

MONTEVIDEO, URUGUAY / ABRIL, 2017

FACULTAD DE INGENIERÍA  
UNIVERSIDAD DE LA REPÚBLICA

---

# **Beneficios de aplicar el Personal Software Process en el desarrollo de software: un estudio de revisión sistemática**

Proyecto de Grado: Vicente Fava, Maite  
Mañana, José Ignacio Villaverde

---

Tutores: Diego Vallespir, Silvana Moreno







## RESUMEN

---

En la actualidad existe una gran necesidad de generar software de calidad de manera efectiva. El *Personal Software Process* (PSP) es un marco de trabajo que busca responder a esa necesidad. Este trabajo define y valida un protocolo de trabajo para la ejecución de una Revisión Sistemática de la Literatura (SLR, por sus siglas en inglés “systematic literature review”) que busca conocer los efectos del PSP en el desarrollo de software. Puntualmente esta SLR se enfoca en los efectos producidos por el PSP en el desarrollo de software, en las estrategias empíricas utilizadas para medir dichos efectos y en las métricas y métodos estadísticos aplicados al análisis de resultados.

Se proponen y ejecutan dos experiencias piloto con el objetivo de afinar y validar el protocolo de trabajo propuesto. Una vez validado, se ejecuta el proceso de selección de la SLR, del cual se obtiene un conjunto de 38 estudios primarios. Las experiencias piloto realizadas permiten afirmar que el protocolo de trabajo propuesto es válido para la ejecución de la SLR planteada. A partir de los datos preliminares se entiende que el PSP tiene un efecto positivo, en términos de exactitud de la estimación en tiempo, defectos por KLOC y productividad medida en LOC/hora, en el desarrollo de software.



## ÍNDICE GENERAL

---

1	INTRODUCCIÓN	1
1.1	Objetivos	2
1.2	Trabajo Realizado	2
1.3	Estructura del documento	3
2	MARCO TEÓRICO SOBRE INGENIERÍA DE SOFTWARE EMPÍRICA	5
2.1	Conceptos de Ingeniería de Software Empírica	5
2.1.1	Contexto y estrategias	5
2.1.2	Estrategias empíricas	6
2.2	Revisiones sistemáticas de la literatura	12
3	MARCO TEÓRICO SOBRE PERSONAL SOFTWARE PROCESS	15
3.1	Estructura del PSP	17
3.2	Métricas	18
3.2.1	Métricas de tiempo	18
3.2.2	Métricas de tamaño	18
3.2.3	Métricas de calidad	18
4	GUIA DE TRABAJO PARA REVISIONES SISTEMÁTICAS	21
4.1	Fases de la investigación	21
4.2	Planificación de la investigación	23
4.2.1	Identificar la necesidad de una SLR	23
4.2.2	Especificar las preguntas de investigación (PI)	23
4.2.3	Desarrollar el protocolo de trabajo	25
4.2.4	Evaluación del protocolo	26
4.3	Ejecución de la investigación	26
4.3.1	Identificación de la investigación	26
4.3.2	Selección de estudios primarios	29
4.3.3	Estudio de calidad de los estudios primarios	29
4.3.4	Extracción de datos	30
4.3.5	Síntesis de datos	31
4.4	Reportar la investigación	31
4.4.1	Especificación de los mecanismos de diseminación de datos	31
4.4.2	Formato del reporte final	32
5	PROTOCOLO DE TRABAJO	33
5.1	Identificar la necesidad de la revisión	33

5.2	Especificación de las preguntas de la SLR	34
5.2.1	Preguntas de investigación (PI)	34
5.2.2	Definición del alcance	34
5.3	Estrategia de búsqueda	35
5.3.1	Cadena de búsqueda	35
5.3.2	Recursos de búsqueda	35
5.3.3	Criterios de selección	35
5.3.4	Proceso de selección	36
5.3.5	Extracción de datos	38
5.3.6	Síntesis de datos extraídos	39
5.3.7	Cronograma de la revisión	40
6	EXPERIENCIAS PILOTO	43
6.1	Introducción	43
6.2	Experiencia Piloto 1	44
6.2.1	Proceso de selección	44
6.2.2	Extracción de datos	47
6.2.3	Síntesis de datos extraídos	47
6.2.4	Análisis cuantitativo	53
6.2.5	Otros hallazgos	56
6.2.6	Conclusiones	60
6.2.7	Modificaciones del protocolo 1	60
6.3	Experiencia piloto 2	62
6.3.1	Proceso de selección	63
6.3.2	Extracción de datos	68
6.3.3	Síntesis de datos extraídos	68
6.3.4	Análisis cuantitativo	73
6.3.5	Conclusiones	73
6.3.6	Modificaciones del protocolo 2	79
7	APLICACIÓN DEL MÉTODO DE INVESTIGACIÓN	81
7.1	Proceso de selección	81
7.2	Análisis cuantitativo	85
7.3	Conclusiones	91
8	CONCLUSIONES Y TRABAJO FUTURO	93
8.1	Conclusiones	93
8.2	Trabajo futuro	96
	BIBLIOGRAFÍA	97
A	ANÁLISIS DE HERRAMIENTAS	99
A.1	Características de las herramientas	99

A.2	Herramientas analizadas	99
A.3	Evaluación de las herramientas	101
A.4	Conclusiones	102
B	EJERCICIOS ESTÁNDARES DEL PSP	103
C	ANTECEDENTES DE SLR SOBRE PSP	105
D	PROTOCOLO DE TRABAJO 1.0	107
D.1	Identificar la necesidad de la revisión	107
D.2	Especificación de las preguntas de la SLR	107
D.2.1	Preguntas de investigación (PI)	107
D.2.2	Definición del alcance	108
D.3	Estrategia de búsqueda	108
D.3.1	Cadena de búsqueda	108
D.3.2	Recursos de búsqueda	109
D.3.3	Criterios de selección	109
D.3.4	Proceso de selección	110
D.3.5	Extracción de datos	111
D.3.6	Síntesis de datos extraídos	112
D.3.7	Cronograma de la revisión	113



## ÍNDICE DE FIGURAS

---

Figura 1	Versiones del PSP	16
Figura 2	Fases del PSP	17
Figura 3	Proceso de la SLR	22
Figura 4	Proceso para obtener el protocolo de la SLR.	33
Figura 5	Etapas del proceso de selección.	36
Figura 6	Detalle de la etapa de selección por título, resumen y conclusiones.	37
Figura 7	Detalle de la etapa de selección por lectura completa.	38
Figura 8	Proceso de selección por estudiante.	45
Figura 9	Métricas presentadas en cada estudio y sus ocurrencias.	48
Figura 10	Proporción de estrategias empíricas utilizadas en los estudios incluidos.	52
Figura 11	Proporción de estudios que utilizan métodos estadísticos y aquellos que no.	53
Figura 12	Cantidad de ocurrencias de los métodos estadísticos.	55
Figura 13	Proporción de estudios incluidos y excluidos.	56
Figura 14	Cantidad de estudios incluidos y totales por año.	57
Figura 15	Proporción de medios de publicación.	57
Figura 16	Proporción de contexto de publicación de los estudios incluidos.	58
Figura 17	Cantidad de estudios agrupados por proceso utilizado por año.	59
Figura 18	Proceso de selección por estudiante de estudios que utilizan investigación-acción	63
Figura 19	Proceso de selección por estudiante de estudios que utilizan encuesta	64
Figura 20	Proceso de selección por estudiante de estudios que utilizan experimento	65
Figura 21	Proceso de selección por estudiante de estudios que utilizan caso de estudio	66
Figura 22	Métricas presentadas en cada estudio y sus ocurrencias.	69
Figura 23	Proporción de estrategias empíricas utilizadas en los estudios incluidos.	72

Figura 24	Proporción de estudios que utilizan métodos estadísticos y aquellos que no.	73
Figura 25	Cantidad de ocurrencias de los métodos estadísticos.	75
Figura 26	Proporción de estudios incluidos y excluidos.	75
Figura 27	Cantidad de estudios incluidos y totales por año.	76
Figura 28	Proporción de medios de publicación.	76
Figura 29	Proporción de contexto de publicación de los estudios incluidos.	77
Figura 30	Cantidad de estudios agrupados por proceso utilizado por año.	78
Figura 31	Estudios incluidos por etapa del proceso de selección.	82
Figura 32	Proporción de estudios incluidos y excluidos.	86
Figura 33	Cantidad de estudios incluidos y totales por año.	88
Figura 34	Proporción de medios de publicación.	88
Figura 35	Etapas del proceso de selección	110
Figura 36	Detalle de la etapa de selección por título, resumen y conclusiones.	111
Figura 37	Detalle de la etapa de selección por lectura completa.	112

## ÍNDICE DE CUADROS

---

Cuadro 1	Información a ser documentada para cada fuente de información. 28
Cuadro 2	Ejemplo de presentación de datos sobre métodos estadísticos. 40
Cuadro 3	Estudios incluidos luego del proceso de selección. 47
Cuadro 4	Métricas presentadas en cada estudio y sus ocurrencias. 49
Cuadro 5	Métodos estadístico empleados por métrica. 54
Cuadro 6	Métodos estadísticos utilizados por estudio. 55
Cuadro 7	Fuente de publicación en los estudios incluidos. 56
Cuadro 8	Instituciones participantes en los estudios incluidos. 58
Cuadro 9	Cadenas de búsqueda utilizadas y su respectiva cantidad de estudios arrojados por el buscador. 62
Cuadro 10	Estudios incluidos luego del proceso de selección. 67
Cuadro 11	Métricas presentadas en cada estudio y su cantidad de ocurrencias 70
Cuadro 12	Métodos estadístico empleados por métrica. 74
Cuadro 13	Métodos estadísticos utilizados por estudio. 74
Cuadro 14	Fuente de publicación de los estudios incluidos. 74
Cuadro 15	Instituciones participantes en los estudios incluidos. 77
Cuadro 16	Estudios incluidos luego del proceso de selección 85
Cuadro 17	Fuente de publicación en los estudios incluidos. 87
Cuadro 18	Instituciones participantes en los estudios incluidos. 91
Cuadro 19	Tabla comparativa de herramientas. 102
Cuadro 20	Ejercicios propuestos para el curso de PSP. 103
Cuadro 21	Ejemplo de presentación de datos sobre métodos estadísticos. 113



## INTRODUCCIÓN

---

La revolución de la información y la tecnología ha significado, entre otras cosas, que el software haya pasado a formar parte de una gran cantidad de productos. El software se encuentra en productos desde tostadoras a transbordadores espaciales. Esto significa que cada vez más software es desarrollado. Este desarrollo no es para nada fácil, de hecho es un proceso altamente creativo. El rápido crecimiento del desarrollo de software ha significado que numerosos proyectos de software se hayan encontrado con problemas en términos de funcionalidades faltantes, exceso de costos, plazos de entrega y baja calidad de los productos [Wohlin et al., 2012].

Aunque la mayoría de las organizaciones industriales han adoptado principios de calidad, la comunidad del software continúa confiando en el *testing* como el principal método de gestión de calidad. Un paso significativo en la mejora de la calidad del software fue dado con el Personal Software Process (PSP). Este proceso se centra en las prácticas de trabajo a nivel individual. El principio detrás de PSP es que para producir software de calidad, cada ingeniero involucrado en el sistema debe realizar un trabajo de calidad. Para producir sistemas de software de alta calidad, cada parte del mismo debe ser de alta calidad. Esto significa que todos los ingenieros deben encargarse de la calidad de su trabajo personal. Para alcanzar esto, el PSP guía a los ingenieros en el proceso de registro y gestión de los defectos de su trabajo. El PSP describe cómo planear y registrar el trabajo, cómo usar un proceso medido y definido, cómo establecer objetivos medibles y cómo registrar el rendimiento en relación a estos objetivos. El PSP indica cómo manejar la calidad desde el principio del desarrollo del producto, cómo analizar los resultados y cómo utilizar dichos resultados para mejorar el proceso a futuro [Humphrey, 2000].

Este trabajo pretende observar los beneficios conocidos del Personal Software Process en el desarrollo de software. Dado que abarcar y procesar toda la literatura relacionada al PSP es un desafío, se plantea realizar una revisión sistemática de la literatura (SLR, por sus siglas en inglés “systematic literature review”). Una SLR es un método para identificar, evaluar e interpretar toda información relevante disponible correspondiente a una serie de preguntas de investigación específicas, área de investigación o fenómeno de interés [Kitchenham and Charters, 2007].

### 1.1 OBJETIVOS

El objetivo principal de este trabajo es recopilar los beneficios conocidos del PSP en el desarrollo de software. Para alcanzar dicho objetivo, se considera un enfoque de investigación basado en evidencias. La ingeniería de software empírica (ISE) es el área de investigación que promueve la utilización del método científico experimental para la generación de conocimiento sobre el proceso de desarrollo de software. Una de las herramientas principales para abordar la ISE son las revisiones sistemáticas de literatura [Kitchenham and Charters, 2007].

Por esto, se decide realizar una SLR con el objetivo de conocer la evidencia existente hasta la fecha acerca de los beneficios del PSP. A su vez, este trabajo pretende evidenciar las dificultades enfrentadas en el proceso para servir como referencia para otros estudiantes o investigadores.

A continuación se presentan los objetivos específicos de este trabajo:

- Estudiar el método de SLR aplicada a la ingeniería de software.
- Definir y validar un protocolo de trabajo para una SLR sobre los beneficios del PSP en el desarrollo de software.
- Ejecutar la SLR anteriormente mencionada.

### 1.2 TRABAJO REALIZADO

Para llevar a cabo la SLR fue necesario realizar un estudio sobre revisiones sistemáticas de la literatura como método de investigación empírica. Una SLR es un método para identificar, evaluar e interpretar toda información relevante disponible correspondiente a una serie de preguntas de investigación específicas, área de investigación o fenómeno de interés [Kitchenham and Charters, 2007]. Se elaboró una guía de trabajo para revisiones sistemáticas adaptada a los objetivos de este trabajo a partir de la guía genérica propuesta por Kitchenham en [Kitchenham and Charters, 2007]. A partir de la guía de trabajo se elaboró un protocolo de trabajo donde se especifican los métodos que se utilizan para llevar a cabo la SLR. Es necesario que este sea predefinido para así reducir la posibilidad de sesgo de la investigación. En el mismo se especifica la motivación para realizar la SLR, las preguntas de investigación, la estrategia de búsqueda, los criterios de inclusión/exclusión, la extracción y síntesis de datos. El protocolo de trabajo se validó mediante la ejecución de dos experiencias pilotos. Luego de la ejecución de la primera experiencia piloto (piloto 1) se realizaron modificaciones al protocolo de trabajo. De esta forma se obtuvo el protocolo 2 y se ejecutó

una segunda experiencia piloto (piloto 2) para contemplar la mayor cantidad de estrategias empíricas posibles (encuesta, casos de estudio, investigación-acción y experimento). Una vez finalizado el piloto 2, se obtuvo la versión final del protocolo. Por motivos de extensión de tiempo del proyecto de grado, no se ejecutó por completo la SLR, sino que se alcanzó hasta finalizar la ejecución del proceso de selección. Del mismo se obtuvo un total de 38 estudios incluidos. A partir del proceso de selección realizado y de las dos experiencias piloto, se respondieron las preguntas de investigación planteadas en la SLR. Además, se realizó un análisis de herramientas disponibles para dar soporte a las SLR (el cual se encuentra en el anexo A)

### 1.3 ESTRUCTURA DEL DOCUMENTO

Este trabajo está organizado de la siguiente manera. En los capítulos 2 y 3 se presenta un marco teórico sobre los conceptos de ingeniería de software empírica y PSP, respectivamente. Luego, en el capítulo 4 se define una guía de trabajo para revisiones sistemáticas adaptada a los objetivos de este trabajo. En el capítulo 5 se presenta el protocolo de trabajo definido tras las experiencias piloto realizadas, las cuales se detallan en el capítulo 6. A continuación, en el capítulo 7 se presenta parte de la aplicación del protocolo anteriormente mencionado. Por último, en el capítulo 8 se presentan las conclusiones y las principales líneas de trabajo futuro. En el anexo A se presenta un análisis de herramientas para dar soporte a las SLR que fueron consideradas al momento de realizar este trabajo. En el anexo B se mencionan los ejercicios estándar del PSP. En el anexo C se presenta una breve descripción de la búsqueda de antecedentes de SLR sobre PSP. Por último, en el anexo D se expone la versión intermedia del protocolo obtenido tras la realización del piloto 1.



## MARCO TEÓRICO SOBRE INGENIERÍA DE SOFTWARE EMPÍRICA

---

En este capítulo se presentan conceptos de Ingeniería de Software Empírica y Revisiones Sistemáticas de la Literatura. Se basa en Guidelines for performing Systematic Literature Reviews in Software Engineering [Kitchenham and Charters, 2007], Métodos de investigación de ingeniería de software [Genero et al., 2014], Conceptos de Ingeniería de Software Empírica [Apa et al., 2010] y Experimentation in Software engineering [Wohlin et al., 2012].

### 2.1 CONCEPTOS DE INGENIERÍA DE SOFTWARE EMPÍRICA

En esta sección se contextualiza la investigación en el área de la Ingeniería de Software empírica y se mencionan los diferentes métodos de investigación. A su vez, se describen las diferentes estrategias empíricas, profundizando en experimentos controlados <sup>1</sup>. Es importante aclarar que de ahora en más se hará referencia a experimentos controlados como experimentos.

#### 2.1.1 Contexto y estrategias

Comprender una disciplina implica aprendizaje, es decir, observación, reflexión y encapsulación de conocimiento, construcción de modelos, experimentación y evolución de los modelos con el tiempo. Desde un punto de vista científico, la investigación en ingeniería de software se centra en conocer la naturaleza de los procesos, de los productos y las interrelaciones entre ellos. Para esto es necesario emplear diferentes métodos de investigación.

La experimentación es un método que se usa para corresponder ideas o teorías con la realidad, proporcionando evidencia que soportan las hipótesis o suposiciones que se creen válidas. La Ingeniería de Software Empírica, se basa en dicho método y se enfoca en mostrar con hechos las especulaciones, suposiciones y creencias existentes sobre la construcción del software.

---

<sup>1</sup> Inicialmente se optó por profundizar en experimentos controlados ya que se pretendía incluir únicamente estudios que utilizaran dicha experiencia empírica. Más adelante en la ejecución de la SLR se decidió ampliar esta condición.

Los resultados obtenidos tras una investigación dependen del entorno, de la configuración y de los objetivos de los proyectos estudiados. Ya que no es posible obtener resultados válidos universalmente, se debe determinar en qué contexto concreto son aplican. Un claro ejemplo es el de la industria del software. Las distintas organizaciones productoras de software difieren en tamaño, naturaleza del cliente, localidad y cultura organizacional, entre otros factores. Todos estos aspectos influyen en el software producido. Por lo tanto, reconocer estos aspectos y la manera en la que se interrelacionan es importante para la utilización eficaz de la investigación en la práctica. Sin embargo, la naturaleza de estas relaciones es poco conocida y en consecuencia, no podemos asumir a priori que los resultados de un estudio en particular sean aplicables fuera del contexto específico en el que fue llevado a cabo.

Los métodos de investigación se pueden clasificar de diversas maneras. A continuación se muestra un tipo de clasificación, en función del nivel de evidencia que proporcionan.

