

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/311948162>

# Calidad en la inspección de requerimientos de software: Una propuesta de medición

Conference Paper · April 2016

CITATIONS

0

READS

525

5 authors, including:



**Carlos Monsalve**

Escuela Superior Politécnica del Litoral (ESPOL)

26 PUBLICATIONS 111 CITATIONS

[SEE PROFILE](#)



**Alex Armijos**

Escuela Superior Politécnica del Litoral (ESPOL)

2 PUBLICATIONS 0 CITATIONS

[SEE PROFILE](#)



**Ruben Ullon**

Escuela Superior Politécnica del Litoral (ESPOL)

2 PUBLICATIONS 1 CITATION

[SEE PROFILE](#)



**Ricardo D. Maya**

Escuela Superior Politécnica del Litoral (ESPOL)

2 PUBLICATIONS 1 CITATION

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



VLIR -Telecommunications Subcomponente 2007-2010 [View project](#)



ESFUERZO PARA INSPECCIÓN DE DOCUMENTOS DE ESPECIFICACIONES DE REQUERIMIENTOS DE SOFTWARE: FACTORES CRÍTICOS Y CONSTRUCCIÓN DE UN MODELO DE ESTIMACIÓN. [View project](#)

# Calidad en la inspección de requerimientos de software: una propuesta de medición

Alex F. Armijos, Carlos Monsalve, Rubén H. Ullón, Ricardo D. Maya, and José A. Romero

Facultad de Ingeniería en Electricidad y Computación (FIEC)  
Escuela Superior Politécnica del Litoral (ESPOL)  
Campus Gustavo Galindo, Km 30.5 vía Perimetral, Guayaquil-Ecuador  
{afarmijos,monsalve}@fiec.espol.edu.ec  
{rullon,rmaya,joanrome}@espol.edu.ec

**Resumen** La inspección de software, dentro del análisis de requerimientos, consiste en un conjunto de pasos ordenados y coordinados que intenta detectar tempranamente defectos para prevenir la construcción de un software que no responda a las necesidades del usuario final. Sin embargo, aunque se sigan estos pasos, su calidad podría verse afectada por el perfil de los inspectores. La calidad en la inspección de software no ha sido estudiada a profundidad; es decir, no se ha definido claramente que mediciones de calidad realizar y cómo realizar comparaciones entre inspectores. El presente estudio propone, con bases estadísticamente probadas, una forma concreta para medir la calidad en trabajos de inspección realizados por inspectores con diferentes perfiles. El experimento fue diseñado basado en la norma IEEE 1028-2008 y consistió en inspeccionar casos de uso para un software de un organismo del gobierno ecuatoriano donde los inspectores tenían dos perfiles: estudiantes universitarios y profesionales de computación. Los resultados obtenidos presentan mejores índices de calidad en los profesionales. La medición de calidad de inspección propuesta podría mejorarse con experimentos adicionales y trabajos a futuro.

## 1. Introducción

Construir software libre de defectos es una aspiración que todo ingeniero o equipo de ingenieros de software desea. Con este fin, en el proceso de desarrollo de software se utilizan distintas técnicas que apoyen este objetivo. Una técnica ampliamente usada y estudiada es la inspección de software que tuvo su origen en 1976 cuando Fagan [1] realizó su primera propuesta enfocada en la inspección de código. Desde ese entonces la técnica de inspección de software ha tenido variantes con experimentos diversos que han mostrado su aplicabilidad en la industria. La importancia de la mencionada técnica radica en que mientras más defectos significativos se encuentren en una etapa temprana de la construcción del software, más posibilidades existen que el producto final responda a las necesidades funcionales. Por lo tanto, encontrar defectos en los Documentos de Especificación

de Requerimientos de Software (DERS) durante el análisis de requerimientos reducirá el riesgo de producir software que no satisfaga las necesidades de usuarios y clientes finales.

Actualmente la norma IEEE 1028-2008 [2] establece una serie de pasos y recomendaciones que se deben cumplir para que una inspección (revisión sistemática de un artefacto) cumpla su cometido de encontrar defectos en los DERS. Sin embargo, estas recomendaciones sólo se enfocan en la detección de defectos mas no en medir la calidad del proceso de inspección; es decir, no plantea ningún indicador que refleje la cantidad de defectos detectados por el inspector que son realmente defectos (eficacia) y el esfuerzo demandado por la inspección (eficiencia). El mencionado esfuerzo corresponde a toda la inspección (revisión del artefacto) aunque este no tuviese defectos. La medición de la calidad de inspección es de utilidad pues permite realizar comparaciones de productividad y cambios en el proceso de inspección si fuesen necesarios (en los pasos o en los inspectores); y así, encontrar cada vez mayor cantidad de defectos en menor tiempo en los artefactos inspeccionados. Siendo la medición de la calidad de inspección de utilidad, entonces también resulta de interés el identificar los factores que podrían incidir en la calidad mencionada. Por un lado algunas métricas de inspección de software han sido resumidas por Nair, Suma y Tiwari [3] mientras que por otro, Monsalve, Ullón y Maya [4] plantean que se debe analizar si la experiencia de los inspectores es un factor que influye en el acierto de los defectos y en el esfuerzo requerido para la inspección.

