criterios (12) puede evaluarse sin necesidad de agrupar las HU. No obstante, la presencia de criterios en set sugiere la importancia de considerar el contexto compartido entre diferentes requisitos, ya que la mayoría de estos criterios en set también corresponde a criterios pragmáticos.

Cabe destacar que se excluyeron aquellos criterios cuya cantidad de menciones en los estudios fuese inferior a diez. Estos sumaron alrededor de 41 criterios, agrupados en el enlace correspondiente (véase criterios). Algunos de estos criterios son similares o pueden interpretarse de manera similar a los más mencionados; sin embargo, se tomó como referencia la forma en que fueron descritos e implementados, lo que llevó a considerarlos como conceptos distintos.

La Fig. 4 ilustra la frecuencia de los criterios más mencionados al considerar todos los estudios analizados. Asimismo, se contabilizaron aquellos artículos que incorporaron explícitamente un marco de calidad o que emplearon determinados criterios señalando su origen en un estándar específico.

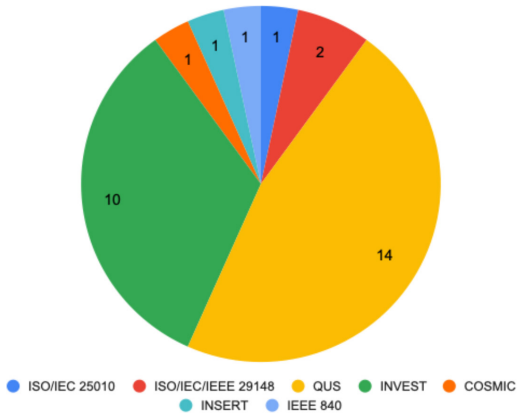


Fig. 4: *Marcos de calidad referenciados.*

Por su parte, la Fig. 5 muestra la frecuencia con la que los distintos marcos de calidad fueron referenciados en los estudios, contabilizando únicamente aquellos que fueron citados de manera explícita, independientemente de si se utilizaron o no todos sus criterios.

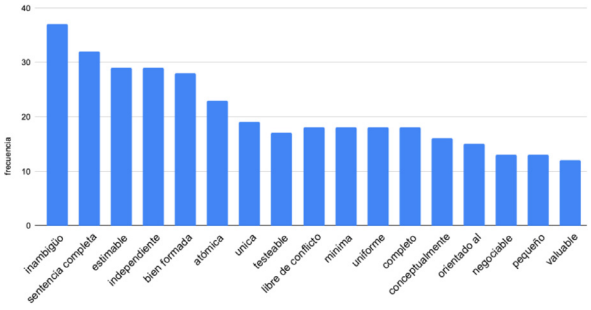


Fig. 5: *Frecuencia de criterios más referenciados.*

Los resultados evidencian nuevamente una marcada tendencia hacia el uso del marco QUS, citado en el 46.7% de los estudios, seguido de INVEST, presente en el 33%. Otros marcos, como COSMIC, aparecen en el 6.7% de los estudios, mientras que dos estándares de ISO e IEEE, así como una actualización de INVEST, presentan una frecuencia menor (3.4% cada uno).

El marco QUS, al ser una propuesta más reciente, puede considerarse más completo, lo que explicaría su uso ligeramente más frecuente. Las investigaciones actuales tienden a adoptar clasificaciones más rigurosas y formalmente definidas, lo que pone de manifiesto la necesidad de ampliar el espectro de criterios considerados dentro de dicho marco de calidad. No obstante, los criterios INVEST, pese a ser una taxonomía más antigua, siguen siendo ampliamente utilizados. Esto evidencia que la calidad de las HU no se reduce a una sola problemática (como la ambigüedad), sino que requiere marcos capaces de abordar múltiples dimensiones.

IV. Conclusiones

El análisis realizado refleja un creciente interés en la verificación de historias de usuario, evidenciado por el aumento de estudios en la última década y, especialmente, durante el año más reciente, 2024. Al examinar la evolución temporal de la problemática, se identifica un incremento progresivo que va desde un par de estudios en 2014 hasta 14 estudios diez años después. Sin embargo, se presentan algunas fluctuaciones en años específicos.

Un hallazgo relevante es que la técnica de historias

de usuario está intrínsecamente vinculada al procesamiento de lenguaje natural. Aspectos como la inambigüedad, la estimación, la adecuada redacción y la correcta conceptualización convierten a esta técnica en una herramienta poderosa para la comunicación asertiva entre stakeholders y desarrolladores, pero también en un objeto de estudio indispensable para asegurar la calidad del proceso y de los productos de software.

Los criterios más empleados para evaluar esta calidad se encuentran integrados principalmente en los marcos QUS e INVEST, destacándose aspectos como la inambigüedad y la independencia. No obstante, dichos aspectos no constituyen las únicas problemáticas. Aunque la ambigüedad es uno de los problemas más recurrentes, otros factores también impactan la calidad: una HU que no es independiente puede afectar los tiempos de entrega en equipos ágiles; una estimación inadecuada puede comprometer la planificación; y una orientación excesiva hacia la solución, en lugar de centrarse en la problemática, puede limitar el diseño de los productos de software. Estos y otros elementos identificados representan la antítesis de los criterios propuestos por los marcos de calidad.