- **Métodos primarios:** aquellos métodos utilizados para realizar estudios con el objetivo de obtener evidencia empírica sobre un tema de interés. Por consiguiente, los estudios llevados a cabo utilizando estos métodos se los denomina estudios primarios.
- **Métodos secundarios:** refiere a los métodos que permiten recopilar de manera sistemática y rigurosa los estudios primarios relacionados con una pregunta de investigación específica, con el objetivo de sintetizar la evidencia disponible para responder a dicha pregunta. En ingeniería de software se utilizan las revisiones y mapeos sistemáticos de la literatura como métodos de investigación secundarios. Este tipo de métodos se utilizan para producir estudios secundarios.

### 2.1.2 *Estrategias empíricas*

Existen 4 tipos principales de técnicas o estrategias de la investigación empírica: encuesta, caso de estudio, experimento e investigación-acción.

#### **Encuesta**

Una encuesta es un método empírico que se utiliza para recopilar información de o sobre personas para describir, comparar o explicar su conocimiento, sus actitudes o su comportamiento. También se pueden utilizar encuestas para la descripción de las características de métodos o herramientas. Son investigacio-

nes que proporcionan una visión general, mediante la recolección de información estandarizada de una población específica o una muestra representativa de la misma. Esta recolección se realiza por medio de un cuestionario o entrevista, para luego generalizar los resultados al resto de la población.

Las encuestas se pueden utilizar ya sea cuando una técnica o herramienta ha sido utilizada o antes de comenzar a hacerlo. Para el caso de la ISE se utiliza este método para determinar cómo reacciona la población frente a una técnica, herramienta o método particular, o para determinar relaciones o tendencias. Es un proceso complejo formado por una serie de actividades bien definidas como son: establecer objetivos, diseñar la encuesta, desarrollar, evaluar y validar el cuestionario, obtener y analizar datos y reportar los resultados.

### **Caso de estudio**

Tradicionalmente se habla de caso de estudio como un método empírico que hace uso de múltiples fuentes de evidencia para investigar una instancia (o un pequeño número de instancias) de un fenómeno contemporáneo relacionado con la ingeniería del software. Este fenómeno puede presentarse dentro de su contexto real, específicamente cuando las fronteras entre el fenómeno y su contexto no pueden definirse claramente.

Los casos de estudio no dan lugar a relaciones causales como ocurre con otras técnicas de investigación (por ejemplo, los experimentos) sino que permiten comprender en mayor profundidad el fenómeno que se está estudiando. Precisamente ahí radica la principal funcionalidad de los casos de estudio: la capacidad de proporcionar resultados de investigación a partir de proyectos puestos en producción. Son estudios de campo u observaciones donde no hay aleatoriedad de variables ni representatividad de la muestra.

En un caso de estudio se identifican los factores clave que pueden afectar la salida de una actividad y se documentan las entradas, las limitaciones, los recursos y las salidas. El nivel de control de la ejecución es menor en los casos de estudio que en los experimentos, debido a que en los primeros no se controla, solo se observa. Las variables no se manipulan, sino que son determinadas por la situación que se está investigando, contrario a lo que ocurre en los experimentos.

### **Experimento**

Un experimento en ingeniería de software es una investigación empírica que manipula una variable del entorno o fenómeno estudiado midiendo el efecto que tiene sobre otra variable. Estos experimentos pueden ser “orientados a las personas” u

“orientados a la tecnología”. En los primeros, los sujetos aplican diferentes tratamientos a los objetos; por ejemplo, dos métodos de inspección se aplican sobre dos códigos fuente. En los “orientados a la tecnología”, generalmente se aplican diferentes herramientas a diferentes objetos, por ejemplo al mismo programa se le aplican dos técnicas de generación de casos de prueba diferentes.

Los experimentos pueden ser utilizados para confirmar el conocimiento convencional, explorar relaciones entre sucesos, evaluar la precisión de modelos, validar métricas, etc. En todas estas aplicaciones de los experimentos lo que se persigue es probar una hipótesis, ya que si los resultados de un experimento contradicen una hipótesis, la misma puede rechazarse.

Los experimentos son generalmente ejecutados en un ambiente de laboratorio, el cual brinda un alto grado de control, manipulando una o más variables y controlando el resto. Es una técnica formal, rigurosa y controlada. Tiene un alto costo en relación a las otras técnicas de investigación, pero a cambio ofrece un control total de la ejecución y es de fácil replicación.

Los objetos sobre los cuales se ejecuta el experimento son llamados *unidades experimentales* u *objetos experimentales*. Aquellas personas que aplican los métodos, o técnicas, a las unidades experimentales, son llamadas *sujetos experimentales*. Finalmente, el resultado de un experimento es llamado *variable de respuesta*, o *variable dependiente*, y es un valor cuantitativo.

Un *parámetro* es cualquier característica que permanezca invariable a lo largo del experimento. Los mismos no deben influir en el resultado del experimento. Los resultados serán particulares a las condiciones definidas por los parámetros. Por otra parte, cada característica del desarrollo de software a ser estudiada y que afecte las variables de respuesta se denomina *factor*. Los factores son intencionalmente modificados durante el experimento. A su vez, los posibles valores que puede tomar cada factor en cada unidad experimental reciben el nombre de *alternativas* o *niveles*.

Cada ejecución del experimento que se realiza en una unidad experimental es llamado *experimento unitario*. Por lo tanto, cada aplicación de una combinación de alternativas de factores por un sujeto experimental en una unidad experimental es un experimento unitario.

Existen tres principios generales de diseño que deben ser tenidos en cuenta. A continuación se describe en qué consiste cada uno. Todos los métodos estadísticos requieren que las observaciones sean de variables independientes aleatorias. Por lo que, tanto las alternativas de los factores como los sujetos tienen que ser elegidos de forma aleatoria. Este principio recibe el nombre de *aleatoriedad*. El principio de *bloqueo* se utiliza cuando se tienen factores que probablemente tengan efectos indeseados en las variables de respuesta y éstos efectos son conocidos

y controlados. Como último principio se encuentra el *balance*. Dicho principio consiste en que la cantidad de unidades experimentales sobre las que se ejecuta el experimento debe ser la misma que a las que no. Esto simplifica y fortalece el análisis estadístico de los datos, aunque no es indispensable.

### Fases de ejecución de un experimento

A continuación se describen brevemente las diferentes fases a realizar durante la ejecución de un experimento.

La primer fase del proceso es la de *definición*, donde se define el problema que se quiere resolver, los objetivos y las metas del experimento. Se define el *objeto de estudio* y el *propósito*, siendo estos la entidad que va a ser estudiada y la intención del mismo, respectivamente. Por último, se debe definir el *contexto*, que es el ambiente en el que se ejecuta el experimento.

La *planificación* es la fase en la que se define como se va a llevar a cabo el experimento, especificando claramente las características del ambiente donde se ejecuta el experimento (por ejemplo, si se va a realizar un proyecto real - *on-line*, o en un laboratorio - *off-line*), las características de los sujetos y si el problema es “real” (existente en la industria) o “de juguete”. Se debe definir si el experimento es válido para un contexto específico o para un dominio general de la Ingeniería de Software.

Usualmente se definen dos hipótesis: la *nula* y la *alternativa*. La primera, denotada por  $H_0$ , asume que no hay una diferencia significativa con respecto a la alternativa, con respecto a las variables dependientes que se están midiendo. La hipótesis nula establece que si hay diferencias entre las observaciones realizadas, éstas son por casualidad y no por producto de la alternativa aplicada. El foco del experimento está puesto en rechazarla. En cambio, la hipótesis alternativa  $H_1$  afirma que existe una diferencia entre las alternativas con respecto a las variables dependientes. Ésta es la hipótesis a probar.

Otro aspecto importante al llevar a cabo un experimento es la selección de los sujetos. Para poder generalizar los resultados al resto de la población, la selección debe ser una muestra representativa de la misma y cuanto mayor sea, menor será el error al generalizar los resultados. Por último, se debe elegir el diseño adecuado y, ya que el mismo afecta el análisis estadístico y viceversa. Por este motivo, se debe tener en cuenta qué análisis estadístico es el mejor para rechazar la hipótesis nula y para aceptar la alternativa.

Luego de diseñar y planificar el experimento, éste debe ser ejecutado para recolectar los datos que se desea analizar. Como primera etapa de la *fase de operación* se seleccionan los sujetos y se preparan los artefactos a ser utilizados. Es necesario

tener prontos los instrumentos, formularios, herramientas y demás para comenzar la ejecución. En cuanto a la recolección de datos, existen diferentes formas de hacerlo: manualmente (mediante formularios), entrevistas o automáticamente. Una vez obtenidos los datos, se debe verificar que fueron recolectados correctamente y que son razonables.

Una vez finalizada la ejecución, comienza la *fase de análisis de los datos* conforme a los objetivos planteados. El análisis se realiza en tres etapas: caracterizar el conjunto de datos usando estadística descriptiva, reducir el conjunto de datos y realizar las pruebas de hipótesis.

La estadística descriptiva es utilizada para entender mejor la naturaleza de los datos y para identificar datos falsos o anormales. Los aspectos fundamentales que examina son: la *tendencia central*, indicando la media de un conjunto de datos; la *dispersión*, midiendo el nivel de desviación de la tendencia central, es decir, que tan diseminados o concentrados están los datos respecto al valor central; y la dependencia entre variables.

La reducción del conjunto de datos es necesaria ya que el resultado de aplicar métodos estadísticos depende de la calidad de los datos. Si los datos no representan lo que se cree, las conclusiones que se derivan de los resultados de los métodos son incorrectas. Errores en el conjunto de datos pueden ocurrir por un error sistemático o por lo que se conoce en estadística con el nombre de *valor atípico* (outlier). Un outlier es un dato mucho más grande o mucho más chico de lo que se puede esperar observando el resto de los datos. En algunos casos se realiza un análisis acerca de estos valores que difieren mucho de la media y se decide quitarlos de los datos a analizar porque no son representativos de la población, ya que fueron causados por algún tipo de anomalía.

El objetivo de la prueba de hipótesis es ver si es posible rechazar cierta hipótesis nula  $H_0$ . Si la hipótesis nula no es rechazada, no se puede decir nada sobre los resultados. Esta prueba se realiza mediante métodos estadísticos. Los mismos son utilizados para validar o rechazar hipótesis sobre modelación probabilística. Tratan de distinguir lo que es razonable o probable de lo que es poco verosímil.

Existen dos tipos de métodos (tests): paramétricos y no paramétricos. Los tests paramétricos están basados en un modelo que involucra una distribución específica. En la mayoría de los casos, se asume que los parámetros involucrados en un test paramétrico están normalmente distribuidos. Si los parámetros no pueden medirse en al menos una escala de intervalo, generalmente no se puede utilizar un test paramétrico. En este caso hay un amplio rango de tests no paramétricos disponible.

Los tests no paramétricos no asumen ninguna distribución de probabilidad específica para los parámetros, por lo que son más generales que los paramétricos.

Un test no paramétrico se puede utilizar en vez de un test paramétrico, pero el caso inverso no siempre puede darse. En la elección entre un test paramétrico y un test no paramétrico hay dos aspectos a considerar:

- **Aplicabilidad:** es importante que tanto las suposiciones sobre las distribuciones de parámetros como las que conciernen a las escalas sean realistas.
- **Poder:** el poder de los tests paramétricos es generalmente mayor que el de los tests no paramétricos. Por lo tanto, los test paramétricos requieren menos datos (experimentos más pequeños), que los tests no paramétricos, siempre que sean aplicables.

Los test paramétricos más usados en experimentos de Ingeniería de Software son ANOVA (ANalysis Of VAriance) y ANOM (ANalysis Of Means). El objetivo principal del ANOVA es contrastar si existen diferencias entre las diferentes medias de las alternativas. En el caso de ANOM, este test no solamente responde a la pregunta de si hay o no diferencias entre las alternativas, sino que cuando hay diferencias, también dice cuáles alternativas son mejores y cuáles peores.

Al concluir el experimento se presentan las conclusiones y los hallazgos como un resumen de todo el experimento, junto con los resultados, problemas y desviaciones respecto al plan.

### Investigación-Acción

Es el proceso de recopilar de forma sistemática datos de una investigación acerca de un sistema actual en relación con algún objetivo, meta o necesidad de ese sistema; de alimentar de nuevo con esos datos al sistema; de emprender acciones por medio de variables alternativas seleccionadas dentro del sistema, basándose tanto en los datos como en las hipótesis; y de evaluar los resultados de las acciones, recopilando datos adicionales [French and Bell, 1996].

Por lo tanto, la investigación-acción (IA) es una forma de investigar de carácter colaborativo que busca unir teoría y práctica entre investigadores y profesionales. Esto se consigue gracias a la intervención de un investigador en la realidad de un grupo de profesionales. Los resultados de esta experiencia deben ser beneficiosos tanto para el investigador como para los practicantes.

Durante una IA se identifican los siguientes pasos:

- **Planificación:** donde se determinan las cuestiones relevantes que guiarán la investigación.
- **Acción:** donde se realiza una variación de la práctica, deliberada y controlada (es cuando el investigador interviene sobre la realidad).

- **Observación:** donde se recoge información, datos y se documenta lo que ocurre.
- **Reflexión:** donde se analizan los resultados de manera tal de realizar un nuevo planeamiento que pueda mejorar las prácticas, modificando éstas como parte del propio proceso investigador.

El proceso de una IA es iterativo, avanzando hacia soluciones cada vez más refinadas.

## 2.2 REVISIONES SISTEMÁTICAS DE LA LITERATURA

Una revisión sistemática de la literatura es un medio para identificar, evaluar e interpretar toda la información disponible sobre cierta temática y responder ciertas preguntas de investigación específicas. Una SLR se considera un estudio secundario y cada uno de los estudios individuales recopilados en una SLR son estudios primarios.

Las SLR difieren principalmente de las revisiones de la literatura tradicionales en que se planifican formalmente y efectúan de manera sistemática y metódica. Una revisión sistemática debe ser replicable de forma independiente, por lo que posee mayor valor científico que una revisión tradicional. Al encontrar, evaluar y resumir toda la evidencia disponible sobre un tema específico de investigación, una SLR puede proporcionar un mayor nivel de validez de sus conclusiones que las que proveen cualquiera de los estudios primarios analizados en la SLR.

Existen diversas razones para realizar una SLR, a continuación se listan las principales:

- Resumir la evidencia existente en relación a una tecnología o método.
- Identificar “vacíos” en investigaciones contemporáneas de forma de sugerir áreas en las que se debería realizar mayor investigación.
- Proveer un contexto de manera de situar apropiadamente nuevas actividades de investigación.

Por otra parte, las revisiones sistemáticas también pueden utilizarse tanto para examinar el grado con el cual ciertas hipótesis teóricas son respaldadas (o no) por evidencia empírica, como para generar nuevas hipótesis.

A continuación se mencionan algunas de las características de las revisiones sistemáticas que las diferencian de una tradicional son:

- Es necesario definir un protocolo que especifique las preguntas de investigación que se buscan responder y los métodos que se utilizarán para realizarla.

- Se basan en una estrategia de búsqueda definida que pretende detectar la mayor cantidad de literatura relevante que sea posible.
- Se debe documentar la estrategia de búsqueda de forma que los lectores puedan, por un lado evaluar la rigurosidad y completitud del proceso, y por el otro replicar el proceso de revisión.
- Requieren criterios de inclusión y exclusión explícitos para evaluar cada potencial estudio primario.

Las SLR tienen numerosas ventajas, por ejemplo su metodología bien definida, la cual disminuye la probabilidad de sesgo (aunque existe la posibilidad de que los estudios primarios con los que se trabaja sí estén sesgados). A su vez, proveen información sobre los efectos de ciertos fenómenos sobre un amplio rango de configuraciones y métodos empíricos por lo que, si los estudios presentan resultados consistentes, las revisiones proveen evidencia de la robustez y validez de dicho fenómeno. En el caso de estudios cuantitativos es posible combinar datos, aumentando la probabilidad de detectar efectos reales que estudios individuales más pequeños no pudieron detectar.

La principal desventaja es que requieren más tiempo y esfuerzo que una revisión tradicional, por lo que uno de los principales retos de la realización de las SLR es lograr un equilibrio entre el rigor metodológico y el esfuerzo requerido.



## MARCO TEÓRICO SOBRE PERSONAL SOFTWARE PROCESS

---

En este capítulo se presentan conceptos del Personal Software Process. Está basado en The Personal Software Process [Humphrey, 2000].

El Personal Software Process (PSP) es un framework bien definido y de carácter individual para ser utilizado por ingenieros al desarrollar software. El proceso consiste en una serie de métodos, formularios y scripts que ayudan a planificar, medir y administrar el trabajo. Está diseñado para ser utilizado con cualquier lenguaje de programación, y puede aplicarse tanto para relevar requerimientos como para realizar tareas de verificación, reparación de errores o definición de procesos.

PSP se centra en las prácticas de trabajo de los ingenieros a nivel individual. El principio detrás de esto es que para producir software de calidad, cada ingeniero involucrado en el sistema debe realizar un trabajo de calidad.

El PSP presenta la forma de planear y registrar el trabajo, mediante un proceso bien definido, estableciendo metas medibles y registrando el rendimiento en relación a estas metas. Además, muestra cómo analizar los resultados de cada trabajo y cómo utilizarlos para mejorar el proceso en el siguiente proyecto.

El diseño del PSP se encuentra basado en los siguientes principios:

- Cada ingeniero es diferente, por lo que cada cual debe planificar su propio trabajo y debe basar sus planes en sus propios datos personales.
- Para mejorar el rendimiento consistentemente, se deben utilizar procesos medibles y bien definidos a nivel personal.
- Para producir un producto de calidad, todos los ingenieros deben sentirse personalmente responsables por la calidad del producto.
- El costo de encontrar y arreglar un error detectado de forma temprana es menor a si se lo detecta avanzado el proyecto.
- Es más eficiente prevenir defectos que encontrarlos y arreglarlos.
- La manera correcta es siempre la más rápida y la menos costosa de realizar una tarea.

Este proceso utiliza una serie de scripts para guiar al desarrollador en su trabajo. Dichos scripts suelen ser cortos y presentan instrucciones precisas. Mientras los ingenieros siguen dicho script para realizar su trabajo, registran el tiempo y los defectos encontrados en un log. Al culminar una actividad, se mide el tamaño del programa y se ingresa esta información en un formulario de resumen. Al terminar, se entrega el producto finalizado, junto con un formulario de resumen completo.

Debido a que éste proceso cuenta con una serie de métodos que generalmente no han sido aplicados nunca por parte de los ingenieros, estos métodos son introducidos en una serie de siete versiones del proceso, como se muestra en la Figura 1. Estas versiones son etiquetadas desde el 0 al 3 y cada una define un conjunto de logs, formularios, scripts y estándares. Los scripts del proceso definen los pasos para cada parte del proceso, mientras que los logs y formularios proveen plantillas para registrar y almacenar datos respectivamente. Finalmente, los estándares guían al ingeniero mientras trabaja.