El presente estudio tiene un doble objetivo: 1) analizar si los perfiles de los inspectores influyen en la calidad de la inspección; y 2) proponer un nuevo método de cálculo de la calidad de inspección que incluye tanto la eficacia de acierto al detectar defectos como la afectación del esfuerzo que toma detectar dichos defectos. El método propuesto es un modelo estadístico, y su aplicación en un caso de estudio y un experimento de inspección de DERS contribuye a responder si la experiencia de los inspectores influye en la calidad de detección de defectos. Tanto el caso de estudio como el experimento incluyen dos tipos de perfiles de los inspectores participantes: estudiantes y profesionales del área de computación. La hipótesis que formulamos es: *La calidad de inspecciones realizadas por profesionales es significativamente más alta que las realizadas por estudiantes.*

A continuación exponemos la estructura de esta investigación. En la sección 2 se expone lo relacionado al proceso de inspección, y los trabajos actuales para medir su calidad. La sección 3 explica el método de cálculo propuesto. La sección 4 explica la metodología con que se diseñó el experimento. La sección 5 analiza los resultados, las secciones 6 y 7 discuten y concluyen el trabajo.

## 2. Inspección y Calidad de Inspección

### 2.1. La Inspección

El proceso de inspección fue inicialmente propuesto por Fagan [1] para detectar defectos en el código construido; sin embargo, el uso de este proceso ha

sido extendido a otros artefactos que intervienen en el proceso de construcción de software. La norma IEEE-1028-2008 detalla los documentos que pueden ser inspeccionados por este proceso. En esta norma también se detallan los roles involucrados: 1) Moderador, 2) Registrador, 3) Lector, 4) Autor y 5) Inspector; así como sus etapas: 1) Preparar el proceso, 2) Planificar el trabajo, 3) Revisar anticipadamente los procedimientos, 4) Revisar anticipadamente el producto a inspeccionar, 5) Preparar la inspección (inspección individual), 6) Inspeccionar (inspección conjunta), 7) Remodelar y 8) Culminar. El *moderador* realiza una labor de guía y de coordinación de las reuniones de inspección, el *registrador* toma nota de los defectos detectados, el *lector* parafrasea el documento, el *autor* explica o modifica el documento (de ser necesario) y el *inspector* detecta los defectos. La *preparación del proceso* define el mejor proceso de inspección para el trabajo deseado. Durante la *planificación del trabajo* se agenda las reuniones y se define el equipo de inspección. La etapa de *revisar anticipadamente los procedimientos* proporciona una introducción sobre el proceso a los participantes (de la inspección). La *revisión anticipada del producto a inspeccionar* aproxima el documento a los participantes. Durante la *preparación de la inspección se hace una inspección individual* del documento. El *inspeccionar* detecta los defectos en inspecciones conjuntas (en reuniones). La etapa de *remodelar* verifica que los puntos encontrados en la reunión de inspección estén concluidos. Finalmente, la etapa de *culminar* es cuando se termina el proceso mediante un criterio de salida.

## 2.2. Métricas de calidad de inspección

El aporte que realiza presente trabajo sobre lo revisado en la literatura (sobre métricas de calidad de inspección) consiste en que las métricas actualmente propuestas no consideran el nivel de impacto de los defectos al momento de definir las métricas de calidad de inspección; y por tanto estas métricas de calidad no penalizan las mediciones de acuerdo al grado del acierto que tuvieron los inspectores al encontrar defectos. Otra diferencia es que ninguna de las métricas propuestas integran en una sola métrica los conceptos de eficacia y eficiencia, sino que miden por separado acierto y eficiencia. Los trabajos relacionados a las métricas de calidad de inspección son: [5–11]. Nair y Suma en [6] proponen métricas de profundidad de inspección (DI) y un indicador de procesos de inspección (IPM). El DI relaciona los defectos encontrados en la inspección versus los encontrados en la fase de pruebas y el IPM es un indicador del esfuerzo. Estas métricas han sido aplicadas en [5–10] y su motivación consiste en partir de un factor de calidad en función de aciertos (efectividad) y cuánto tiempo toma encontrar defectos (eficiencia). Estas métricas incluyen parámetros de inspección como tiempo de preparación, tiempo de inspección, nivel de experiencia, cantidad de inspectores y complejidad. Por otro lado Saito *et al.* [11] midieron trabajos de inspección con un puntaje de calidad orientado a medir el acierto pero sin incluir el esfuerzo dentro de este.

### 2.3. Inspecciones y experiencia de los inspectores

Existen estudios relacionados con el mejoramiento de la efectividad del proceso de inspección que proponen cambios al proceso: [12,13,13–18]. McCarthy *et al.* [12] realizaron un análisis costo beneficio (de las reuniones colaborativas de inspección), sugiriendo que se omitan dichas reuniones ya que no son necesarias para el éxito del proceso. Por otro lado, Eickelmann *et al.* [13] crearon un simulador con datos de inspección recolectados (sobre inspecciones) por Motorola para identificar los factores que intervienen en la efectividad de la inspección, y concluyeron que si las horas de inspección son incrementadas, su efectividad tiende a disminuir. Esto también lo confirma Raz y Yaung [17] con la variante que el tamaño del producto inspeccionado también influye en la efectividad.

Otros estudios explican cómo influye la experiencia del inspector en la inspección de requerimientos. Estos estudios realizan comparaciones entre inspectores profesionales y estudiantes universitarios: [12,19–22]. En un estudio comparativo de métodos de inspección, Porter y Votta [19] indican que los resultados de las inspecciones realizadas por estudiantes son similares a la de los profesionales, sugiriendo que por tanto es preferible hacer experimentos con estudiantes porque hacerlos con profesionales resulta costoso. En otro estudio, Albayrak y Carver [20] concluyeron que la experiencia en los inspectores tiene una relación positiva sobre la efectividad de la inspección. Por otro lado, Carver *et al.* [21] propusieron que los inspectores novatos participen como observadores de otros procesos de inspección para mejorar su nivel de acierto y así realizar procesos de inspección más efectivos.