El panorama global de la literatura analizada muestra una clara concentración de estudios en países tecnológicamente avanzados, lo cual evidencia una brecha en la investigación que podría abordarse en regiones en desarrollo. Asimismo, se identifica la necesidad de ampliar el espectro de criterios considerados, más allá de aquellos tradicionalmente empleados, e incorporar tecnologías emergentes para su evaluación automática. Este trabajo contribuye a sentar las bases para futuras iniciativas orientadas a mejorar la calidad de las historias de usuario y, en consecuencia, la eficacia del desarrollo ágil de software.

Como trabajo futuro, se plantea la ampliación de este estudio mediante un enfoque cualitativo sustentado en una revisión sistemática de la literatura. Dicho enfoque permitirá identificar con mayor precisión las técnicas y tecnologías empleadas para la verificación de historias de usuario, así como su relación con los marcos de calidad comúnmente utilizados, tales como QUS e INVEST.

De manera complementaria, se propone explorar el papel de los grandes modelos de lenguaje en actividades de verificación. Esta línea futura contempla tanto la identificación de estudios previos que hayan empleado modelos generativos para evaluar criterios de calidad como el diseño e implementación de experimentos propios orientados a la verificación automática de historias de usuario.

V. Referencias

[1] "IEEE Standard Glossary of Software Engineering Terminology" IEEE Std 610.12-1990, pp. 1-84, 1990.

[2] R. S. Pressman, "Software Engineering: A Practitioner's Approach", Seventh Edition, McGraw Hill Higher Education, 2010.

[3] S. Ashmore and K. Runyan, "Introduction to Agile Methods", Addison-Wesley, Pearson Education, Inc., 2015.

[4] K. Beck, M. Beedle, A. van Bennekum, A. Cockburn, W. Cunningham, M. Fowler, J. Grenning, J. Highsmith, A. Hunt, R. Jeffries, J. Kern, B. Marick, R. C. Martin, S. Mellor, K. Schwaber and J. Sutherland, "Manifesto for Agile Software Development," 2001. [Online]. Available: http://agilemanifesto.org/.

[5] J. Shore, "The Art of Agile Development", 2nd Edition, O'Reilly Media, Inc., 2021.

[6] M. Cohn, "User Stories Applied: For Agile Software Development", Addison Wesley Longman Publishing Co., Inc., 2004.

[7] X. Wang, L. Zhao, Y. Wang and J. Sun, "The Role of Requirements Engineering Practices in Agile Development: An Empirical Study", Communications in Computer and Information Science, 2014.

[8] N. Godbole, "Software Quality Assurance Principles And Practice", Alpha Science, 2004.

[9] G. Lucassen, F. Dalpiaz, J. M. Van der Werf and S. B. Brinkkemper, "Improving agile requirements: the Quality User

Story framework and tool," Requirements Engineering, 2016.

[10] L. Montgomery, D. Fucci, A. Bouraffa, L. Scholz and W. Maalej , "Empirical research on requirements quality: a systematic mapping study", Requirements Engineering, vol. 27, no. 2, pp. 183-209, 2022.

[11] A. R. Amna and G. Poels, "Systematic Literature Mapping of User Story Research", IEEE Access, 2022.

[12] A. R. Amna and G. Poels, "Ambiguity in user stories: A systematic literature review", Information and Software Technology, 2022.

[13] A. Hendriana, T. Raharjo and A. Nurfitriani, "Approaches in Determining User Story Quality through Requirement Elicitation: A Systematic Literature Review", Indonesian Journal of Computer Science, 2023.

[14] K. Petersen, R. Feldt, S. Mujtaba and M. Mattsson, "Systematic Mapping Studies in Software Engineering", Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering, vol. 17, 2008.

[15] K. Petersen, S. Vakkalanka and L. Kuzniarz, "Guidelines for conducting systematic mapping studies in software engineering: An update", Information and Software Technology, vol. 64, pp. 1-18, 2015.

[16] B. A. Kitchenham, D. Budgen and P. Brereton, "Evidence-Based Software Engineering and Systematic Reviews", Taylor & Francis Group, 2015.

[17] O. I. Lindland, G. Sindre and A. Sølvberg, "Understanding quality in conceptual modeling", IEEE Software, vol. 11, no. 2, pp. 42-49, 1994.

[18] K. Schwaber and J. Sutherland, "Scrum Guide", 2020. [Online]. Available: https://scrumguides.org/revisions.html.

[19] G. Lucassen, F. Dalpiaz, J. M. Van der Werf and S. B. Brinkkemper, "The Use and Effectiveness of User Stories in Practice," in Requirements Engineering: Foundation for Software Quality, 2016.