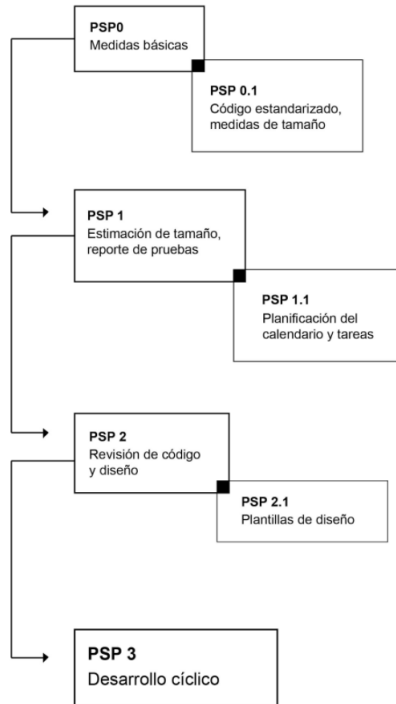


Figura 1: Versiones del PSP

## 3.1 ESTRUCTURA DEL PSP

Este proceso consta de tres grandes fases: *Planificación*, *Desarrollo* y *Postmortem* (Figura 2). La fase de Planificación tiene como entrada los requerimientos, elaborando a partir de ellos un plan de desarrollo. La fase de desarrollo engloba cuatro grandes etapas: Diseño (y su correspondiente revisión), Codificación (y su correspondiente revisión), Compilación y Pruebas. Durante esta fase se realiza la construcción del software y se registran diferentes métricas de tiempo, tamaño y calidad. Finalmente, en la fase de Postmortem, se realiza un resumen analizando las métricas registradas para una futura planificación y mejora. Dicho resumen es entregado en conjunto con el producto final.

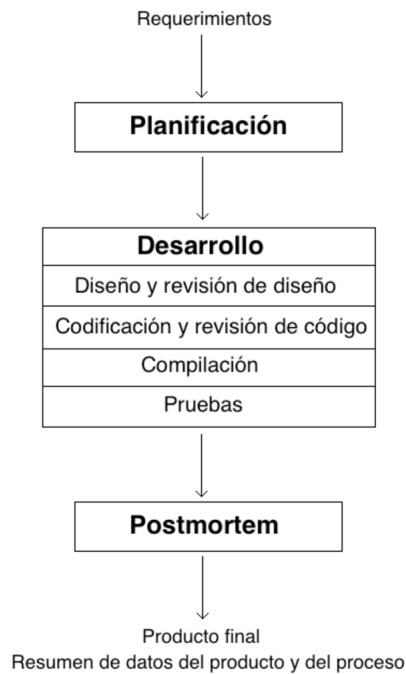


Figura 2: Fases del PSP

## 3.2 MÉTRICAS

El PSP presenta diferentes maneras de medir, evaluar y administrar la calidad de los programas, ayudando así a examinar la calidad de los mismos desde diferentes perspectivas. Mientras que cada una por separado no pueden indicar adecuadamente la calidad general del programa, el agregado de todas provee un indicador de calidad considerablemente confiable.

### 3.2.1 *Métricas de tiempo*

Para registrar el tiempo, se utiliza un log, donde se anota el tiempo en el que se comenzó a realizar una tarea, el tiempo en el que se detiene el desarrollo de la tarea y cualquier lapso de tiempo invertido en interrupciones. Al realizar un registro preciso del tiempo, se puede ver el esfuerzo real invertido en cada tarea del proyecto ya que considerando que los tiempos de interrupción son esencialmente aleatorios, ignorar este tiempo podría representar un gran error - aleatorio - en los datos y reduciría la exactitud de la estimación.

### 3.2.2 *Métricas de tamaño*

Por otra parte, se debe estimar el tamaño del producto que se planea desarrollar, ya que el tiempo que llevará realizarlo está determinado en gran parte por el tamaño del producto. Una vez finalizado el producto, se mide el tamaño del mismo, utilizándolo para producir estimaciones más exactas a futuro. Sin embargo, para que esto sea útil, la medida de tamaño y de tiempo de desarrollo deben poder relacionarse. Si bien líneas de código (LOC) es la principal métrica de tamaño de PSP, cualquier métrica que provea una relación razonable entre tiempo de desarrollo y tamaño de producto puede ser utilizada.

**Líneas de código (LOC):** es la cantidad de líneas de código que tiene el programa, PSP utiliza el concepto líneas de código lógicas para realizar esta medición. Dado que es un concepto que puede ser definido de diferentes maneras, los ingenieros deben definir precisamente de qué forma desean medir las LOC.

### 3.2.3 *Métricas de calidad*

En cuanto al manejo de defectos, se necesita información sobre los defectos existentes, las fases en las cuales fueron inyectados, las fases en las que fueron encontrados y arreglados, y cuánto tiempo tomó hacerlo. Al utilizar PSP, esta información se registra para cada defecto encontrado en el log correspondiente. Un defecto puede tratarse de un error ortográfico, un problema de puntuación o un código incorrecto. Pueden presentarse en el programa, en el diseño, o hasta en los requerimientos o cualquier tipo de documentación.

Como se mencionó anteriormente, el objetivo principal es encontrar y solucionar defectos antes de realizar la primera compilación o prueba unitaria. El proceso del PSP incluye pasos en la fase de diseño y de revisión de código en el que cada ingeniero debe inspeccionar su trabajo antes de que sea inspeccionado por otro, de que sea compilado e incluso testeado. El concepto detrás de esto es que las personas suelen realizar el mismo error repetidas veces, por lo que analizando la información respecto a los defectos que han sido cometidos anteriormente y definiendo las acciones requeridas para encontrarlos, se puede detectar y solucionar defectos con mayor eficiencia.

A continuación se mencionan y describen las principales métricas de calidad utilizadas en PSP:

**Densidad de defectos (DDS):** refiere a la cantidad de defectos encontrados cada mil líneas de código (KLOC) nuevas o modificadas de un programa. Esta métrica se utiliza tanto para el proceso entero como para cada fase del mismo.

**Tasa de revisión:** es la cantidad de LOC revisadas por hora. En PSP, los ingenieros recolectan información de las revisiones con el objetivo de determinar qué tan rápido deben revisar sus programas con el fin de encontrar todos o casi todos los defectos.

**Proporción de tiempo de desarrollo:** refiere a la proporción de tiempo utilizado por un ingeniero entre dos fases de desarrollo. Se utilizan 3 proporciones de tiempo de desarrollo: tiempo de diseño sobre tiempo de codificación, tiempo de revisión de diseño sobre tiempo de diseño y tiempo de revisión de código sobre tiempo de codificación. El porcentaje de tiempo de diseño sobre tiempo de codificación es el más útil para determinar la calidad del diseño. Cuando los ingenieros pasan menos tiempo en diseño que en codificación, están realizando la mayoría del diseño mientras codifican. Esto significa que el diseño no está documentado y que no puede ser revisado y/o inspeccionado, por lo tanto, seguramente el diseño sea pobre.

**Proporción de defectos:** compara los defectos encontrados en una fase con los defectos encontrados en otra fase. Las principales proporciones de defectos son los defectos encontrados en revisión de código sobre los defectos encontrados en compilación; y los defectos encontrados en revisión de diseño sobre los defectos en pruebas unitarias. La proporción de defectos encontrados en revisión de código sobre los defectos encontrados en compilación es una métrica de calidad de código. Cuando un ingeniero encuentra más del doble de defectos en la etapa de revisión de código que en la de compilación, generalmente significa que realizó una buena revisión de código o que no encontró todos los defectos en la fase de compilación. Por otra parte, el porcentaje de defectos encontrados en revisión de diseño sobre los defectos encontrados en pruebas unitarias es una métrica de calidad de diseño. Si un ingeniero encuentra el doble de defectos en la etapa de revisión de diseño que en la de pruebas unitarias, probablemente haya hecho una aceptable revisión de diseño.

**Yield:** es medido de dos formas, el yield de una fase mide el porcentaje de defectos que son encontrados y removidos en la fase. Para calcular dicho porcentaje se toma como defectos totales la cantidad de defectos existentes antes de comenzar la fase. Por su parte, el yield del proceso mide el porcentaje de defectos removidos antes de la primera compilación.

**Defectos por hora (DRR):** a partir de los datos recolectados por medio de PSP, los ingenieros pueden calcular la cantidad de defectos inyectados por hora y cantidad de defectos removidos por hora. Con dichas métricas, el ingeniero puede planificar cuanto tiempo debe asignar a la revisión de código a partir del tiempo que le llevó el desarrollo. Por ejemplo, si un ingeniero inyecta 4 defectos por hora y remueve 8 defectos por hora, sabe que debe pasar 30 minutos en revisión de código por hora de codificación.

**A/FR:** se calcula dividiendo el tiempo de revisión de diseño y código por el tiempo de compilación y pruebas. A/FR provee una buena forma de evaluar la calidad, indica el grado con el cual el ingeniero intento encontrar y corregir defectos de forma temprana.

En este capítulo se presenta una guía para ejecutar la SLR. En la misma se presentan los lineamientos generales seguidos en este trabajo con el objetivo de introducir la metodología utilizada en la SLR. Es una adaptación del artículo Guidelines for performing Systematic Literature Reviews in Software Engineering [Kitchenham and Charters, 2007]. Además, se basa en la presentación realizada por Fernando Acerenza [Acerenza, 2015] en el marco de los Seminarios de Ingeniería de Software en la Facultad de Ingeniería, UdelaR.

#### 4.1 FASES DE LA INVESTIGACIÓN

A continuación se presenta, a modo de esquema, las fases que se deben seguir al realizar una SLR:

1. Planificar la investigación
  - 1.1. Identificar la necesidad de la SLR
  - 1.2. Puesta en marcha
  - 1.3. Especificar las preguntas de investigación
  - 1.4. Desarrollar el protocolo de trabajo
  - 1.5. Evaluar el protocolo
2. Ejecutar la investigación
  - 2.1. Identificación de la investigación
  - 2.2. Selección de estudios primarios
  - 2.3. Estudio de calidad de los estudios primarios
  - 2.4. Extracción de datos
  - 2.5. Síntesis de datos.
3. Reportar la investigación
  - 3.1. Especificación de los mecanismos de diseminación
  - 3.2. Formateo del reporte final

### 3.3. Evaluación del reporte final

Todas estas actividades se consideran de carácter obligatorio e imprescindibles a excepción de las actividades *Puesta en marcha de la SLR* y *Evaluación del reporte final*.

Debido a que esta investigación no tiene un carácter comercial se obviara la primera actividad del primer punto mencionado (Puesta en marcha de la SLR). Esto se debe a que durante la misma se realiza un documento de especificación del trabajo requerido en los casos en que una organización desee realizar una investigación pero no cuente con el tiempo o los recursos necesarios para hacerlo. Dicho documento es luego entregado a los investigadores a cargo de conducirla. A su vez, Kitchenham menciona en su guía [Kitchenham and Charters, 2007] que no existen evidencias de SLR que ejecuten este paso en el área de la ingeniería de software.

Respecto al *estudio de calidad de los estudios primarios*, se optó por no realizarlo ya que se busca cubrir la mayor cantidad de estudios en esta investigación. Por lo tanto no se busca descartar algunos de ellos por no cumplir ciertos parámetros de calidad.

Por otra parte, al tratarse de un proyecto de grado, la evaluación del reporte quedará a cargo de un tribunal. Por esta misma razón, no se pretende la publicación de la investigación en ningún medio, por lo que no será necesaria la etapa de *especificación de los mecanismos de diseminación*. La especificación de los mecanismos de diseminación consta de la realización de posters, páginas web, folletos, artículos de prensa, entre otras.

La Figura 3 presenta el proceso que se sigue en esta SLR.

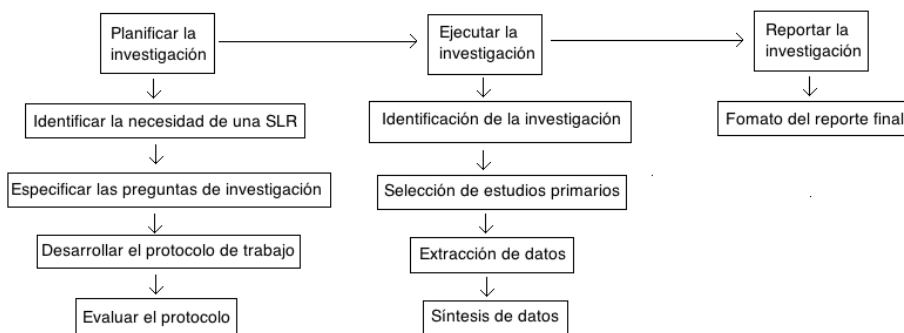


Figura 3: Proceso de la SLR

## 4.2 PLANIFICACIÓN DE LA INVESTIGACIÓN

### 4.2.1 *Identificar la necesidad de una SLR*

Al realizar una revisión sistemática, los investigadores deben asegurarse de que es necesario realizar la misma. Para esto, se deberá identificar y evaluar cualquier SLR sobre el fenómeno de interés. En particular, en la guía de la CRD (Centre for Reviews and Dissemination) [Khan et al., 2001] se establecen ciertas preguntas a tener en cuenta a la hora de evaluar las SLR existentes. Las principales a tener en cuenta son:

- ¿Cuáles son los objetivos de la investigación?
- ¿Qué fuentes fueron utilizadas para obtener los estudios primarios?
- ¿Qué criterios de calidad fueron utilizados para evaluar dichos estudios?

Dichas preguntas deben ser respondidas para cada SLR encontrada y comparadas contra la que se pretende realizar. De esta forma se puede evaluar si ya existe una investigación que contemple las inquietudes presentadas anteriormente.

### 4.2.2 *Especificar las preguntas de investigación (PI)*

Las preguntas de investigación son una parte central en la investigación debido a que dirigen el curso de la misma, ya que:

- El proceso de búsqueda debe identificar estudios primarios que respondan las PI.
- La extracción de datos debe asegurarse de extraer los datos necesarios para responder dichas preguntas.
- El análisis de los datos debe hacerse de tal forma que las preguntas sean respondidas.

A continuación se presentan seis tipos de preguntas genéricas sobre Ingeniería de Software (IS) que se pueden responder mediante una SLR:

- ¿Cuáles son los efectos de cierta tecnología?
- ¿Cuál es la frecuencia de éxito o fracaso de un proyecto?
- ¿Cuál es el costo y los factores de riesgo asociados a cierta tecnología?

- ¿Cuál el impacto de cierta tecnología en cuanto a confiabilidad, performance y costo?
- ¿Cuál es la relación costo/beneficio tras emplear ciertas tecnologías o aplicaciones?

### Estructura de las PI

Generalmente, se suelen definir las PI a partir de los siguientes conceptos:

- **Población:** en los experimentos de IS, la población puede ser cualquiera de los siguientes:
  - Un rol específico de IS (testers, managers, etc.)
  - Una categoría para un ingeniero de software (senior, junior, etc.)
  - Un área de aplicación (sistemas de información, sistemas de control, etc.)
  - Un grupo de la industria (compañías de telecomunicación, compañías de IT, etc.)
- **Intervención:** metodología, herramienta, tecnología o procedimiento que se utilizan para resolver algo específico, por ejemplo, tecnologías para realizar tareas específicas como especificación de requerimientos, testing o estimación de costos.
- **Comparación:** metodología, herramienta, tecnología o procedimiento contra el que la intervención será comparada. Normalmente el grupo contra el que la comparación se realizará recibe el nombre de grupo de control, mientras que al intervenido se le otorga el nombre de grupo intervenido. En particular, “no utilizar la intervención” es inadecuado como comparación, ya que las técnicas de ingeniería de software usualmente requieren entrenamiento, por lo que el efecto de la técnica puede ser confundido con el efecto del entrenamiento.
- **Salida:** la salida del estudio debe relacionarse con factores de importancia para los involucrados como incremento de la confiabilidad, reducción en los costos de producción y reducción del *tiempo de salida al mercado*.
- **Contexto:** contexto en el cual toma lugar la comparación (por ejemplo, academia o industria), los participantes del estudio (por ejemplo, estudiantes, académicos, etc.) y las tareas realizadas por los participantes (por ejemplo gran escala, pequeña escala).

- **Diseño experimental:** restricción de los estudios primarios a un tipo específico o un conjunto de los mismos. Esto normalmente se realiza en disciplinas, como la medicina, donde la numerosa cantidad de artículos es un problema importante. Sin embargo, este no es el caso de Ingeniería de Software por lo que, en general, no es necesario restringir.

#### 4.2.3 *Desarrollar el protocolo de trabajo*

El protocolo de trabajo debe especificar los métodos que se utilizan para llevar a cabo la SLR. Es necesario predefinir el mismo para así reducir la posibilidad de sesgo de la investigación. A continuación se mencionan los principales componentes del protocolo (que incluyen todos los elementos de la investigación, más información de planificación adicional):

- Antecedentes: justificación de la necesidad de realizar la investigación.
- Preguntas de investigación que se pretende responder.
- Estrategia para realizar la búsqueda de estudios primarios incluyendo la cadena de búsqueda y los recursos en dónde se realizará la misma (librerías digitales, conferencias, etc.).
- Criterios de selección para determinar cuáles estudios serán incluidos o excluidos de la SLR. Es recomendable realizar una experiencia piloto sobre un conjunto de estudios primarios para validar dichos criterios.
- Proceso de selección: descripción de cuántos investigadores evaluarán cada potencial estudio primario y como se resolverán las discrepancias y aplicación de los criterios de selección.
- Evaluación de la calidad: se deberá contar con *checklists* para evaluar la calidad de los estudios primarios. Esto es útil en caso de que haya resultados contradictorios en artículos diferentes, ya que la diferencia en calidad entre ambos podría explicar la diferencia de resultados.
- Estrategia de extracción de datos: se define el proceso de obtención de la información requerida de cada estudio primario. Si los datos requieren manipulación o asunciones e inferencias, el protocolo debe especificar un proceso de validación apropiado. Además se debe definir qué' datos serán extraídos.

- Síntesis de los datos extraídos: debe clarificar si se realizará meta-análisis y, en caso de realizarse, cuáles técnicas se utilizarán. Se debe definir cómo se trabajará con los datos extraídos para responder las PI.
- Estrategia de diseminación: se define los mecanismos de diseminación a emplear.
- Cronograma del proyecto: se elabora el calendario de trabajo que se seguirá durante la investigación.

#### 4.2.4 *Evaluación del protocolo*

Dado que el protocolo es un elemento crítico de la SLR, el mismo debe ser evaluado. Como consideraciones generales, se puede tener en cuenta los siguientes puntos para realizar la evaluación:

- ¿La cadena de búsqueda se deriva de las PI?
- ¿Los datos a ser extraídos aportan información para que las PI sean respondidas?
- ¿El procedimiento de análisis de datos es apropiado para responder las PI?

### 4.3 EJECUCIÓN DE LA INVESTIGACIÓN

#### 4.3.1 *Identificación de la investigación*

El enfoque general para generar una estrategia de búsqueda es definir cada faceta de la investigación (población, intervención, comparación, etc.) y, a partir de esto, construir una lista de sinónimos, abreviaciones y delecteos alternativos. Finalmente se construye la cadena de búsqueda utilizando los booleanos AND y OR. Fuentes usuales para realizar la búsqueda:

- Librerías digitales
- Lista de referencia de estudios primarios relevantes
- Diarios (por ejemplo, IBM Journal or Research and Development), literatura gris (como reportes técnicos y trabajos en proceso) y conferencias
- Internet

También puede considerarse contactar a los investigadores de estudios primarios relevantes por consejos sobre fuentes alternativas de material apropiado.