Existen otros trabajos [16,23–28] que exponen propuestas para mejorar la calidad de la inspección; sin embargo, ninguno de ellos realizan mediciones de esa calidad.

### 2.4. Adaptación de la Norma IEEE 1028-2008

El proceso para las inspecciones realizadas en el presente trabajo fue diseñado tomando como referencia la norma IEEE 1028-2008. Algunas etapas fueron consolidadas en una por lo que exponemos las etapas adaptadas y su correspondencia con lo indicado en la norma: 1) Planificación (Preparar el proceso y Planificar el trabajo); 2) Arranque (Revisión anticipada de los procedimientos y Revisión anticipadamente del producto a inspeccionar); 3) Preparación (Preparar la inspección); 4) Inspección; 5) Explicación (Remodelar); 6) Cierre (Culminar).

En la *preparación* fue realizada la inspección individual e intervino un miembro con el rol de *inspector*. En la *inspección* fue realizada la inspección conjunta e intervinieron: 1) un *inspector* que también hacía de *secretario* (conocido en la norma como *registrador*) y 2) un *inspector* que también participaba como *lector*. En la *explicación* los resultados de las inspecciones son revisados con el *autor*. El rol de *moderador* se mantuvo tanto para la *preparación* como para la *inspección*.

### 3. Método de Calidad propuesto

#### 3.1. Calidad integral de inspección

La medición de calidad de inspección propuesta incluye en un mismo puntaje tanto un componente de esfuerzo como otro de eficacia de la detección de defectos; es decir, este trabajo propone una medición de la calidad integral. El primer componente está basado en el esfuerzo requerido para detectar los defectos (según el nivel de impacto de los defectos), mientras que el segundo está basado en el nivel de acierto en la detección de defectos (según el tipo de defecto y su nivel de impacto). Para ello, definimos un puntaje  $Q$  que será la calidad de inspección promedio para un grupo de inspecciones realizadas cada una sobre un artefacto. Este puntaje  $Q$  se define en el rango 0 a 100, siendo 100 el puntaje obtenido si todos los supuestos defectos encontrados por el inspector son aciertos y 0 para el caso en que el inspector no presenta acierto alguno. Un defecto encontrado por un inspector se considera acertado cuando este ha sido confirmado por el autor del artefacto inspeccionado, tanto en *Tipo de Defecto* (ortografía, redacción, ambigüedad, inconsistencia e incompleto) como en *Nivel de Impacto* (crítico o menor).

j	Tipo de Defecto	Nivel de Impacto	Valor Priorización	Penalidad
0	Acertado	Acertado	38.19	1.00
1	No Acertado	Acertado	33.40	0.87
2	Acertado	No Acertado	27.72	0.73
3	No Acertado	No Acertado	0.69	0.02

Cuadro 1: Penalidad a aplicar de acuerdo al desacierto del defecto.

Por cada desacierto ocurrido en la inspección se aplica una penalidad al puntaje cuyo valor podemos observar en el Cuadro 1. Estas penalidades se obtuvieron a partir del juicio de expertos en priorizar los tipos de aciertos y desaciertos en los resultados de inspección, tal como se explica en el siguiente párrafo. Por ejemplo, si el inspector acierta el tipo de defecto pero desacierta el nivel de impacto, entonces se le aplica la penalidad número 2 que es de 0.73. Los valores  $j$  indentifican la penalidad a aplicar.

Los valores de penalidades fueron calculados mediante una relación lineal aplicada sobre los valores de priorización, asignando 1 al valor de priorización 38.19 (acierto en ambos criterios). Los valores de priorización fueron determinados a partir de una matriz de priorización generada a partir de una encuesta realizada a 11 expertos a los que se consultó cuál de los criterios (tipo de defecto y nivel de impacto) tenía más peso si se acertaba (o desacertaba) en cada defecto detectado. Para ello se realizaron los siguientes pasos: 1) Crear una matriz de priorización de importancia sobre los tipos de desaciertos; 2) Realizar un pilotaje del llenado de esta matriz por tres expertos; 3) Llenado de la matriz de priorización por 11 expertos mediante una encuesta electrónica; y 4) Definir los valores

de penalización de acuerdo a las priorizaciones ingresadas por los expertos. Llamaremos  $P_j$  al elemento  $j$  del arreglo  $P$  que contiene las penalidades definidas en el Cuadro 1; donde  $j$  pertenece al rango  $0..3$ .

$$TCC_i * X_i + TMC_i * Y_i = 100 \quad (1)$$

La ecuación (1) es la base para determinar el aporte de cada defecto acertado (si no hubiese penalidad, por eso suma 100) al puntaje de calidad  $Q$ ; siendo el subíndice  $i$  la inspección realizada en un artefacto, donde  $i$  se define en el rango  $0..n-1$ , siendo  $n$  artefactos inspeccionados. Es decir, cada uno de los  $n$  artefactos es inspeccionado una vez.  $X_i$  es el puntaje por acertar cada defecto de impacto crítico en la  $i$ ésima inspección realizada; y  $TCC_i$  es la cantidad total de verdaderos defectos de impacto crítico que tiene el artefacto confirmados por el autor en la  $i$ ésima inspección realizada. Por otro lado,  $Y_i$  es el puntaje por acertar cada defecto de impacto menor en la  $i$ ésima inspección realizada y  $TMC_i$  es la cantidad total de defectos de impacto menor confirmados por el autor en la  $i$ ésima inspección realizada.  $X$ ,  $Y$ ,  $TCC$  y  $TMC$  son arreglos de tamaño  $n$ .

$$K = \frac{X_i}{Y_i} \quad (2)$$

Para introducir la afectación del esfuerzo dentro del puntaje  $Q$  fueron separados por un lado los puntajes de los defectos de impacto crítico de los defectos de impacto menor. El motivo de esta división se debe al supuesto que encontrar un defecto de impacto crítico no toma el mismo esfuerzo que encontrar uno de impacto menor por lo tanto, no puede otorgarse el mismo puntaje. En la sección 5.1 explicamos la correlación que existe entre el esfuerzo y el nivel de impacto de los defectos detectados al realizar las inspecciones. Para utilizar esta correlación, definimos la constante  $K$  como la relación que existe entre el esfuerzo por encontrar un defecto de impacto crítico y el esfuerzo por encontrar uno de impacto menor. La sección 3.2 explica cómo calcular esta constante. Una vez obtenido este valor de  $K$ , podemos definir la relación entre el puntaje por acertar un defecto de impacto crítico ( $X_i$ ) y uno de impacto menor ( $Y_i$ ), para cada una de las inspecciones realizadas. Esta relación de puntajes se encuentra definida en la ecuación (2). Cada inspección tiene un par  $X_i, Y_i$  porque estos valores otorgan puntos a cada defecto y la cantidad de defectos detectados varía por cada inspección; por lo tanto, estos valores deben ser calculados para cada inspección. El método de cálculo sólo considera los falsos positivos, es decir defectos encontrados que el autor no confirmó. No se consideran los falsos negativos (defectos no detectados que el autor hubiese confirmado).

### 3.2. Cálculo de la constante $K$

Cada constante  $K$  puede definirse dependiendo de los grupos de inspecciones que deseemos comparar. Es decir, que para este trabajo vamos a definir dos constantes  $K$ : 1) desde el esfuerzo en las inspecciones realizadas por los estudiantes; y 2) desde el esfuerzo en las inspecciones realizadas por los profesionales.

La hipótesis planteada en este trabajo deberá ser contrastada para ambas  $K$ ; es decir, encontrar sus valores y realizar una prueba de hipótesis para cada una.

Cada valor de  $K$  puede obtenerse con los coeficientes de una regresión lineal múltiple que explique el esfuerzo (variable dependiente) versus la cantidad de defectos críticos y menores (variables independientes). Es decir, cada  $K$  se calculará dividiendo el coeficiente correspondiente a los defectos de impacto crítico sobre el correspondiente a los defectos de impacto menor, coeficientes resultantes de la mencionada regresión.

### 3.3. Cálculo de los puntajes $X$ y $Y$

Los puntajes para defectos de impacto crítico están dados por el arreglo  $X$ , mientras que los correspondientes a los de impacto menor están dados por  $Y$ . Cada  $X_i$  y  $Y_i$  (siendo  $i$  el subíndice correspondiente una inspección realizada sobre un artefacto en el rango de  $0..n-1$ ) es hallado resolviendo el sistema de ecuaciones (1) y (2) para cada inspección realizada (una sobre cada artefacto).

### 3.4. Cálculo del puntaje de calidad integral $Q$

$$Q = \frac{\sum_{i=0}^{n-1} Q_i}{n} \quad (3)$$

El puntaje de calidad integral  $Q$  como se indica en la ecuación (3), está dado por el promedio de los puntajes de calidad calculados para cada una de las  $n$  inspecciones y cuyos valores están dados por  $QI_i$ .

$$QI_i = QC_i + QM_i \quad (4)$$

Para calcular  $QI_i$  (puntaje de calidad para cada inspección), como se indica en la ecuación (4), debemos obtener por un lado los puntajes de calidad de los defectos de impacto crítico (dado por  $QC_i$ ) y por otro los de impacto menor (dado por  $QM_i$ ) de cada  $i$ -ésima inspección. El valor de estos puntajes de calidad separados por el nivel de impacto se obtienen de las ecuaciones (5) y (6) respectivamente.

$$QC_i = \sum_{j=0}^3 X_i * P_j * DC_{ij} \quad (5)$$

$$QM_i = \sum_{j=0}^3 Y_i * P_j * DM_{ij} \quad (6)$$

Llamaremos  $DC_{ij}$  al elemento de la  $i$ -ésima fila y de la  $j$ -ésima columna de la matriz  $DC$  que contiene la cantidad de defectos detectados por el inspector y que los identificó como de impacto crítico de la  $i$ -ésima inspección y de la  $j$ -ésima penalidad de desacuerdo (penalidad enumerada en el Cuadro 1). La matriz



$DC$  es utilizada para determinar cuántos defectos deben ser castigados con una determinada penalización. De la misma manera  $DM$  será la matriz que contendrá la cantidad de defectos de impacto menor, siendo del mismo tamaño de  $DC$  con los mismos subíndices  $i$  y  $j$ . En las ecuaciones (5) y (6) son aplicadas las penalidades cuyos valores se encuentran en el arreglo  $P$ , castigando los puntajes  $X_i$  y  $Y_i$  de la  $i$ -ésima inspección. Ejemplos de estos cálculos se presentan en el Cuadro 2 en la sección 5.