### Sesgo de las publicaciones

La posibilidad de que la investigación, una vez finalizada, presente cierto sesgo es un problema con el que cuenta toda SLR y se debe al hecho de que resultados positivos tienden a ser más publicados que los negativos. Otro problema se presenta cuando determinadas técnicas o herramientas son patrocinadas por grupos influyentes en la industria, ya que pocas compañías se atreverán a publicar resultados negativos al respecto y habrá un gran incentivo hacia la publicación de artículos que apoyen dicha técnica o herramienta. Ya que el sesgo en las publicaciones puede llevar a un sesgo en la investigación, se identificaron las siguientes estrategias para atacar dicho problema:

- Escaneo de literatura gris
- Escaneo de conferencias
- Contactar expertos e investigadores trabajando en el área por resultados sin publicar

### Documentación de la investigación

El proceso de la investigación debe ser transparente y replicable por lo que debe documentarse al máximo posible y con un nivel de detalle tal que los lectores sean capaces de evaluar la minuciosidad de la misma. En caso de que ocurra algún cambio, el mismo debe ser documentado y justificado. Además, los resultados obtenidos antes de aplicar el criterio de selección deben ser registrados para posible re-análisis.

En el cuadro 2 se presenta la información a ser documentada para cada fuente de información.

Además, los investigadores deben justificar las decisiones tomadas sobre librerías digitales, revistas y conferencias sobre las que se realizará la búsqueda.

[Brereton et al., 2007] identifica siete fuentes electrónicas relevantes para la IS:

- IEEE Xplore [IEE]
- ACM Digital library [acm]
- Google scholar [sch]

Fuente de información	Documentación
Librería digital	Nombre de la base de datos Estrategia de búsqueda de la base de datos (cadena de búsqueda utilizada, filtros aplicados, etc.) Fecha de la búsqueda Años cubiertos en la búsqueda
Revistas	Nombre de la revista Años investigados
Conferencias	Título de la conferencia Nombre de la presentación Nombre de la conferencia (en caso de que fuera publicada).
Estudios no publicados	Grupos de investigación contactados (nombres y detalles del contacto) Sitios web de investigación utilizados (Fecha y URL)
Otras fuentes	Fecha de búsqueda/contacto URL Cualquier información pertinente a la búsqueda

Cuadro 1: Información a ser documentada para cada fuente de información.

- CiteSeerX library [cit]
- Inspec [ins]
- ScienceDirect [sci]
- El Compendex [eng]

En la guía realizada por Kitchenham, a su vez, se recomienda considerar SpringerLink [spr], para acceder a revistas tales como Empirical Software Engineering o Springer Conference Proceedings, y Scopus [sco].

#### 4.3.2 Selección de estudios primarios

Una vez que se obtienen los estudios primarios resultantes de la búsqueda se deben evaluar según su relevancia.

Para evitar la probabilidad de sesgo, el criterio de selección debe decidirse en la etapa de definición de protocolo, aunque puede refinarse durante el proceso de búsqueda.

Los criterios de inclusión/exclusión deben basarse en las PI y se recomienda evaluarlos en una experiencia piloto con un subconjunto de los estudios primarios resultantes para asegurar que no existen interpretaciones ambiguas (por parte de los investigadores) y que, de hecho, clasifica los estudios correctamente.

A menos que un estudio se pueda excluir leyendo únicamente el título y el resumen del mismo, se debería obtener una copia completa del artículo. [Brereton et al., 2007] destacan que normalmente dichas secciones son muy pobres como para seleccionar (o no) un estudio, por lo que es recomendable leer, además, las conclusiones del mismo.

Por otra parte, en su guía, Kitchenham recomienda que, una vez excluidos todos los estudios primarios irrelevantes (es decir, aquellos que ni siquiera están relacionados a la ingeniería de software como área o con la temática de la SLR en general), se debe mantener una lista de los estudios excluidos (resultado de la aplicación de los criterios de inclusión/exclusión) junto con su justificación.

Los desacuerdos entre los investigadores sobre si un estudio debería ser incluido o no deben ser discutidos y resueltos. Aquellos estudios que sean poco claros se tratarán como casos sensibles y deberán ser documentados adecuadamente, de manera de registrar este hecho.

#### 4.3.3 Estudio de calidad de los estudios primarios

Es crítico evaluar la calidad de los estudios primarios para:

- Investigar si las diferencias en calidad proveen una explicación para diferencias en resultados de estudios diferentes.
- Como forma de otorgar cierto “peso” a la importancia de cada estudio individual cuando se realiza la síntesis de datos.
- Interpretar las conclusiones y la fuerza de las inferencias.

Sin entrar en detalle, es importante aclarar que por calidad de un estudio primario nos referimos al concepto sugerido por la guía de la CRD [Khan et al., 2001], en la que relaciona la calidad con el grado con el cual un estudio minimiza su sesgo y maximiza la validez del estudio y su aplicación fuera del mismo.

#### 4.3.4 *Extracción de datos*

El objetivo de esta etapa es diseñar formularios de extracción de datos para registrar adecuadamente la información que los investigadores obtendrán a partir de los estudios primarios. Nuevamente, para reducir la posibilidad de sesgo, se recomienda que los formularios sean definidos durante la definición del protocolo y, además, se realice una experiencia piloto con un subconjunto de estudios.

Estos formularios deben ser diseñados para recolectar toda la información necesaria para responder las PI y realizar el estudio de calidad (si lo hubiera). Si el estudio de calidad se utiliza como criterio de inclusión/exclusión, entonces deben realizarse dos formularios separados (ya que la información recolectada por el formulario de calidad debe utilizarse antes de extraer los datos de los estudios finales de la SLR). Por el contrario, si los criterios de calidad fuesen utilizados como parte del análisis de los datos, basta con realizar un único formulario. Es recomendable realizar una experiencia piloto en una muestra de los estudios primarios de manera de evaluar problemas tanto técnicos como de completitud y usabilidad.

Cada formulario debe proveer cierta información estándar, incluyendo:

- Nombre del investigador
- Fecha de la extracción de datos
- Título, autor, fuente y detalles de la publicación
- Notas adicionales

Siempre que sea posible, la extracción de datos debe ser realizada por dos investigadores (o más) independientemente. Dichos datos deben ser comparados

y las discrepancias resueltas (al igual que para el proceso de selección de estudios primarios, es recomendable definir el método de resolución de antemano y los casos poco claros considerarlos de análisis sensible). Un formulario adicional es necesario para marcar y corregir errores o discrepancias. En caso de que cada estudio primario no pueda ser evaluado por más de un investigador, se sugiere que todos los investigadores evalúen una muestra de los estudios primarios de forma de evaluar y corroborar la consistencia entre investigadores.

### **Consideraciones especiales**

Incluir publicaciones múltiples de los mismos datos en una SLR puede representar un sesgo considerable para el estudio, por lo tanto, es de suma importancia chequear que diferentes reportes no referan a un mismo estudio. En caso de la existencia de publicaciones duplicadas, debe usarse la más completa.

Si se incluye información de estudios en progreso, se debe proveer la información de calidad correspondiente que asegure la misma.

En caso de manipular datos (para obtener otros a partir de los que sí están presentados en el estudio), primero deben de ser presentados de la forma en que fueron publicados y, los datos obtenidos a partir de manipulación, deben de tratarse de forma sensible.

#### *4.3.5 Síntesis de datos*

Las actividades referentes a la síntesis de datos deben ser especificadas en el protocolo. La información extraída debe ser tabulada de manera consistente con las PI. Las tablas deben estar estructuradas de forma de resaltar similitudes y diferencias entre los resultados de los estudios.

## **4.4 REPORTAR LA INVESTIGACIÓN**

### *4.4.1 Especificación de los mecanismos de diseminación de datos*

Es de suma importancia comunicar los resultados de la SLR de manera efectiva. Debido a esto, la mayoría de las guías recomiendan planear la estrategia de diseminación durante la preparación del protocolo. Como no se pretende publicar esta investigación en ningún medio, se obviará esta etapa.

#### 4.4.2 *Formato del reporte final*

El reporte de la investigación normalmente cuenta con las siguientes secciones: título, autores, resumen, antecedentes, preguntas de investigación, métodos de la investigación, estudios incluidos y excluidos, resultados, discusión, conclusiones, agradecimientos, conflicto de intereses (si hubiera), referencias y apéndices.

## PROTOCOLO DE TRABAJO

Este capítulo presenta el protocolo de trabajo elaborado a partir de las pautas definidas en la guía de trabajo [capítulo 4]. Además, fue validado mediante la ejecución de dos experiencias piloto denominadas piloto 1 y piloto 2.

En la Figura 4 se presenta el proceso realizado para obtener la versión final del protocolo de trabajo.

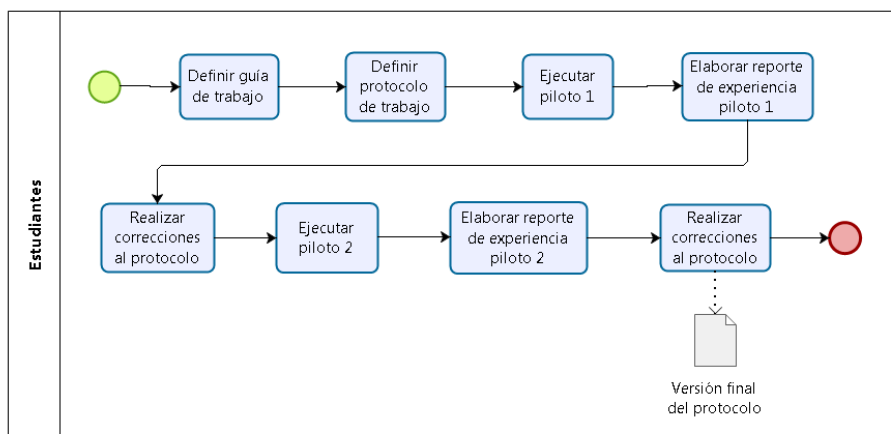


Figura 4: Proceso para obtener el protocolo de la SLR.

### 5.1 IDENTIFICAR LA NECESIDAD DE LA REVISIÓN

Existe un interés particular por investigar y recopilar la información existente respecto al *Personal Software Process* y sus efectos en el desarrollo de software. Debido a esto, se realizó una búsqueda para constatar si existían, de hecho, estudios sistemáticos sobre esta temática. Para esto, se utilizaron las siguientes librerías digitales: Scopus, ScienceDirect, IEEE Xplore, SpringerLink, CiteSeerX y ACM.

Para el caso de Scopus, ScienceDirect, IEEE Xplore, CiteSeerX y ACM se aplica la siguiente cadena de búsqueda "*Personal Software Process AND Systematic*

*literature review*” sobre el título, resumen y palabras clave de los estudios. Para el caso particular de SpringerLink se realizó la búsqueda sobre texto completo (utilizando la misma cadena de búsqueda que en los demás buscadores), ya que es la única opción que ofrece la librería. En el anexo C se detallan las búsquedas realizadas.

Ninguna de las librerías arrojó resultados luego de la búsqueda, a excepción de SpringerLink. Éste último arrojó 3 resultados, ninguno de los cuales se relacionaba con la temática establecida. Por lo tanto, se considera necesario la necesidad de realizar una Revisión Sistemática de la Literatura sobre los efectos del Personal Software Process en el desarrollo de software.

Las librerías digitales anteriormente mencionadas se escogieron siguiendo las pautas de la guía de trabajo desarrollada, por lo tanto se considera justificación suficiente para determinar que no se tiene conocimiento de ninguna revisión sistemática sobre la temática a estudiar.

## 5.2 ESPECIFICACIÓN DE LAS PREGUNTAS DE LA SLR

### 5.2.1 Preguntas de investigación (PI)

RQ.1. ¿Qué efectos produce aplicar el PSP en el desarrollo de software?

RQ.2. ¿Qué medidas se utilizan a la hora de evaluar los efectos producidos al aplicar PSP?

RQ.3. ¿Qué estrategias empíricas se utilizan, y cómo se describen, para evaluar los efectos producidos al aplicar PSP?

RQ.4. ¿Qué métodos estadísticos se utilizan para analizar los resultados?

### 5.2.2 Definición del alcance

- **Población:** toda aquella persona sobre la cual se realizó un estudio empírico, ya sea estudiante o profesional que utilizó PSP.
- **Intervención:** utilización del Personal Software Process, en cualquiera de sus niveles o formas de definición.
- **Comparación:** utilización de niveles o formas de definición diferentes a la presentada en la intervención o la no utilización del PSP.
- **Salida:** resumen cuantitativo de los resultados obtenidos tras utilizar el proceso.

- **Contexto:** industria y academia.
- **Diseño experimental:** experiencias empíricas.

### 5.3 ESTRATEGIA DE BÚSQUEDA

#### 5.3.1 *Cadena de búsqueda*

La cadena de búsqueda se aplica sobre el título, resumen y palabras clave de los estudios indexados en los recursos de búsqueda. A continuación se presenta dicha cadena:

*“Personal Software Process”*.

La razón por la cual se opta por aplicarla sobre título, resumen y palabras clave (en lugar de sobre texto completo) se debe a que si la cadena *Personal Software Process* no se encuentra en ninguna de las secciones anteriormente mencionadas, se concluye que el estudio no trata sobre la temática de la revisión y se descarta automáticamente. Por otra parte, no se incluye la sigla PSP como alternativa (es decir, que la cadena de búsqueda resultante fuese *“Personal Software Process OR PSP”*) porque se admitirían estudios que incluyen dicha sigla pero que hacen referencia a algo diferente que al *Personal Software Process*. A su vez, si se exige como condición necesaria que tanto la frase como la sigla estuviesen presentes (en cuyo caso la cadena de búsqueda sería *“Personal Software Process AND PSP”*), se perderían estudios que nunca utilizan la abreviación, pero sí la expresión *Personal Software Process*.

#### 5.3.2 *Recursos de búsqueda*

La librería digital utilizada para realizar la SLR es Scopus. La razón por la cual se opta por ésta se debe a que dicho buscador indexa las demás librerías mencionadas. Sin embargo, han existido casos puntuales en los que otras librerías han devuelto estudios que no se encontraban presentes en Scopus.

#### 5.3.3 *Criterios de selección*

Se incluyen aquellos estudios en los cuales:

IC1. El estudio trata sobre la aplicación de PSP y sus efectos en el desarrollo de software, presenta métricas para evaluar los mismos y/o métodos estadísticos para analizar los resultados obtenidos.

Se excluyen aquellos estudios que:

- EC1. No traten sobre aplicar PSP.
- EC2. No presenten información sobre el impacto en el software al aplicar PSP, métricas utilizadas ni métodos estadísticos aplicados.
- EC3. No constan de un carácter empírico.
- EC4. Ya hayan sido seleccionados previamente.
- EC5. No abordan un estudio primario.
- EC6. No han sido publicados <sup>2</sup>.
- EC7. Los resultados solo están disponibles en formato de resumen o presentación PowerPoint.
- EC8. El idioma no es inglés.

#### 5.3.4 Proceso de selección

El proceso de selección consta de 3 etapas, como se muestra en la Figura 5, en las que, dependiendo en la que se encuentre, se descartan estudios mediante ciertas consideraciones.

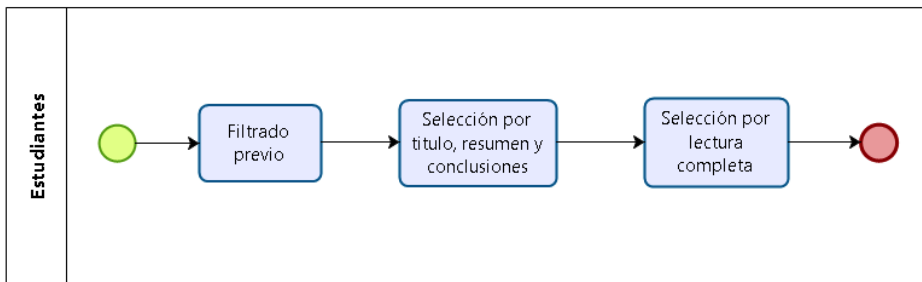


Figura 5: Etapas del proceso de selección.

<sup>2</sup> En caso de poder contactarse con el autor del estudio o conseguirlo mediante una fuente externa se incluiría el estudio (en caso que no violase ningún otro criterio).

El proceso comienza con los estudios resultantes de aplicar la cadena de búsqueda sobre los recursos mencionados previamente. La primera etapa del proceso de selección es *filtrado previo*, que consiste en la lectura y posible eliminación de aquellos estudios no relacionados con la ingeniería de software. También se eliminan actas de conferencias o congresos donde se presenten estudios (que estén relacionados con la ingeniería de software o no) debido a que dichas actas proveen un resumen de la presentación y no el estudio realizado. Los estudios a filtrar en esta etapa son asignados de forma equitativa a los tres estudiantes. Para realizar dicho filtrado, los estudiantes leen título, resumen y conclusiones de los estudios que le fueron asignados y deciden si descartarlos o no. En caso de que se presenten dudas sobre si descartar un estudio, se lo incluye. De esta manera se lo puede analizar en mayor profundidad en las siguientes etapas.

La siguiente etapa, *selección por título, resumen y conclusiones*, se puede observar en la Figura 6. Esta etapa tiene como entrada los estudios resultantes del *filtrado previo*. Cada estudiante, de manera paralela e independiente, realiza una lectura del título, resumen y conclusiones de dichos estudios, aplicando los criterios de inclusión/exclusión. Luego de su lectura se registra en la planilla de inclusión/exclusión si dicho estudio es incluido o excluido. En caso de ser excluido, se registra en la planilla su justificación. Si un estudio es incluido por al menos un estudiante, éste será considerado para la etapa final del proceso.

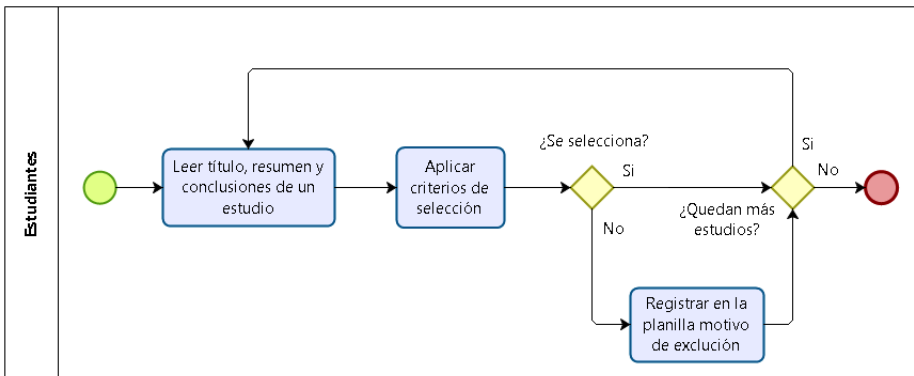


Figura 6: Detalle de la etapa de selección por título, resumen y conclusiones.