## 4. Metodología

Para probar la hipótesis en estudio exponemos el siguiente diseño metodológico.

### 4.1. Lineamientos de las inspecciones

Siguiendo la línea de la hipótesis planteada, es necesario realizar inspecciones utilizando inspectores profesionales por un lado e inspectores estudiantes por otro. Para ambos grupos de inspecciones se respetaron los siguientes lineamientos: 1) Asignación equitativa de cantidad de páginas de los artefactos (DERS) para cada inspector; 2) Cada sesión de inspección no duró más de 2 horas; 3) Los inspectores recibieron talleres sobre el proceso de inspección, uso del formulario de inspección y explicación de los documentos a inspeccionar (artefactos); 4) Los artefactos inspeccionados siempre provienen de un mismo grupo de DERS; 5) Se utilizó el mismo formulario (validado y probado en inspecciones previas); y 6) Cada defecto encontrado fue señalado en el artefacto y registrado en el formulario con un comentario. El autor sólo confirmó los defectos encontrados por los inspectores, es decir, él no detectó ningún defecto durante la *explicación*.

### 4.2. Artefactos inspeccionados

Las inspecciones realizadas por los profesionales fueron llevadas a cabo en un caso de estudio donde fueron inspeccionados 209 documentos de especificaciones de casos de uso pertenecientes a una institución del gobierno ecuatoriano para un proyecto de un sistema financiero nacional. En estas inspecciones fueron ejecutadas todas las etapas indicadas en la sección 2.4 y participaron dos grupos de profesionales en ciencias de la computación (grupo profesional 1 y 2). Los 209 casos de uso fueron asignados equitativamente (misma cantidad de páginas) para los 8 inspectores profesionales en ciencias de la computación. Como lo indican las etapas, el proceso de inspección fue preparado, las capacitaciones y aproximación a los casos de uso fueron realizadas y se procedió a realizar la inspección individual (etapa de *preparación*). Luego, en cuatro grupos de inspección los documentos de casos de uso fueron revisados en la inspección conjunta (etapa de *inspección*) donde los autores de los artefactos confirmaron por cada defecto detectado si el tipo de defecto y su nivel de impacto eran aciertos. En esta etapa intervinieron los siguientes roles: *secretario*, *lector* y *moderador*. A continuación,

en la etapa de *explicación*, el autor realizó la confirmación de los defectos detectados por los inspectores. Esta confirmación fue el punto de referencia para penalizar el desacierto al aplicar el nuevo método de cálculo.

Las inspecciones realizadas por los estudiantes fueron llevadas a cabo en un experimento en una escuela politécnica ecuatoriana durante una sesión de clases de la asignatura de Ingeniería de Software con máximo 2 horas de duración (el mismo tiempo máximo que tuvieron las sesiones de inspección realizadas en el caso de estudio). Estas inspecciones solo consideraron las etapas de *planificación*, *arranque* y *preparación* (inspección individual). Debido a que el experimento sólo se podía realizar en una sesión de clases de 2 horas (por restricciones de la asignatura) era imposible inspeccionar los 209 documentos de casos de uso ya que habían sólo 23 estudiantes. Por lo tanto, para garantizar la igualdad de condiciones y poder comparar las inspecciones en el experimento fue necesario: 1) inspeccionar una muestra de 54 documentos casos de uso (tamaño de muestra calculado con el programa STATS) y asignar similar carga de inspección a cada inspector (19 páginas de casos de uso); y 2) considerar en las inspecciones realizadas por profesionales sólo los defectos detectados hasta la etapa de *preparación* (sólo se compararon hasta las inspecciones individuales).

Los 54 casos de uso fueron escogidos al azar respetando la distribución del tamaño de los 209 casos de uso: 79.90 % pequeños (menos de 10 páginas), 19.14 % medianos (entre 11 y 15 páginas) y 0.96 grandes (más de 16 páginas).

### 4.3. Instrumentos utilizados

Los instrumentos utilizados para medir las inspecciones fueron: 1) Formulario de inspección; 2) Documentos de especificación de casos de uso a inspeccionar; y 3) Método de cálculo de la calidad de inspección ( $Q$ ).

## 5. Resultados

Acorde al diseño metodológico anteriormente expuesto, la calidad fue calculada para cada documento tanto para el caso de estudio como para el experimento. En esta sección se presentan los resultados para su posterior análisis y discusión. Un ejemplo de estos cálculos realizados son mostrados en el Cuadro 2.

i	Caso Uso	X	Y	TCC	TMC	$DC_j$				$DM_j$				QC	QM
						0	1	2	3	0	1	2	3		
0	No. 1	8.84	1.36	9.00	15.00	3.00	0.00	0.00	0.00	9.00	0.00	0.00	0.00	26.51	12.27
1	No. 2	35.06	5.41	1.00	12	1.00	0.00	0.00	0.00	8.00	1.00	1.00	0.00	35.06	51.95
2	No. 3	34.18	5.27	2.00	6.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	34.18	0.00

Cuadro 2: Ejemplo de cálculos realizados para obtener el puntaje  $Q$  en inspecciones profesionales del grupo 1 para  $K$  6.48

### 5.1. Esfuerzo y nivel de impacto

Para aplicar este método de cálculo propuesto es necesario confirmar que existe una correlación entre el esfuerzo(variable dependiente) y las cantidades de defectos de impacto crítico y menor (variables independientes). Para ello, se ejecutó una prueba de Pearson entre las variables mencionadas resultando:  $p < 0,05$  y  $r = 0,493$  (para inspecciones estudiantes) y  $p < 0,05$  y  $r = 0,364$  (para inspecciones profesionales). Estos valores confirman la existencia de la correlación.

### 5.2. Constantes K

	Profesionales			Estudiantiles		
Variables	Coefficiente	t	valor p	Coefficiente	t	valor p
Intercepto	0.287299	7.952	1.05e-13	0.478567	6.629	2.28e-08
Crítico	0.009637	5.100	7.47e-07	0.082238	3.469	0.00109
Menor	0.008319	2.775	0.006	0.012697	2.319	0.02453

Cuadro 3: Regresión lineal múltiple en inspecciones estudiantiles y profesionales

Una vez realizada la confirmación de correlación en la sección 5.1, aplicamos regresiones lineales múltiples cuyos coeficientes resultantes se muestran en el Cuadro 3. Los valores de la prueba F de estas regresiones fueron: 1) inspecciones realizadas por los estudiantes  $F(2, 50) = 11,61$ ,  $p < 0,001$  y  $R^2 = 0,3171$ ; 2) inspecciones realizadas por los profesionales  $F(2, 214) = 16,43$ ,  $p < 0,001$  y  $R^2 = 0,1331$ . Los modelos solo explican los datos al 32 % y 13 % respectivamente. El  $K$  correspondiente a los estudiantes fue  $K = 0,082238/0,012697 = 6,48$ . Como indicamos en la sección 3.2 dividimos entre el coeficiente de la regresión lineal múltiple del defecto de impacto crítico y el de impacto menor. Lo mismo fue realizado para el  $K$  de los profesionales:  $K = (0,009637/0,008319) = 1,15$ .

### 5.3. Puntajes de Calidad Obtenidos

	Media (Q)		Mediana		Max.		Desv.Est.		$Q_1$		$Q_3$	
Inspectores / K	6.48	1.15	6.48	1.15	6.48	1.15	6.48	1.15	6.48	1.15	6.48	1.15
Estudiantes	7.41	11.00	2.36	5.62	45.17	55.94	10.90	14.49	0.00	0.00	17.94	11.50
Profesionales1	40.51	38.86	38.24	42.08	100.00	100.00	35.89	33.63	0.14	0.06	68.29	73.90
Profesionales2	44.40	49.96	49.35	58.39	100.00	100.00	36.30	35.89	3.50	1.79	77.46	75.24

Cuadro 4: Resultados de Puntos de Calidad (la media corresponde a la Q final)

El Cuadro 4 describe los puntajes calculados para ambas  $K$ . Existen inspecciones realizadas por profesionales que no fueron penalizadas (100 % acierto) y obtuvieron 100 puntos. Las inspecciones de los estudiantes no llegaron a 56 puntos. Las  $Q$  (que corresponde a la media) de los profesionales son superiores a la de los estudiantes y lo mismo se repite con las medianas.

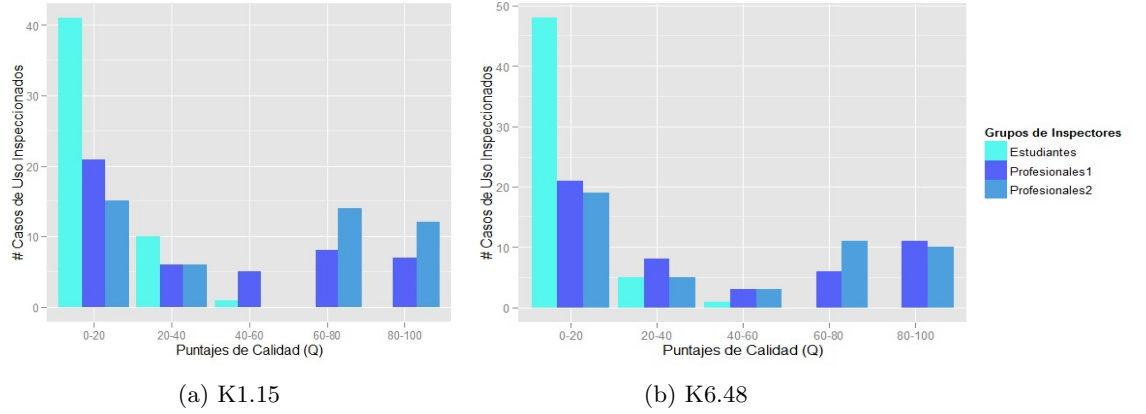


Figura 2: Puntajes de calidad Q calculados

En las Figuras 2a y 2b mostramos los números de casos de uso (inspecciones) según el puntaje obtenido. Para ambas K, la mayor cantidad de inspecciones obtuvieron puntajes inferiores a 20, sobretodo con las realizadas por estudiantes. En el resto de rangos de puntajes, los profesionales concentraron sus puntajes entre 60 y 100, donde no aparecen inspecciones estudiantes.

#### 5.4. Hipótesis en estudio

La hipótesis en estudio es: *La calidad de inspecciones realizadas por profesionales es significativamente más alta que las realizadas por estudiantes*. Para poder probar esta hipótesis primero realizaremos la prueba de normalidad y homocedasticidad.