La etapa final, *selección por lectura completa*, se puede observar en la Figura 7. Esta etapa tiene como entrada los estudios incluidos en la *selección por título, resumen y conclusiones*. Cada estudiante, de manera paralela e independiente, realiza la lectura completa de dichos estudios, aplicando los criterios de inclusión-

n/exclusión. Luego de su lectura se registra en la planilla de inclusión/exclusión si dicho estudio es incluido o excluido. En caso de ser excluido, se registra en la planilla su justificación. Aquellos estudios para los que existe acuerdo por parte de los estudiantes sobre su inclusión o exclusión, son registrados en la planilla. Por otra parte, se debe llegar a un acuerdo para aquellos en los que existe discrepancia. En caso de no alcanzar una resolución por parte de los 3 estudiantes, la decisión final queda a cargo de los tutores.

Finalmente, como resultado de la tercer etapa, se obtiene una colección de estudios finales, con los cuales se procede a realizar la extracción de datos.

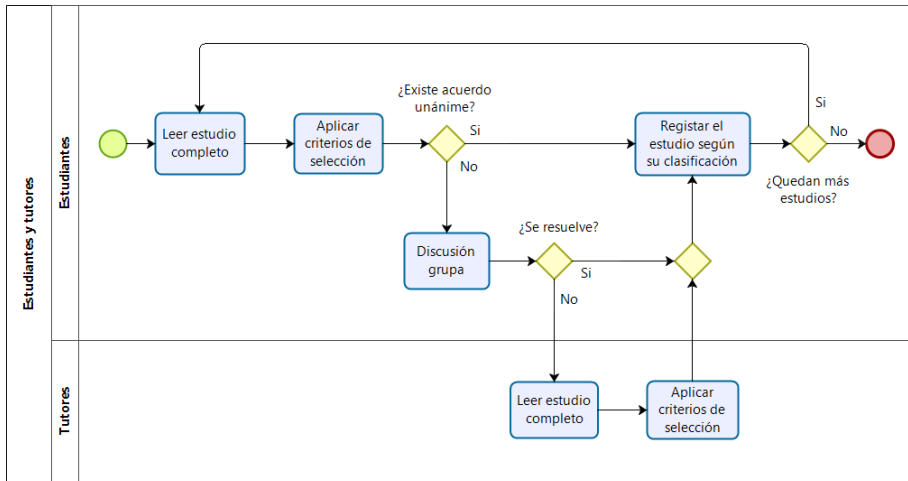


Figura 7: Detalle de la etapa de selección por lectura completa.

5.3.5 *Extracción de datos*

A continuación se lista la información extraída de los estudios obtenidos.

- Fuente de publicación.
- Año de publicación.
- Autor.
- Contexto (academia/industria).
- Estrategia empírica (experimento, caso de estudio, encuesta, etc.).

- Objetivos del estudio, donde se mencionan brevemente los objetivos del estudio.
- Descripción del estudio empírico, donde se describe el diseño del mismo.
- Descripción de los ejercicios, donde se menciona cada ejercicio a realizar en el experimento, lenguaje de programación utilizado, entre otros datos. Este campo aplica únicamente para aquellos estudios que realicen experimentos.
- Tipos de pregunta, donde se describen brevemente las preguntas realizadas en la encuesta. Este campo aplica únicamente para aquellos estudios que realicen encuestas.
- Conclusiones del estudio, donde se registran únicamente las conclusiones relevantes para la SLR y las mismas podrían ser deducidas de los resultados presentados en el estudio.
- Niveles de PSP empleados, donde se describe cada proceso utilizado durante el experimento.
- Métricas utilizadas para evaluar los efectos en el desarrollo de software.
- Métodos estadísticos utilizados para analizar los resultados.
- Tamaño de la muestra.
- Características de la muestra, donde se detalla el nivel de experiencia, universidad a la que pertenecen (en caso de que corresponda), consciencia sobre el experimento, etc.

A su vez, para cada estudio, se registrará la siguiente información:

- Nombre del responsable de la extracción de los datos
- Fecha de extracción
- Espacio para notas adicionales

#### 5.3.6 *Síntesis de datos extraídos*

Para responder la primera pregunta de investigación (**RQ.1**) se listan las conclusiones encontradas.

Para el caso de la **RQ.2** se tabulan las diferentes métricas utilizadas en los diferentes estudios y se agruparán por categorías (por ejemplo: esfuerzo, productividad, entre otras).

En cuanto a la **RQ.3**, se describen las diferentes estrategias utilizadas para evaluar el PSP.

Finalmente, para la **RQ.4** se listan los métodos estadísticos utilizados, señalando la estrategia empírica utilizada y la métrica evaluada. En el cuadro 2 se describe la tabla que se utiliza para presentar dicha información.

Estrategia empírica	
Métrica 1	Método estadístico empleado
Métrica 2	Método estadístico empleado

Cuadro 2: Ejemplo de presentación de datos sobre métodos estadísticos.

A continuación se detallan otros datos extraídos que se presentan en forma gráfica de forma complementaria, además de los anteriormente mencionados:

- Cantidad de estudios por año
- Porcentaje de estudios incluidos y excluidos durante el proceso de selección
- Porcentaje de estudios incluidos y excluidos tras realizar cada paso del proceso de selección
- Cantidad de estudios por Contexto
- Cantidad de estudios por estrategia empírica
- Porcentaje de procesos utilizados
- Cantidad de procesos utilizados por año

### 5.3.7 *Cronograma de la revisión*

1. Experiencia piloto
  - 1.1. Selección de estudios primarios sobre buscador Scopus
  - 1.2. Extracción de datos
  - 1.3. Síntesis de datos
  - 1.4. Diseminación de datos
  - 1.5. Conclusiones parciales
  - 1.6. Reporte de experiencia piloto

- 1.7. Correcciones del protocolo
2. Aplicación del protocolo final
  - 2.1. Selección de estudios primarios sobre todos los recursos de búsqueda
  - 2.2. Extracción de datos
  - 2.3. Síntesis de datos
  - 2.4. Diseminación de datos
  - 2.5. Conclusiones
  - 2.6. Reporte final
3. Formateo del reporte final
  - 3.1. Marco teórico
  - 3.2. Resumen



## EXPERIENCIAS PILOTO

---

En este capítulo se presentan las experiencias piloto realizadas para validar el protocolo de trabajo.

### 6.1 INTRODUCCIÓN

Las experiencias piloto son importantes para evaluar distintos aspectos de la aplicación del protocolo de trabajo. Permiten validar la cadena de búsqueda, comprobar si los datos a extraer aportan información para responder las preguntas de investigación y si la síntesis es adecuada para responderlas. Además, permiten a los estudiantes familiarizarse con el proceso, verificando que tanto ellos como los tutores sigan los mismos criterios de trabajo.

Se determinó que durante la ejecución de los pilotos, el proceso de selección fuera ejecutado hasta que se cumpla cierta condición de parada, especificada en la sección correspondiente de cada piloto. La decisión de definir una condición de parada se debe a que se desea trabajar simplemente con una muestra de los estudios y no con la totalidad de ellos. De esta forma, se puede ejecutar el protocolo de trabajo de forma rápida y realizar las mejoras necesarias al mismo.

En el proceso de selección de cada piloto, las etapas *filtrado previo*, *selección por título*, *resumen y conclusiones* y selección por *lectura completa* fueron realizadas de forma consecutiva para cada estudio. Una vez que se daba por finalizada una de las etapas (anteriormente mencionadas), se continuaba con la siguiente. Al resolver si el estudio se incluía o no, se procedía con el siguiente. Si un estudio no era eliminado en el *filtrado previo*, inmediatamente se continuaba con la etapa de *selección por título, resumen y conclusiones*. En caso de ser seleccionado en dicha etapa, inmediatamente se procedía a la etapa de selección por *lectura completa* para ese estudio. De esta forma no se leyeron estudios que luego no fueran a ser seleccionados por el hecho de tener una condición de parada.

Una vez obtenidos dichos estudios por parte de cada estudiante, se dio paso a una etapa de discusión en donde el objetivo fue llegar a una selección final de estudios. Para los casos en los que existieron discrepancias, los conflictos fueron resueltos con los tutores. Luego de cumplir con la condición de parada, se dio por culminada la etapa de selección.

## 6.2 EXPERIENCIA PILOTO 1

Este piloto sigue el proceso definido en la primera versión del protocolo (presente en el anexo D), junto con las modificaciones mencionadas en la introducción del capítulo. De ahora en adelante, se denomina a este protocolo como *protocolo 1*.

La cadena de búsqueda utilizada para el piloto 1 fue “*Personal Software Process*”. La misma fue aplicada sobre texto completo en el buscador Scopus el 11 de julio del 2016, arrojando 907 estudios. Por otro lado, la condición de parada fue alcanzar 4 estudios que cumplan con el criterio de selección.

### 6.2.1 Proceso de selección

Luego de aplicar la cadena de búsqueda, los estudios fueron ordenados por relevancia y se aplicaron los criterios de inclusión/exclusión. Los estudiantes siguieron un mismo listado de estudios de manera de respetar el orden de lectura (y así leer los mismos estudios).

En la figura 8 se presentan los resultados del proceso de selección de cada estudiante. El eje vertical representa los estudios analizados por los estudiantes. El análisis realizado por cada estudiante está representado por barras, utilizando un color distinto para cada uno de ellos. En el eje horizontal se indican las 3 etapas del proceso de selección, *filtrado previo*, *selección por título, resumen y conclusiones* y *selección por lectura completa*, junto con la resolución de inclusión de cada estudio. Si la barra finaliza en una fase, significa que el estudio fue descartado en esa fase. En cambio, si la barra se extiende hasta el valor *incluido* significa que el estudio fue incluido.

Una vez finalizado el proceso de selección individual, se procedió a realizar la discusión grupal sobre cada uno de los estudios que tuvieron discrepancias. Dado que algunos estudios seleccionados por el estudiante B fueron descartados en la etapa de discusión grupal, dicho estudiante debió retomar su proceso de selección individual de manera de poder continuar con la discusión grupal.

Durante la discusión grupal existieron dudas entre los estudiantes acerca de la inclusión o no de ciertos estudios. Por lo tanto, se debió consultar a los tutores sobre los siguientes estudios:

- **Estudio 7:** fue discutido con los tutores ya que fue incluido únicamente por el estudiante B. A pesar de que realizaba una comparación entre 2 niveles de PSP, se trataba de la experiencia personal del autor. Dado que la experiencia personal de un autor no califica como una experiencia empírica, se resolvió excluirlo.

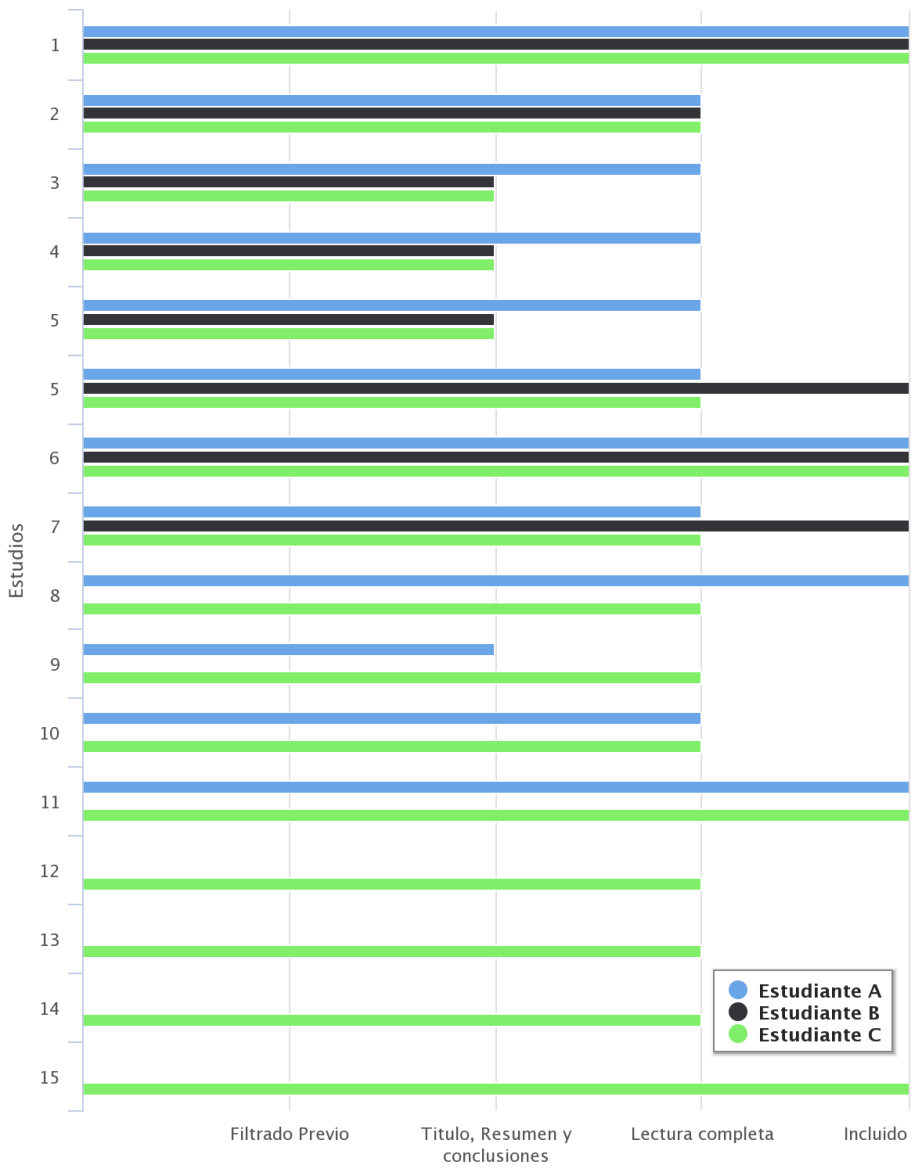


Figura 8: Proceso de selección por estudiante.

- **Estudio 9:** fue discutido con los tutores ya que fue incluido únicamente por el estudiante B ya que se mencionan y describen métricas. Sin embargo, no aporta ningún otro tipo de información. Se resolvió que los estudios de este tipo (únicamente mencionan definiciones y no constaran de un carácter empírico) deben ser excluidos ya que no presentan información sobre los efectos del PSP en el desarrollo de software.
- **Estudio 10:** fue discutido con los tutores ya que fue incluido únicamente por el estudiante C dado que, a pesar de realizar una comparación entre la aplicación del PSP contra la no aplicación del mismo, se consideró que aporta información sobre los efectos del PSP en el desarrollo de software. Se resolvió que sería necesario incluir los estudios de este tipo. Debido a esto, fue necesario modificar la Definición de alcance del protocolo 1 (ya que se establecía que se compararía únicamente contra la utilización de niveles o formas de definición de PSP).
- **Estudio 11:** fue discutido con los tutores aunque fue excluido por los 3 estudiantes ya que trataba sobre la utilización de una herramienta para soportar la aplicación de PSP. Se resolvió que la utilización de una herramienta es considerada una variación del PSP por lo tanto el estudio 11 fue incluido.
- **Estudio 13:** fue discutido con los tutores ya que fue excluido por el estudiante B debido a que no comparaba dos aplicaciones del PSP. Al igual que el estudio 10, se resolvió su inclusión por los mismos motivos.

Como resultado de la discusión grupal junto con el apoyo de los tutores, se incluyeron los estudios detallados en el cuadro 3.

A continuación se describe brevemente cada uno de los estudios seleccionados:

- **Estudio 2:** el estudio presenta una variante al PSP llamado Hybrid Personal Software Process y la compara contra PSP. Dicha variante es una combinación entre PSP y pair programming.
- **Estudio 8:** el estudio presenta un experimento para estudiar los factores críticos que afectan la calidad del software y la productividad utilizando PSP.
- **Estudio 10:** investiga la implementación de PSP en la construcción de un software siguiendo el modelo MVC. El objetivo es crear una guía para aplicar la metodología de PSP que mejore el rendimiento personal de los ingenieros o estudiantes que utilicen el modelo MVC.

ID	Título	Autores
2	Software defect detection and process improvement using personal software process data	Gopichand M., Swetha V., Ananda Rao A.
8	Critical factors affecting personal software processes	Zhong X., Madhavji N.H., El Emam K.
10	Implementing personal software process in undergraduate course to improve model-view-controller software construction	Nachiengmai W., Ramingwong S.
11	The development of a design tool for personal software process (PSP)	Chaiyo Y., Ramingwong S.

Cuadro 3: Estudios incluidos luego del proceso de selección.

- **Estudio 11:** se estudia la utilización de una herramienta con el fin de simplificar y minimizar la etapa de diseño del PSP.

### 6.2.2 Extracción de datos

La extracción de datos se realizó utilizando planillas de soporte al proceso. Se creó una planilla para realizar la extracción de datos que se describe en detalle en la sección extracción de datos del protocolo 1 [Fava et al., 2017].

### 6.2.3 Síntesis de datos extraídos

Para responder la primera pregunta de investigación (**RQ.1**) se extraen las conclusiones de cada estudio. Las conclusiones se presentan por estudio (utilizando el identificador presentado en el cuadro 3 para referenciarlos).

#### ESTUDIO 2

- La utilización del *Hybrid Personal Software Process* (HPSP) fue más eficiente y dió resultados más predecibles en términos de tiempo de desarrollo y LOC que el PSP.
- La cantidad de retrabajo fue menor en el caso de HPSP en comparación con el PSP.

ESTUDIO 10

- La utilización de PSP adaptado para MVC mejoró el rendimiento en general (en términos de exactitud de la estimación en tiempo y tamaño, densidad de defectos y productividad medida en LOC/hora) en comparación a no utilizarlo.

ESTUDIO 11

- La incorporación de una herramienta para dar soporte en la etapa de diseño del PSP no influyó sensiblemente en el rendimiento de los estudiantes.

Para responder la (RQ.2) se extraen las métricas presentadas en cada experimento y se presentan en el cuadro 4. En la columna *ocurrencias* se detalla la cantidad de estudios que hacen uso de cada métrica. En dicho cuadro las métricas se clasificaron siguiendo la clasificación utilizada en [Humphrey, 2000], creándose la categoría correspondiente para las métricas que no entran en ninguna de las categorías presentadas para PSP. De forma complementaria, en la figura 9 se muestra una representación gráfica de dicho cuadro.

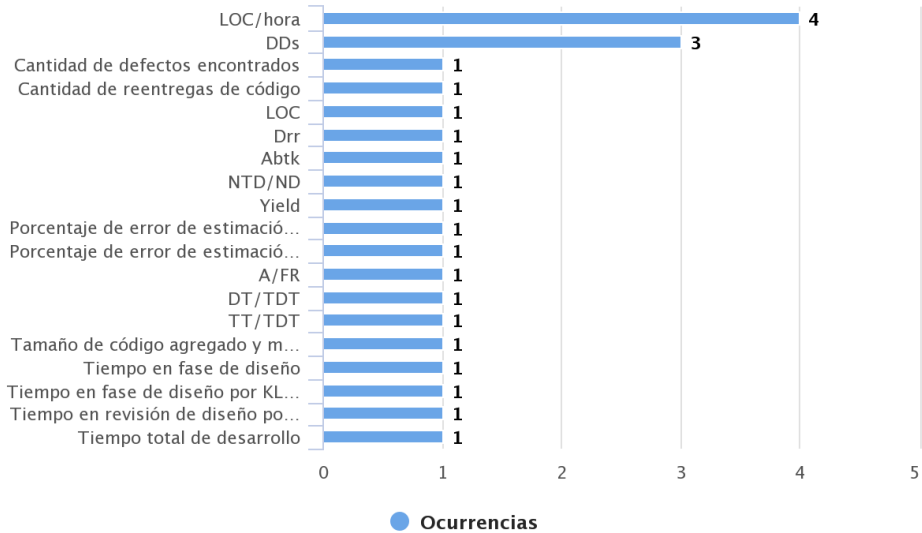


Figura 9: Métricas presentadas en cada estudio y sus ocurrencias.