La prueba de Shapiro Wilk (prueba de normalidad) mostró un valor  $p = 1,564e - 12$  para  $K = 6.48$  y  $p = 3,356e - 11$  para  $K = 1.15$ , ambos inferiores al nivel de significación 0.05; por lo tanto, concluimos que los puntajes calculados no siguen una distribución normal. La prueba de Levene (prueba de homocedasticidad) mostró un valor  $p = 2,2e - 16$  para  $K = 6.48$  y  $p = 2,517e - 12$  para  $K = 1.15$ , ambos inferiores al nivel de significación 0.05; por lo tanto, concluimos que los puntajes calculados no tienen varianzas homogéneas.

Dado que los puntajes calculados no cumplen la normalidad, tampoco con la homocedasticidad, vamos a comparar 3 grupos (estudiantes vs profesionales grupo 1 y estudiantes vs profesionales grupo2), es pareada y de una cola, entonces para la prueba de hipótesis escogimos el Wilcoxon signed rank test. Si definimos a  $MD_{p1}$  como la mediana del puntaje de las inspecciones de los profesionales grupo 1,  $MD_{p2}$  como la mediana del puntaje de las inspecciones de los profesionales grupo 2 y  $MD_e$  como la mediana del puntaje de las inspecciones de los estudiantes; entonces las hipótesis nulas serán:  $Ho1 : MD_{p1} \leq MD_e$  y  $Ho2 : MD_{p2} \leq MD_e$ . Los valores p del test fueron: 1)  $p = 1,646e - 07$  para  $Ho1$  y  $K=6.48$ ; 2)  $p = 2,754e - 08$  para  $Ho2$  y  $K=6.48$ ; 3)  $p = 5,013e - 06$  para  $Ho1$

y  $K=1.15$ ; y 4)  $p = 1,143e - 07$  para  $H_02$  y  $K=1.15$ . Todos los valores  $p$  son inferiores al nivel de significación 0.05 por lo que rechazamos las  $H_0$  y concluimos que los puntajes  $Q$  de las inspecciones realizadas por profesionales si son superiores.

## 6. Discusión

Hemos comprobado que la calidad de inspección  $Q$  de los profesionales es mayor que la de los estudiantes. Esto podría parecer obvio; sin embargo, no teníamos una medida integral de calidad de inspección que corrobore qué tan distantes se encuentran estos dos perfiles de inspectores. A simple vista, podríamos pensar que la experiencia del inspector es un factor que influye en la calidad con que los documentos de especificación de casos de uso fueron inspeccionados. Sin embargo, se puede observar en las Figuras 1a y 1b que una gran cantidad de documentos inspeccionados por profesionales obtuvieron una medición de calidad de inspección inferior a 20 puntos, lo que indica que, seguramente, existen otros factores a más de la experiencia del inspector que influyen en las inspecciones.

Analizar los factores adicionales es una tarea a realizar a futuro que plantea sus propios retos. Podríamos partir analizando cuáles son los tipos de defectos que los inspectores profesionales no acertaron. La formación profesional no está asociada, necesariamente a la experiencia profesional y por lo tanto podría ser otro factor influyente. Alineado con lo investigado por Albayrak y Carver [20], podría ser que los inspectores de otras áreas ajenas a la computación sean más efectivos en la detección de defectos sobre documentos de especificación de casos de uso. Por lo tanto, a futuro también se podría realizar un estudio que compare los tipos de defectos involucrados en las inspecciones realizadas por profesionales de distintas áreas.

## 7. Conclusión

En este trabajo se desarrolló un caso de estudio donde participaron profesionales en ciencias de la computación realizando inspecciones de documentos de especificaciones de casos de uso de un sistema financiero nacional perteneciente a un ministerio del gobierno ecuatoriano. Luego, tomando estos mismos documentos se desarrolló un experimento en el que estudiantes universitarios (de una universidad politécnica ecuatoriana) hicieron trabajos de inspección sobre una muestra de estos documentos. La calidad de inspección de ambos grupos de inspecciones fueron comparadas. Para realizar las inspecciones, se siguió un proceso de inspección basado en la norma IEEE 1028-2008. Debido a las limitantes de tiempo, para el caso de las inspecciones realizadas por estudiantes sólo se consiguió llegar hasta la etapa de preparación (inspección individual), realizada en una sesión de clase de 2 horas. Para los profesionales, se consiguieron ejecutar todas las etapas pero solo se compararon con las inspecciones de estudiantes con los defectos encontrados hasta la etapa de preparación (inspección individual). Los defectos detectados por cada grupo de inspectores fueron

clasificados según el acierto con el objetivo de penalizar por cada desacierto la calidad de inspección de cada documento. Cada inspección de un artefacto obtuvo una métrica de calidad que fue usada para realizar comparaciones entre los dos grupos de inspecciones. La significancia de las diferencias entre las métricas de calidad fue comprobada mediante una prueba Kruskal Wallis. Es importante estudiar si factores adicionales a la experiencia del inspector influyeron en que los estudiantes tengan menores métricas de calidad que los profesionales, o por qué ciertos profesionales obtuvieron bajas calidad en la inspección de algunos de artefactos. Trabajos futuros podrían responder si el dominio de formación profesional (no solo la experiencia del inspector) podría influir en la detección de defectos. Finalmente, otros trabajos pueden replicar el uso del método de cálculo de calidad propuesto y realizar adaptaciones (como ampliar los niveles de impacto) que permitan su generalización.

**Agredecimientos.** Los autores agradecen a la Ing. Cristina Barreto, estudiante de maestría de la Escuela Superior Politécnica del Litoral (ESPOL), por su valiosa contribución en el orden de priorización de los criterios al acertar los defectos en las inspecciones.

## Referencias

1. Fagan: Design and code inspections to reduce errors in program development. In: The IBM Systems Journal. Volume 15. (1976) 182—211
2. IEEE: Ieee standard for software reviews and audits. In: IEEE Standard for Software Reviews and Audits. IEEE STD (2008) 1–52
3. Tiwari, T.G.N.V.S.P.K.: Significance of depth of inspection and inspection performance metrics for consistent defect management in software industry. The Institution of Engineering and Technology (2012) 524–535
4. Carlos Monsalve, Rubén Ullón, R.M.y.J.R.: Dataseries: an efficient, flexible data format for structured serial data. Revista Latinoamericana de Ingeniería de Software **3** (2015) 111–116
5. Tiwari, T.G.N.V.S.P.K.: Significance of depth of inspection and inspection performance metrics for consistent defect management in software industry. IET Software **6** (2012) 524–535
6. Gopalakrishnan Nair, T.R., S.V.: Defect management using pair metrics, di and ipm. CrossTalk, J. Def. Softw. Eng **24** (2011) 22–27
7. Gopalakrishnan Nair, T.R., S.V.: A paradigm for metric based inspection process for enhancing defect management. ACM SIGSOFT SEN **35** (2010) 1
8. Li, Q., S.F.B.B.W.Q.: Improving the roi of software quality assurance activities: an empirical study. LNCS (2010) 357–368
9. Suma V., T.R.G.N.: Four-step approach model of inspection (fami) for effective defect management in software development. InterJRI Science and Technology **3** (2011)
10. T.R.Gopalakrishnan Nair, S.V.: Estimation of characteristics of a software team for implementing effective inspection process through inspection performance metric. American Society for Quality (ASQ) Journal, Software Quality Professional **13** (2011)
11. Saito, S. Takeuchi, M..H.M..K.T..A.M.: Requirements clinic: Third party inspection methodology and practice for improving the quality of software requirements

- specifications. Requirements Engineering Conference (RE), 2013 21st IEEE International (2013)
12. McCarthy, Porter, S.V.: An experiment to assess cost-benefits of inspection meetings and their alternatives: a pilot study. Software Metrics Symposium 3rd International (1996)
  13. Eickelmann, N.S. ; Schaumburg, I.R.F..J.B..A.A.: An empirical study of modifying the fagan inspection process and the resulting main effects and interaction effects among defects found, effort required, rate of preparation and inspection, number of team members and product 1st pass quality. Software Engineering Workshop, 2002. Proceedings. 27th Annual NASA (2002)
  14. Amira A. Alshazly, Ahmed M. Elfataty, M.S.A.: Detecting defects in software requirements specification. Alexandria Engineering Journal **53** (2014) 513–527
  15. Albayrak, O.: An experiment to observe the impact of uml diagrams on the effectiveness of software requirements inspections. Empirical Software Engineering and Measurement in 3rd International Symposium on (2009)
  16. Schilling, W.: Teaching software inspection effectiveness: An active learning exercise. Frontiers in Education Conference (FIE) (2012)
  17. Tzvi Raz, A.T.Y.: Factors affecting design inspection effectiveness in software development. Information and Software Technology **39** (1997) 297–305
  18. Kollanus, S.: Experiences from using icmm in inspection process assessment. Software Quality Journal **17** (2009) 177–187
  19. Adam Porter, L.V.: Comparing detection methods for software requirements inspections: A replication using professional subjects. Empirical Software Engineering **3** (1998) 355–379
  20. Özlem Albayrak, J.C.C.: Investigation of individual factors impacting the effectiveness of requirements inspections: a replicated experiment. Empirical Software Engineering **19** (2014) 241–266
  21. Jeffrey Carver, Forrest Shull, V.B.: Can observational techniques help novices overcome the software inspection learning curve? an empirical investigation. Empirical Software Engineering **11** (2006) 523–539
  22. McMeekin, D.A.;von Konsky, B..R.M..C.D.: The significance of participant experience when evaluating software inspection techniques. Software Engineering Conference, 2009. ASWEC '09. Australian (2009) 200–209
  23. MacDonald, F.;Miller, J..B.A..R.M..W.M.: A review of tool support for software inspection. Computer-Aided Software Engineering Proceedings., Seventh International Workshop on (1995)
  24. Harjumaa, L.;Tervonen, I..V.P.: Improving software inspection process with patterns. Quality Software, 2004. QSIC 2004. Proceedings. Fourth International Conference on (2004)
  25. Stefan Wagner, T.S.: Software quality economics for defect-detection techniques using failure prediction. Newsletter ACM SIGSOFT Software Engineering **30** (2005) 1–6
  26. Mirosław Staron, Ludwik Kuzniarz, C.T.: An empirical assessment of using stereotypes to improve reading techniques in software inspections. Newsletter ACM SIGSOFT Software Engineering **30** (2005) 1–7
  27. Stefan Biffl, Paul Grünbacher, M.H.: Factors affecting design inspection effectiveness in software development. Automated Software Engineering **13** (2006) 373–394
  28. Gursimran S. Walia, J.C.C.: Using error abstraction and classification to improve requirement quality: conclusions from a family of four empirical studies. Empirical Software Engineering **18** (2012) 625–658