<b>Métrica</b>	<b>Ocurrencias</b>
<b>Calidad</b>	
Cantidad de defectos encontrados	1
Cantidad de reentregas de código	1
Defectos removidos cada KLOC (DDS) <sup>1</sup>	3
Número de defectos removidos por hora (DRR)	1
Número promedio de fases revisadas para resolver un defecto (Abtk)	1
Porcentaje de defectos detectados durante la fase de test sobre el número total de defectos (NTD/ND)	1
Porcentaje de defectos removidos antes de la primera compilación (Yield)	1
Porcentaje de tiempo de revisión de código y diseño sobre tiempo de compilación y test (A/FR)	1
Porcentaje de tiempo empleado en la fase de diseño sobre el tiempo total de desarrollo (DT/TDT)	1
Porcentaje de tiempo empleado en la fase de test sobre el tiempo total de desarrollo (TT/TDT)	1
<b>Tamaño</b>	
LOC	1
Porcentaje de error de estimación de tamaño	1
Tamaño de código agregado y modificado	1
<b>Tiempo</b>	
Porcentaje de error de estimación de tiempo	1
Tiempo en fase de diseño	1
Tiempo en fase de diseño por KLOC	1
Tiempo en revisión de diseño por KLOC	1
Tiempo total de desarrollo	1
<b>Productividad</b>	
LOC/hora	4

<sup>1</sup> En algunos estudios se refiere a esta métrica como defectos removidos cada KLOC y en otros defectos por KLOC. Se decidió agruparlas bajo el mismo nombre ya que son equivalentes.

Cuadro 4: Métricas presentadas en cada estudio y sus ocurrencias.

Para responder la **RQ.3** se extraen las estrategias empíricas empleadas en cada estudio y en la figura 10 se presenta la gráfica asociada.

#### EXPERIMENTO

A continuación se describe en detalle la ejecución del experimento para cada estudio de este tipo:

- **Estudio 2:** se crearon 4 grupos, dos grupos de un estudiante y dos grupos de dos estudiantes. Los grupos de un estudiante utilizaron PSP y los de dos estudiantes utilizaron HPSP. Los programadores fueron asignados a sus respectivos grupos de modo que el promedio de GPA (Grade Point Average) de cada equipo sea prácticamente el mismo. Los equipos que utilizaron PSP escribieron su programa en una hoja de papel o en un editor de texto antes de pasarlo a un compilador. Todos los ejercicios fueron realizados en la universidad a modo de poder confiar en los tiempos registrados. Los lenguajes utilizados fueron elegidos por los propios programadores (algunos utilizaron C, otros C++ y otros Eiffel). Se decidió usar un conjunto de 4 ejercicios propuestos por Humphrey en [Humphrey, 1995]. Estos fueron:

##### EJERCICIO 1

Escribir un programa que calcule la mediana y la desviación estándar de un conjunto de  $n$  números reales<sup>3</sup>.

##### EJERCICIO 2

Escribir un programa que calcule los parámetros de la regresión lineal.

##### EJERCICIO 3

Escribir un programa que cuente las LOC lógicas de un programa, omitiendo comentarios y líneas en blanco.

##### EJERCICIO 4

Escribir un programa que cuente las LOC totales de un programa, las LOC en cada objeto y el número de métodos en cada objeto.

- **Estudio 8:** el experimento se realizó sobre una muestra de 53 estudiantes. Se le pidió a cada estudiante que entregue 8 proyectos utilizando PSP, uno por semana. El lenguaje utilizado fue C++. Se utilizaron planillas de producto y de proceso para recolectar la información de desarrollo. No especifica la naturaleza de los proyectos.

---

<sup>3</sup> El estudio utiliza el término *estime* en lugar de *calcule*. Sin embargo, se considera que el término adecuado es *calcular* ya que se trata de un cálculo matemático.

- **Estudio 10:** se crearon 3 grupos de 6 personas basándose en su experiencia previa (dada por los cursos a los cuales atendieron). El experimento duró 9 días de workshops, cada día dividido en 3 etapas: discusión de *post-mortem*, presentación del ejercicio a realizar e implementación del ejercicio. Cada grupo fue responsable del Modelo, Vista y Controlador respectivamente. Para cada ejercicio se hizo una selección randómica de cada grupo para formar un nuevo equipo. A nivel de herramientas se utilizó el framework PHP CodeIgniter. Se rediseñaron 8 ejercicios del curso original de PSP para soportar el patrón MVC:

EJERCICIO 0

Escribir un programa que cuente las LOC de un programa<sup>4</sup>.

EJERCICIO 1

Escribir un programa que calcule la mediana y la desviación estándar.

EJERCICIO 2

Escribir un programa que cuente las líneas y partes<sup>5</sup> de un programa.

EJERCICIO 3

Escribir un programa que liste datos de empleados.

EJERCICIO 4

Escribir un programa que reporte las calificaciones de los estudiantes online.

EJERCICIO 5

Escribir un programa que implemente un mini horóscopo.

EJERCICIO 6

Escribir un programa que implemente un mini diccionario thai-  
inglés.

EJERCICIO 7

Escribir un programa que implemente un mini memo<sup>6</sup>.

- **Estudio 11:** participaron del estudio 6 estudiantes. Cada uno de ellos realizó 7 ejercicios utilizando PSP estándar. En los 3 primeros ejercicios no se requirió proceso de diseño, en los ejercicios 4 y 5 se siguió un proceso de diseño, utilizando las plantillas estándar de PSP. En los dos últimos se utilizó la herramienta presentada en el estudio. La

---

<sup>4</sup> El estudio no especifica si son LOC totales o lógicas.

<sup>5</sup> El estudio no brinda más detalle.

<sup>6</sup> El estudio no brinda más detalle.

mayoría de los estudiantes desarrollaron los ejercicios en C++ y PHP. Los ejercicios realizados fueron los estándar de PSP, sin especificar (dichos ejercicios se pueden observar en el anexo B).

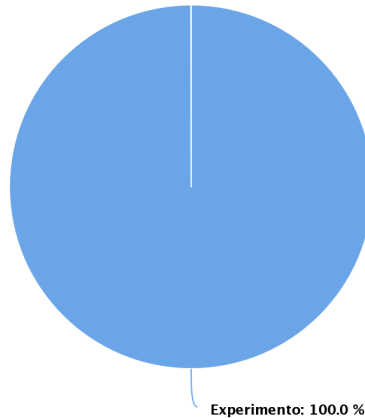


Figura 10: Proporción de estrategias empíricas utilizadas en los estudios incluidos.

Para responder la **RQ.4** se extraen los métodos estadísticos utilizados en cada estudio y son mencionados brevemente. Solamente los estudios 8 y 11 incluyen métodos estadísticos. En la figura 11 se observa la proporción entre los estudios incluidos que presentan dichos métodos y aquellos que no.

La prueba F de Fisher (F-test) se aplica, en el contexto del estudio en la que se presenta, para confirmar que los valores de las métricas siguen un comportamiento cuadrático. Dado que presentan dicho comportamiento, se realiza un test de regresión cuadrático para estudiar la relación entre cierto conjunto de métricas detalladas en la Tabla 4. A partir de esto se puede determinar si existe una relación entre las métricas o no.

En el contexto del estudio que aplica ANOVA, se comparan las medias de las métricas obtenidas (específicamente *Tiempo en fase de diseño/KLOC*, *Defectos/KLOC* y *Tiempo en revisión de diseño/KLOC*) para los 2 tipos de ejecuciones realizadas (con la herramienta de soporte para la fase de diseño y sin ella).

De manera complementaria, en el cuadro 5 se detalla, para cada estrategia empírica utilizada en los estudios incluidos, el método estadístico empleado y la métrica sobre la cual se aplicó. En el cuadro 6 se listan los métodos estadísticos utilizados y el identificador de estudio en el cual se aplica (correspondiéndose dicho identificador con el utilizado en la planilla de extracción [Fava et al., 2017]).

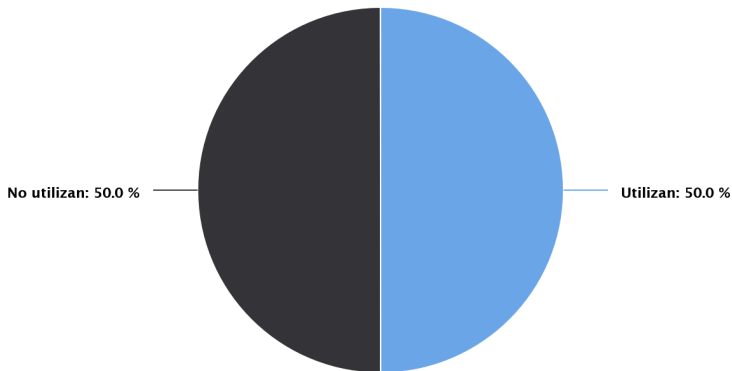


Figura 11: Proporción de estudios que utilizan métodos estadísticos y aquellos que no.

Finalmente, en la figura 12 se presenta gráficamente la cantidad de ocurrencias de cada método estadístico.

#### 6.2.4 *Análisis cuantitativo*

En esta sección se presenta un resumen cuantitativo del proceso de selección y de la extracción de datos del piloto 1. Para realizar este resumen se tomó en cuenta los estudios con identificadores comprendidos entre 2 y 11 inclusive, ya que fueron los 10 estudios discutidos durante el proceso de selección. En la figura 13 se representa el porcentaje de estudios incluidos y excluidos durante el proceso de selección. Además, se incluye una gráfica de la cantidad de estudios totales y seleccionados por año (Figura 14).

La figura 15 presenta el porcentaje de los medios de publicación de los estudios. De forma complementaria, en el cuadro 7 se detallan los lugares de publicación de los estudios.

A partir de la figura 16 se puede observar el porcentaje de los estudios incluidos diseminados por contexto, mientras que en el cuadro 8 se pueden observar las instituciones participantes. Es importante tener en cuenta que algunos estudios fueron realizados por varias instituciones en conjunto.

La figura 17 presenta los estudios agrupados por proceso utilizado por año. Para esto se utilizaron 4 grupos: PSP con herramienta, PSP (sin especificar nivel), adaptaciones del PSP y PSP en alguno de sus niveles. Es importante destacar

<b>Estrategia empírica: Experimento</b>			
<b>Relaciones estudiadas</b>		<b>Método estadístico</b>	<b>ID</b>
DDS	A/FR, Yield, Abtk, DT/TDT, NTD/ND, TT/TDT	Regresión cuadrática	8
DRR	A/FR, Yield, Abtk, DT/TDT, NTD/ND, TT/TDT	Regresión cuadrática	8
LOC/hora	A/FR, Yield, Abtk, DT/TDT, NTD/ND, TT/TDT	Regresión cuadrática	8
Abtk	A/FR, Yield	Regresión cuadrática	8
NTD/ND	A/FR, Yield	Regresión cuadrática	8
TT/TDT	A/FR, Yield	Regresión cuadrática	8
<b>Media de métricas comparadas</b>		<b>Método estadístico</b>	<b>ID</b>
Tiempo en fase de diseño por KLOC con la herramienta de diseño	Tiempo en fase de diseño por KLOC sin la herramienta de diseño	ANOVA	11
Defectos cada KLOC con la herramienta de diseño	Defectos cada KLOC sin la herramienta de diseño	ANOVA	11
Tiempo en revisión de diseño por KLOC con la herramienta de diseño	Tiempo en revisión de diseño por KLOC sin la herramienta de diseño	ANOVA	11

Cuadro 5: Métodos estadístico empleados por métrica.

Método estadístico	ID
Regresión cuadrática	8
F-test	8
ANOVA	11

Cuadro 6: Métodos estadísticos utilizados por estudio.

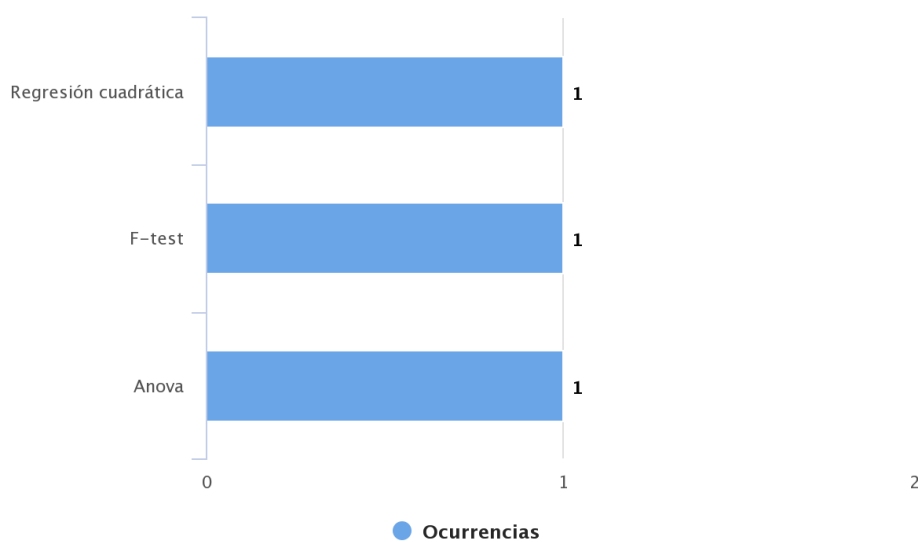


Figura 12: Cantidad de ocurrencias de los métodos estadísticos.

que en un mismo estudio se puede utilizar más de un proceso (por dos grupos distintos, por ejemplo).

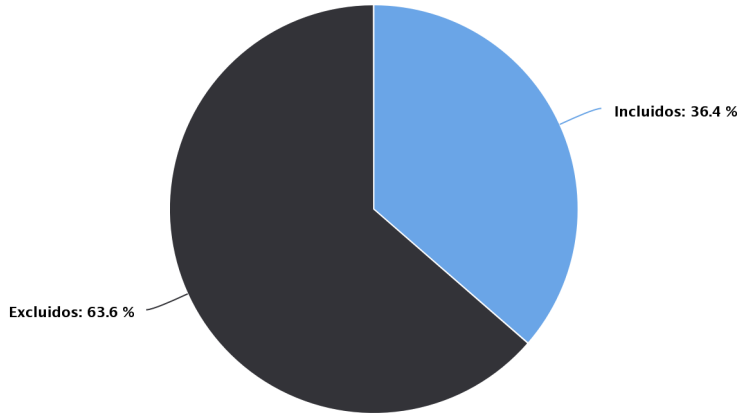


Figura 13: Proporción de estudios incluidos y excluidos.

<b>Fuente de publicación</b>	<b>Estudios</b>
2010 IEEE International Conference on Communication Control and Computing Technologies, ICCCT 2010	1
2013 10th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, ECTI-CON 2013	1
IEEE Software	1
Lecture Notes in Electrical Engineering	1

Cuadro 7: Fuente de publicación en los estudios incluidos.

#### 6.2.5 *Otros hallazgos*

Durante la extracción de datos se registró información interesante para el área de Ingeniería de software empírica pero no directamente relacionada con las preguntas de investigación. Por lo tanto, se decidió presentarla en forma de apartado.

El Estudio 8 concluye lo siguiente respecto a las métricas presentadas:

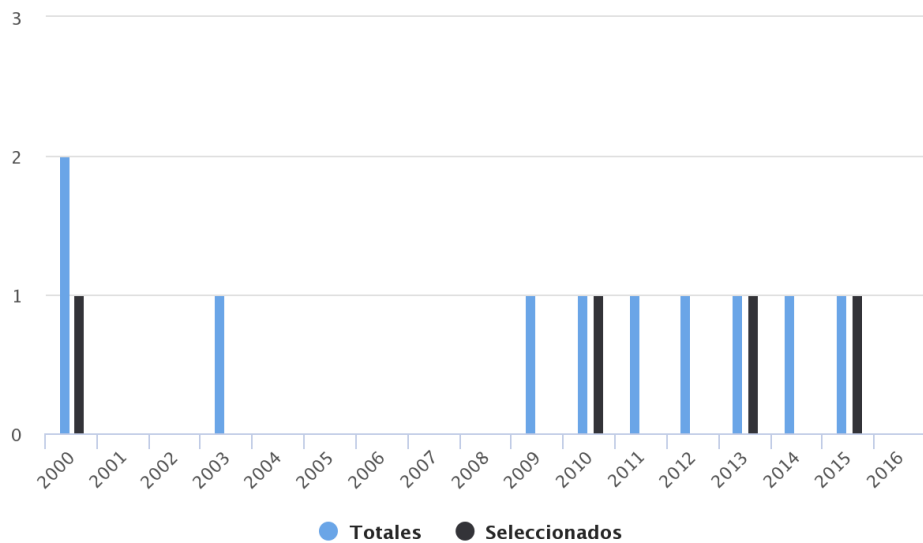


Figura 14: Cantidad de estudios incluidos y totales por año.

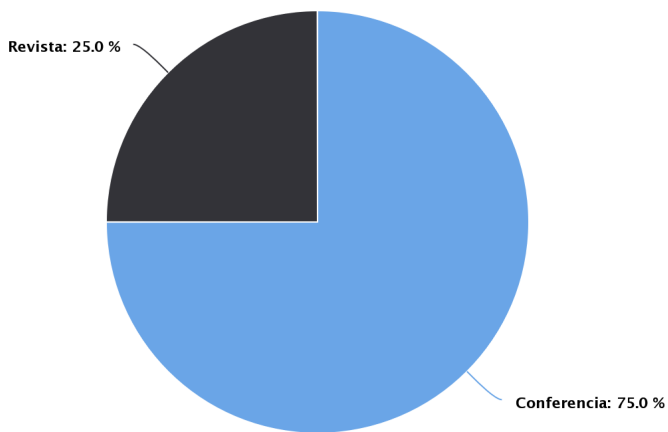


Figura 15: Proporción de medios de publicación.

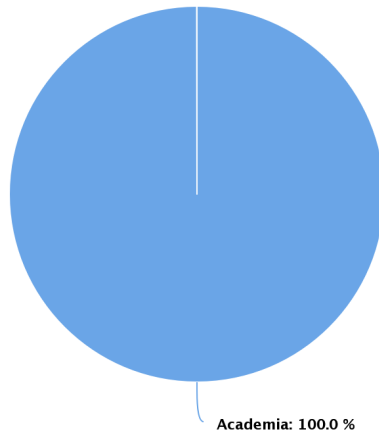


Figura 16: Proporción de contexto de publicación de los estudios incluidos.

<b>Institución</b>	<b>Estudios</b>
Department of Computer Engineering, Faculty of Engineering, Chiang Mai University, Chiang Mai, Thailand	2
Department of Computer Science, University of Quebec, Montreal, Que., Canada	2
Dept. of CSE, BVRIT, Narasapur, Medak, India	1
Dept. of CSE, JNTUA, Anantapur, India	1
JNTU Anantapur, Anantapur, India	1
National Research Council, Ottawa, Ont., Canada	1

Cuadro 8: Instituciones participantes en los estudios incluidos.

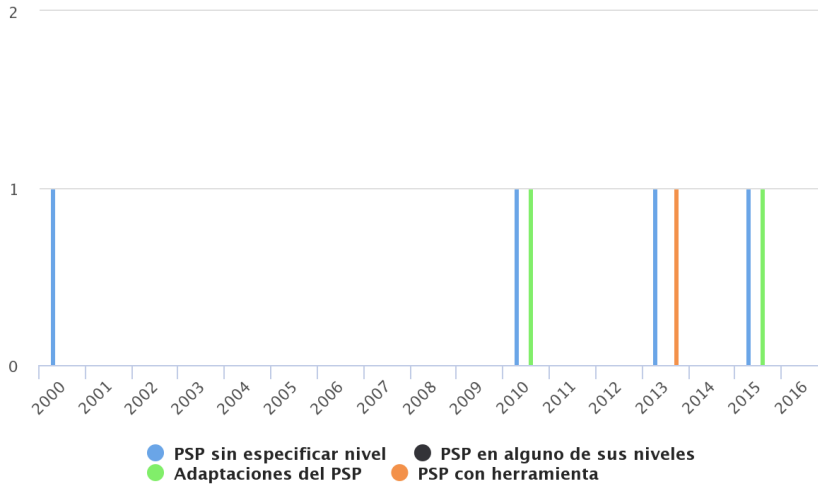


Figura 17: Cantidad de estudios agrupados por proceso utilizado por año.

- El número de defectos removidos<sup>7</sup> cada KLOC (Dds) está fuertemente relacionado con el porcentaje de defectos removidos antes de la primera compilación (Yield). Un alto valor de Yield junto con un bajo valor de Dds, podría implicar código de buena calidad (alto valor de Yield demuestra una buena detección de errores). Por otro lado, un bajo valor de Yield y de Dds podría implicar código de baja calidad debido a una mala detección de errores.
- El número de defectos removidos por hora (Drr) está fuertemente relacionado con el porcentaje de tiempo de revisión de diseño y código sobre el tiempo de compilación y test (A/FR) y Yield. Un alto valor de Drr generalmente implica que los defectos han sido relativamente sencillos de resolver (lo que implicaría que fueron detectados y resueltos en un punto cercano a su origen). Por lo tanto, un alto valor de Drr implica que al momento de culminar el producto, el mismo se encuentra relativamente libre de defectos. Entonces, un alto valor de A/FR y Yield afecta positivamente el valor de Drr.
- La cantidad de LOC por hora está fuertemente relacionada con A/FR y Yield. Un mayor tiempo en la fase de revisión y un mayor tiempo empleado en la detección de errores antes de la primera compilación implica una

<sup>7</sup> En el estudio no se especifica en qué fase fueron removidos los defectos, por lo tanto se asume que fue durante todo el proceso.

mayor cantidad de LOC por hora. Esto se debe a que es menos costoso, en términos de tiempo, detectar un error en la fase en que se inyectó.

### 6.2.6 Conclusiones

Las conclusiones presentadas en esta sección hacen referencia únicamente al análisis del proceso y no así al análisis de los datos. Esto se debe a que el objetivo del piloto 1 no fue obtener conclusiones sobre los datos extraídos. A pesar de no aportar información relevante se incluyen gráficas como la de estrategia empírica (Figura 10) y la de contexto (Figura 16). Estas fueron útiles para refinar el proceso e incluir nuevas gráficas que no fueron consideradas anteriormente por la falta de conocimiento en el tema. Por la misma razón, fue necesario agregar nuevos datos a extraer que inicialmente no fueron considerados.

Durante la etapa *extracción de datos* se hallaron datos interesantes que no forman parte de ninguna de las respuesta a las preguntas de investigación. Por este motivo, para incluir dicha información, se determinó agregar un sección denominada *otros hallazgos*.

Para responder la **RQ.4** se presentaron dificultades debido al poco conocimiento en el área de la estadística. Por esta razón, se presenta únicamente una breve descripción de la utilización de cada método estadístico presentado.

Por otra parte, la ejecución del piloto sirvió para verificar que los estudiantes sigan los mismos criterios de trabajo. Se modificó la cadena de búsqueda de manera que los estudios obtenidos tras aplicarla sean de mayor relevancia para la investigación. Además, también fueron adaptados los campos a extraer durante el proceso de extracción de datos para hacerlo de forma más eficiente. Por último, se agregaron nuevas figuras para facilitar la visualización del análisis de datos. Estas modificaciones se describen en detalle en la siguiente sección. Por otra parte, la ejecución del piloto fue útil para familiarizarse con el proceso de ejecución. Sin embargo, se consideró necesario realizar un segundo piloto con el fin de refinar la extracción de datos para los estudios en los cuales la estrategia empírica empleada fuera Encuesta, Caso de estudio e Investigación-Acción. Este también se realizó sobre estudios que realicen Experimentos de manera de confirmar el resultado obtenido tras el piloto 1.

### 6.2.7 Modificaciones del protocolo 1

En esta sección se detallan las modificaciones pertinentes en cada sección del protocolo 1.

### Definición de alcance

Durante el proceso de selección se evidenció la necesidad de también incluir aquellos estudios que comparen la utilización del PSP en cualquiera de sus formas o niveles contra la no aplicación de PSP. Esto es así ya que se detectaron varios estudios que aportarían información para responder las preguntas de investigación que realizaban dicha comparación y estaban siendo excluidos debido a la definición de alcance.

### Cadena de búsqueda

Luego de ejecutado el piloto 1, se decidió que la cadena de búsqueda se aplique únicamente sobre *título, resumen y palabras clave* y no sobre *todos los campos disponibles* del buscador. La razón de esto es que, luego de ejecutar el proceso de selección se evidenció que si la cadena “*Personal Software Process*” no se encuentra en ninguna de las secciones anteriormente mencionadas, se concluye que el estudio no trata sobre la temática de la revisión y es descartado automáticamente por parte de los estudiantes.

### Criterios de inclusión/exclusión

Durante la etapa de selección, al revisar los criterios de inclusión y exclusión en conjunto con los tutores se resolvió modificar el criterio sobre el idioma de manera de no sesgar la investigación a la región hispanoamericana. Por lo tanto, el criterio EC8 pasó de ser “*el idioma no es inglés ni español*” a “*el idioma no es inglés*”.

### Extracción de datos

Durante la etapa de extracción de datos se decidió incluir nuevos datos a extraer: *descripción del experimento o del caso de estudio, descripción de los ejercicios, características de la muestra y tipos de pregunta*. Los tres primeros estaban incluidos en el campo descripción del estudio, sin embargo, a modo de asegurar la presencia de dicha información y a su vez, disponer de ella de forma separada, se optó por dividirlo en tres campos distintos. Por otra parte, *tipos de pregunta* hace referencia a una breve descripción de las preguntas realizadas cuando la estrategia empírica es del tipo encuesta.

## Síntesis de datos extraídos

Al realizarse la síntesis de datos extraídos, se añadieron nuevas representaciones gráficas de dichos datos. Estas son: *ocurrencias de las métricas utilizadas, ocurrencias de los métodos estadísticos utilizados, porcentaje de estudios que utilizan métodos estadísticos, porcentaje de estudios por medio de publicación y procesos utilizados por año*. Se consideró que en una ejecución con más estudios podría aportar información interesante de visualizar.

### 6.3 EXPERIENCIA PILOTO 2

Este piloto sigue el proceso definido en el protocolo resultante de realizar el piloto 1. Este protocolo contiene las modificaciones detalladas en la sección 6.2, además de las modificaciones mencionadas en la introducción del capítulo. Se denomina a este protocolo como *protocolo 2*.

El objetivo de la experiencia piloto 2 es validar el protocolo 2. Particularmente, refinarlo para asegurar que contemple la mayor cantidad de estrategias empíricas posibles (encuestas, casos de estudio, investigación-acción y experimentos). Por tal motivo se modificó la cadena de búsqueda para la experiencia piloto 2 utilizando 4 cadenas de búsqueda independientes. En el cuadro 9 se listan las cadenas de búsqueda utilizadas y los resultados arrojados para cada una de ellas. Para todos los casos se aplicó “*Personal Software Process*” sobre título, resumen y palabras clave; mientras que “*Action research*”/“*Survey*”/“*Experiment*”/“*Case study*” se aplicó sobre *todos los campos disponibles*. Esta decisión fue tomada intentando no omitir resultados valiosos ya que la estrategia empírica empleada podría no ser mencionada hasta avanzado el estudio. Las cadenas de búsqueda fueron aplicadas en el buscador el 9 de octubre del 2016.

Cadena de búsqueda	Cantidad de estudios
“Personal Software Process” AND “Action research”	2
“Personal Software Process” AND “Survey”	24
“Personal Software Process” AND “Experiment”	48
“Personal Software Process” AND “Case study”	68

Cuadro 9: Cadenas de búsqueda utilizadas y su respectiva cantidad de estudios arrojados por el buscador.

La condición de parada fue alcanzar 2 estudios seleccionados para cada una de las cadenas de búsqueda, en caso de existir al menos 2.

### 6.3.1 Proceso de selección

El proceso de selección se basó en 4 búsquedas completamente independientes (detalladas en el cuadro 9), por lo tanto los resultados se exponen de forma individual.

En las figuras 18, 19, 20 y 21 se presentan los resultados del proceso de selección de cada estudiante para la estrategia empírica investigación-acción, encuesta, experimento y caso de estudio respectivamente. En la figura 18 se muestra que los dos estudios que utilizan investigación-acción fueron descartados por todos los estudiantes en la etapa de *selección por título, resumen y conclusiones*. En la figura 19 se ve que 11 de los estudios que utilizan encuestas fueron descartados por los estudiantes durante la etapa de *título, resumen y conclusiones*, mientras que los 13 restantes fueron descartados durante la etapa de *lectura completa*. Por otra parte, en la figura 20 se muestra que 2 de los estudios que utilizan experimentos fueron incluidos. 6 de estos estudios fueron descartados en la etapa de *título, resumen y conclusiones* y 4 en la etapa de *lectura completa*, leyendo únicamente 12 estudios de los 48 disponibles. Finalmente, en la figura 21 se puede ver que se leyeron 12 estudios que utilizan caso de estudio: 2 de ellos fueron descartados en la etapa de *título, resumen y conclusiones* y 8 en la etapa de *lectura completa*, dejando un total de 2 estudios incluidos.

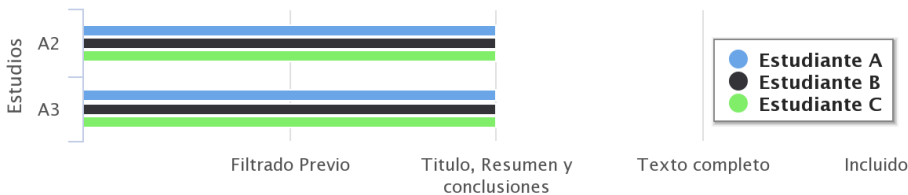


Figura 18: Proceso de selección por estudiante de estudios que utilicen investigación-acción

Una vez finalizado el proceso de selección individual, se procedió a realizar la *discusión grupal* sobre cada uno de los estudios que hubo discrepancias. A diferencia del piloto 1, casi no se presentaron diferencias entre la selección de estudios realizada por los estudiantes. Dado que un estudio seleccionado por el estudiante C fue descartado en la etapa de discusión grupal, dicho estudiante

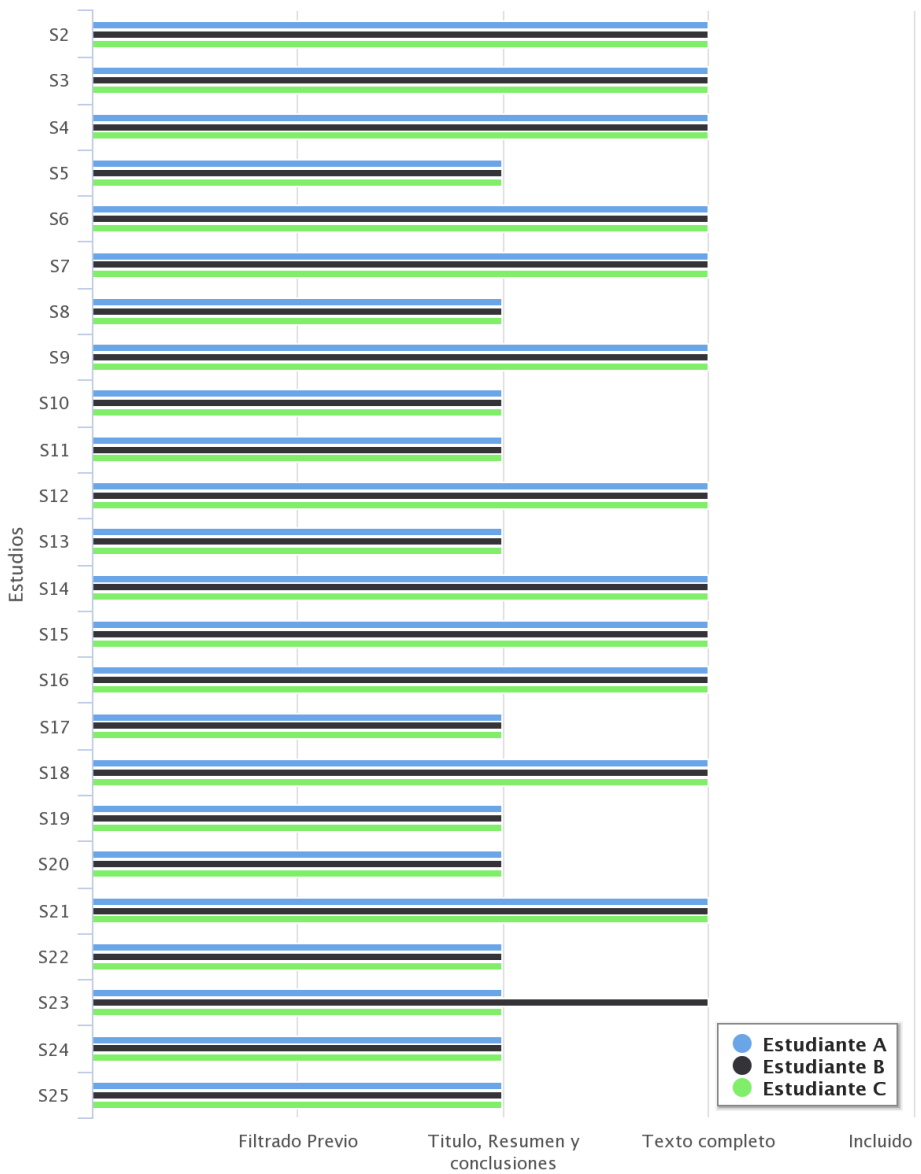


Figura 19: Proceso de selección por estudiante de estudios que utilicen encuesta

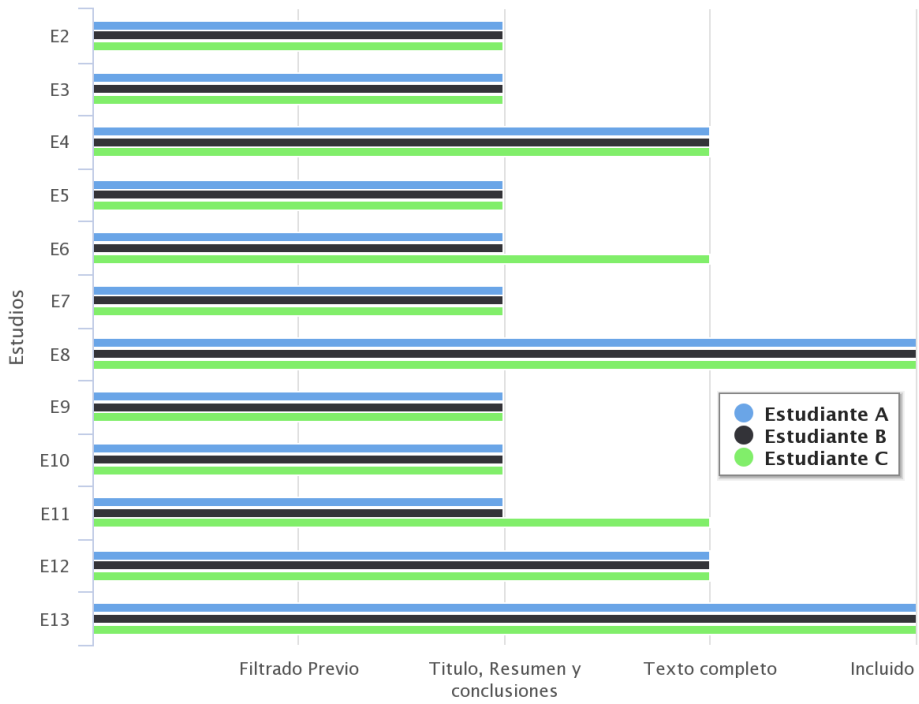


Figura 20: Proceso de selección por estudiante de estudios que utilicen experimento

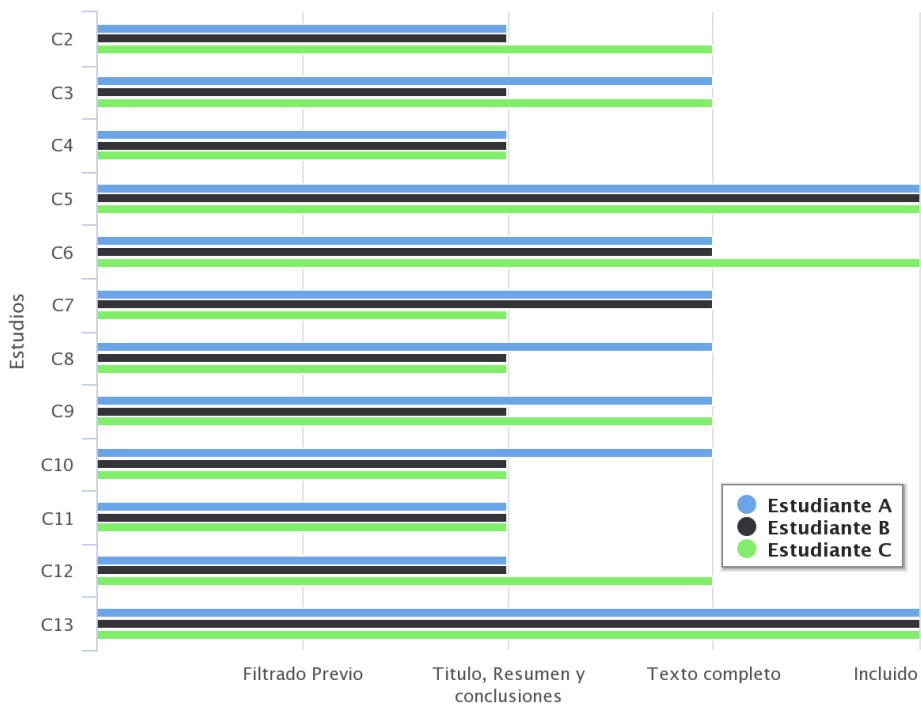


Figura 21: Proceso de selección por estudiante de estudios que utilicen caso de estudio

debió retomar su proceso de selección individual de manera de poder continuar con la discusión grupal.

Luego de finalizada la discusión grupal, se seleccionaron los estudios que se pueden observar en el cuadro 10. La letra que se encuentra en el identificador de los estudios hace referencia a la estrategia empírica de la búsqueda de la cual proviene el estudio (**A** - action research, **C** - case study, **E** - experiment, **S** - survey). Como se puede observar en las figuras 18 y 19, no se incluyeron estudios que utilicen investigación-acción ni encuestas ya que no cumplieron los criterios de inclusión/exclusión.

ID	Título	Autores
E8	Applying the Personal Software Process in CS1: An experiment	Hou Lily, Tomayko James
E13	An experiment measuring the effects of Personal Software Process (PSP) training	Prechelt L., Unger B.
C5	Results of applying the personal software process	Ferguson P., Humphrey W.S., Khajenoori S., Macke S., Matvyva A.
C13	A Case of Engineering Quality for Mobile Healthcare Applications Using Augmented Personal Software Process Improvement	Ghayyur S.A.K., Awan D., Sikander Hayat Khoyal M.

Cuadro 10: Estudios incluidos luego del proceso de selección.

A continuación se describe brevemente cada uno de los estudios seleccionados:

- **Estudio E8:** el estudio presenta un experimento para comparar el uso del PSP contra el no uso del mismo. Se crearon dos grupos, donde uno utilizó PSP y el otro no lo hizo. Realizaron una serie de ejercicios individuales y un proyecto final en grupo de dos estudiantes.
- **Estudio E13:** el estudio presenta un experimento para analizar los efectos del PSP en estudiantes con experiencia previa en PSP (grupo P), comparándolo contra otro grupo de estudiantes, sin experiencia en PSP (grupo N).
- **Estudio C5:** el estudio presenta los resultados de aplicar PSP en la industria a partir de datos recolectados en 3 empresas: Advanced Information

Services, Inc., Motorola Paging Products Group, y Union Switch & Signal Inc.

- **Estudio C13:** se presenta un PSP “aumentado”: Architectural Augmented Personal Process (AAPP) para mejorar la calidad de sistemas de salud en aplicaciones móviles. El estudio explora los efectos de adoptar metodologías de arquitectura de software en el PSP y la efectividad del AAPP. Se generaron dos sistemas y dos grupos por sistema donde un grupo aplicó AAPP y el otro PSP.

### 6.3.2 *Extracción de datos*

Se realizó de la misma forma que el piloto 1 pero utilizando las planillas detalladas en el protocolo 2 [Fava et al., 2017].

### 6.3.3 *Síntesis de datos extraídos*

Para responder la primera pregunta de investigación (**RQ.1**) se extraen las conclusiones encontradas. Las conclusiones se presentan por estudio (utilizando el identificador presentado en el cuadro 10 para referenciarlos).

#### ESTUDIO E8

- El tiempo total de desarrollo promedio de los ejercicios fue similar para los estudiantes que utilizaron PSP y los que no lo utilizaron. El tiempo total de desarrollo individual del proyecto también fue similar, pero el tiempo invertido en la etapa de integración de las partes individuales fue bastante menor en los grupos que utilizaron PSP. Esto probablemente se debe a que los estudiantes que utilizaron PSP realizaron revisión de código y testeó antes de la integración, por lo tanto, dicha etapa fue más sencilla.

#### ESTUDIO C5

- De las 3 empresas, AIS es la única compañía que expone datos comparando PSP con la no utilización de PSP. De ella se obtiene que luego de entrenar a los desarrolladores en PSP y aplicarlo, se mejoró significativamente la estimación de tiempo por componente, disminuyeron considerablemente los defectos por KLOC y aumentó levemente la productividad medida en LOC/hora.

## ESTUDIO C13

- Aunque para ambos sistemas el grupo que utilizó PSP tuvo un menor valor de esfuerzo (en fase de diseño de arquitectura y total), esta diferencia se hace menos significativa a medida que se aumenta el tamaño del sistema (en términos de LOC). Esto se traduce a que a mayor esfuerzo durante la fase de diseño de arquitectura, menor es el retrabajo requerido durante el resto del proyecto. Lo mismo sucede con el costo (durante la fase de diseño de arquitectura y total). En ambos sistemas el grupo utilizando PSP tuvo un costo menor, pero al aumentar el tamaño del sistema, la diferencia de costos se volvió menos significativa. Para el caso de tiempo de salida al mercado ocurre lo mismo.

No se presentan conclusiones para el estudio E13 ya que la mayoría de los integrantes del grupo de PSP no utilizaron el proceso.

Para responder la **RQ.2** se extraen las métricas registradas en cada estudio siguiendo la clasificación utilizada en [Humphrey, 2000]. Las mismas se pueden observar en el cuadro 11. De forma complementaria, en la figura 22 se muestra una representación gráfica de dicho cuadro.

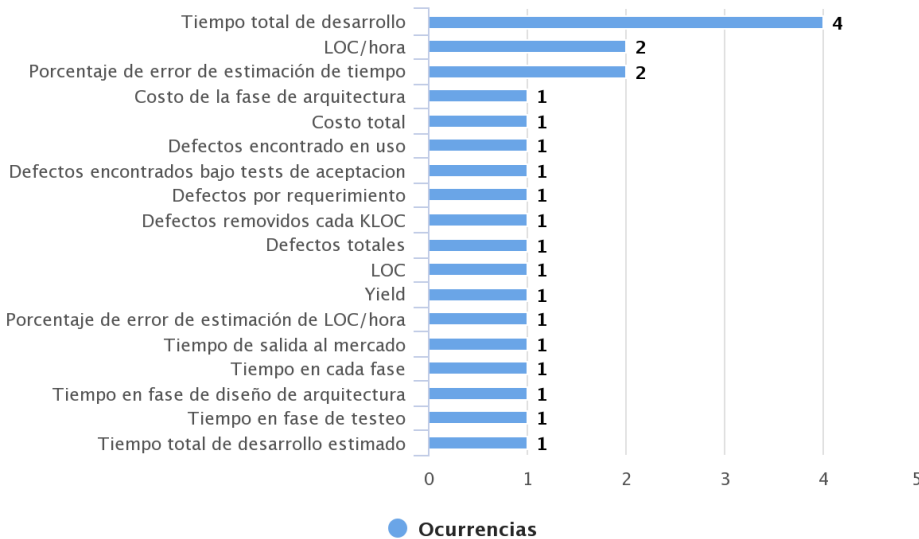


Figura 22: Métricas presentadas en cada estudio y sus ocurrencias.

<b>Métrica</b>	<b>Ocurrencias</b>
<b>Calidad</b>	
Defectos encontrado en uso	1
Defectos encontrados bajo tests de aceptacion	1
Defectos por requerimiento	1
Defectos removidos cada KLOC	1
Defectos totales	1
Porcentaje de defectos removidos antes de la primera compilación (Yield)	1
<b>Costo</b>	
Costo de la fase de arquitectura	1
Costo total	1
<b>Tamaño</b>	
LOC	1
<b>Tiempo</b>	
Tiempo de salida al mercado	1
Tiempo en cada fase	1
Tiempo en fase de diseño de arquitectura	1
Tiempo en fase de testeo	1
Tiempo total de desarrollo	4
Tiempo total de desarrollo estimado	1
Porcentaje de error de estimación de tiempo	2
<b>Productividad</b>	
LOC/hora	2
Porcentaje de error de estimación de LOC/hora	1

Cuadro 11: Métricas presentadas en cada estudio y su cantidad de ocurrencias

Para responder la **RQ.3** se extraen las estrategias empíricas empleadas en cada estudio y en la figura 23 se presenta la gráfica asociada.

#### EXPERIMENTO

A continuación se describe en detalle la ejecución del experimento para cada estudio de este tipo:

- **Estudio E8:** se crearon dos grupos de estudiantes, uno utilizó un conjunto de componentes de PSP mientras que el otro grupo no aplicó ningún proceso, simplemente registró el tiempo que le llevó cada tarea. El experimento constó de 8 ejercicios<sup>8</sup> individuales y un proyecto. El proyecto fue realizado en grupos de dos estudiantes, donde cada uno realizó su parte y luego se integraron. Se utilizó C++. El estudio fue realizado sobre 130 estudiantes de primer y segundo año de la Carnegie Mellon University. No todos tenían experiencia con PSP. Además, algunos no tenían experiencia programando.
- **Estudio E13:** se llevó a cabo durante 1997 y 1998. Se formaron 2 grupos donde uno había asistido a un curso de PSP y el otro no. Al grupo de los estudiantes de PSP no se les pidió específicamente que lo utilizaran. Ambos grupos tenían que realizar la misma tarea y trabajaron bajo las mismas condiciones. El lenguaje de programación se dejó a elección de los estudiantes. Todos debieron implementar un programa de forma individual el cual consistía en realizar un programa que codifique dígitos numéricos en su correspondiente secuencia de palabras de acuerdo a una tabla determinada. El experimento comenzó con 48 estudiantes de los cuales 8 abandonaron, dejando un total de 40 participantes. 24 de ellos realizaron un curso de PSP y 16 realizaron otro curso de carácter técnico (no metodológico). Todos eran hombres, estudiantes de posgrado de Computer Science que se encontraban en su octavo semestre de la universidad y tenían 8 años de experiencia en programación.

#### CASO DE ESTUDIO

A continuación se describe en detalle la ejecución del caso de estudio para cada estudio de este tipo:

- **Estudio C5:** las compañías AIS, Motorola y US&S utilizaron PSP y juntaron datos para mostrar sus efectos. Cada una midió los resultados de varios proyectos implementados con PSP. En todos los casos los

---

<sup>8</sup> El estudio no brinda más detalle.

proyectos eran parte de la operativa normal de la compañía y no fueron diseñados para el caso de estudio en particular. Los proyectos en Motorola y US&S involucraron mantenimiento de software y mejoras, mientras que en AIS se aplicó en un nuevo proyecto y mejoras. Los proyectos fueron programados en C y C++. Los participantes fueron ingenieros entrenados en PSP.

- **Estudio C13:** se implementó primero una pequeña aplicación para *smartp hones* en forma de piloto. Los resultados obtenidos fueron utilizados para refinar el proceso. Se estudiaron 2 sistemas independientes de distinto tamaño. En cada sistema había dos grupos: uno utilizando PSP y otro AAPP. El desarrollo fue sobre una aplicación para *smartp-hones* para el manejo de un hospital. El caso de estudio fue ejecutado en la Preston University de Pakistán con ingenieros entrenados en PSP.

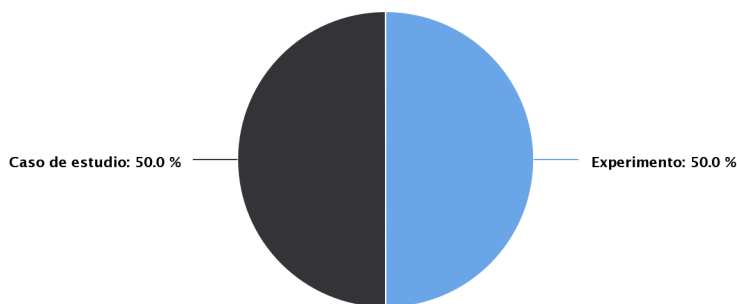


Figura 23: Proporción de estrategias empíricas utilizadas en los estudios incluidos.

Para responder la **RQ.4** se extraen los métodos estadísticos utilizados en cada estudio y se mencionan brevemente. Solamente el estudio E13 presentó métodos estadísticos. En el cuadro 12 se listan los métodos estadísticos utilizados y el identificador del estudio en el cual se aplica (correspondiéndose dicho identificador con el utilizado en la planilla de extracción [Fava et al., 2017]). En la figura 24 se observa la proporción entre los estudios incluidos que presentan métodos estadísticos. En la figura 25 se observa la cantidad de ocurrencias de cada método estadístico.

Se aplicó la prueba de rangos con signo de Wilcoxon (*Wilcoxon rank sum test*, test no paramétrico) para comparar las medianas y *Bootstrap resampling test* para comparar promedios sin asumir una distribución normal sobre ciertas

métricas detalladas en el cuadro 13. A partir de dichos tests se puede observar si la diferencia existente entre los valores de cada grupo del experimento es puramente accidental y no existe una diferencia real o no. Se utiliza la notación *grupo P* para el grupo con experiencia previa en PSP. De la misma manera, se utiliza la notación *grupo N* para el grupo sin experiencia previa en PSP.

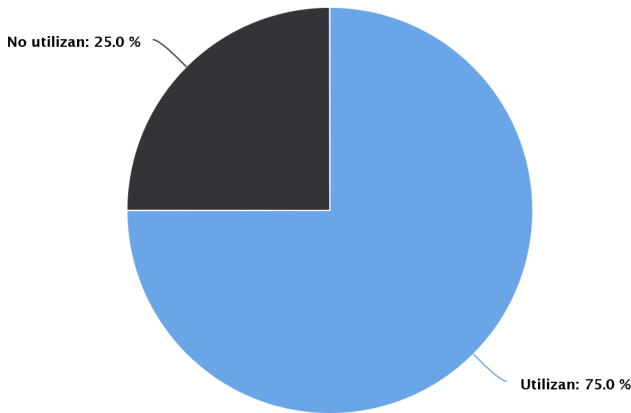


Figura 24: Proporción de estudios que utilizan métodos estadísticos y aquellos que no.

#### 6.3.4 *Análisis cuantitativo*

Al igual que para el piloto 1, se presenta un resumen cuantitativo del proceso de selección y de la extracción de datos del piloto 1. Dicho resumen se puede observar en las figuras 26, 27, 28, 29, 30 y en los cuadros 14 y 15.

#### 6.3.5 *Conclusiones*

El objetivo de este piloto fue validar el protocolo 2 de forma de obtener un protocolo consolidado y robusto para la ejecución de la revisión.

Durante el proceso de selección existió una única discrepancia entre los estudiantes sobre la inclusión o exclusión de los estudios, la cual pudo ser resuelta sin necesidad de consultar con los tutores. Por lo tanto, se considera que los estudiantes lograron familiarizarse con el proceso y confirmar que trabajan bajo los mismos criterios entre sí.

<b>Estrategia empírica:</b> Experimento			
<b>Diferencias estudiadas</b>		<b>Método estadístico</b>	<b>ID</b>
LOC por hora/ LOC por hora estimada para el grupo P	LOC por hora/ LOC por hora estimada para el grupo N	Wilcoxon rank sum test y Bootstrap resampling test	E13
Tiempo total de desarrollo/ Tiempo total de desarrollo estimado para el grupo P	Tiempo total de desa- rrollo/ Tiempo total de desarrollo estima- do para el grupo N	Wilcoxon rank sum test y Bootstrap resampling test	E13
Tiempo total de desarrollo estimado para el grupo P	Tiempo total de desarro- llo estimado para el gru- po N	Wilcoxon rank sum test y Bootstrap resampling test	E13

Cuadro 12: Métodos estadístico empleados por métrica.

<b>Método estadístico</b>	<b>ID</b>
Wilcoxon rank sum test	E13
Bootstrap resampling test	E13

Cuadro 13: Métodos estadísticos utilizados por estudio.

<b>Fuente de publicación</b>	<b>Estudios</b>
Computer	1
IEEE Transactions on Software Engineering	1
Mobile Information Systems	1
Proceedings of the Conference on Integrating Technology into Computer Science Education, ITiCSE	1

Cuadro 14: Fuente de publicación de los estudios incluidos.

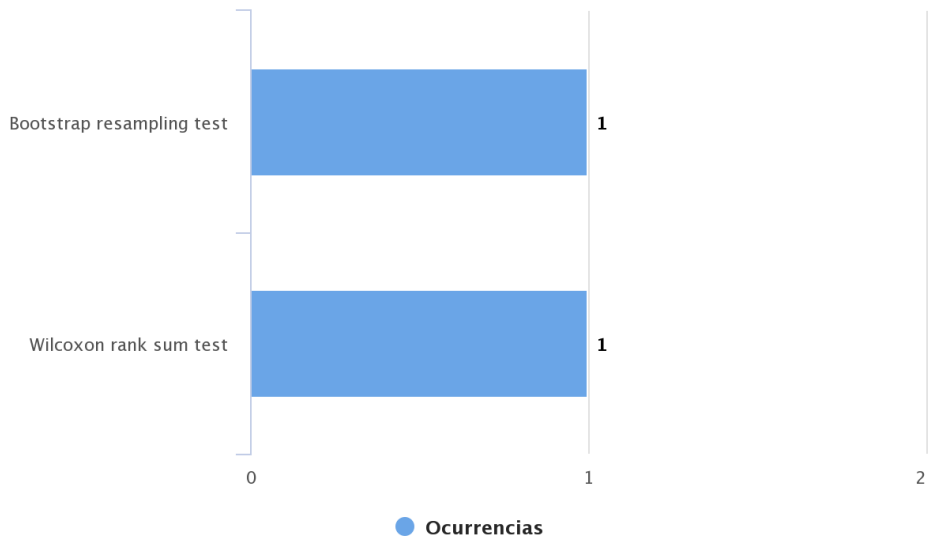


Figura 25: Cantidad de ocurrencias de los métodos estadísticos.

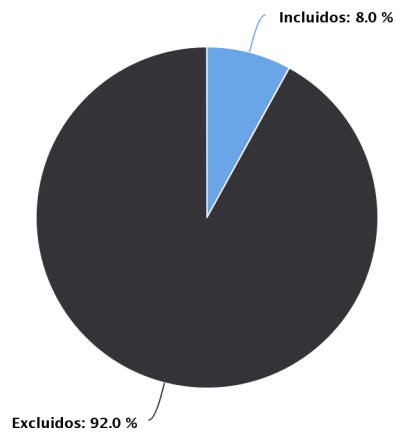


Figura 26: Proporción de estudios incluidos y excluidos.

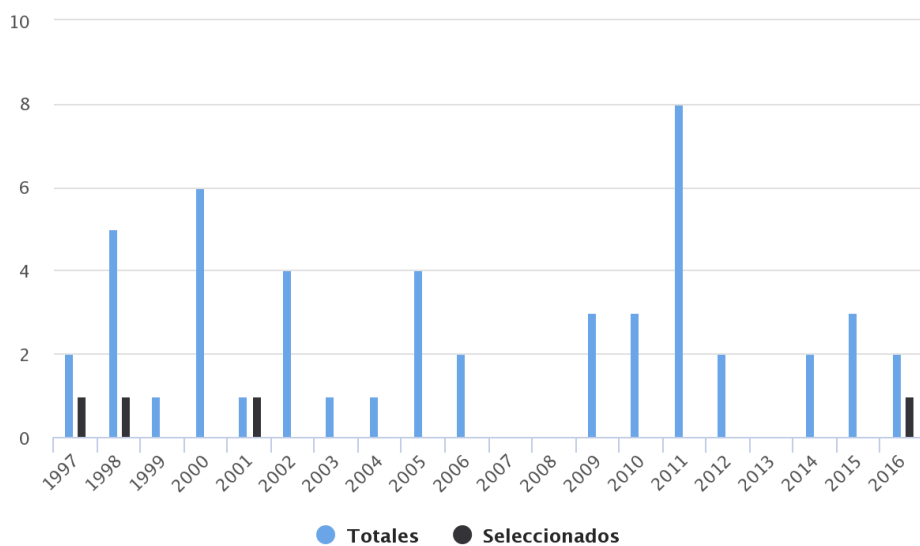


Figura 27: Cantidad de estudios incluidos y totales por año.

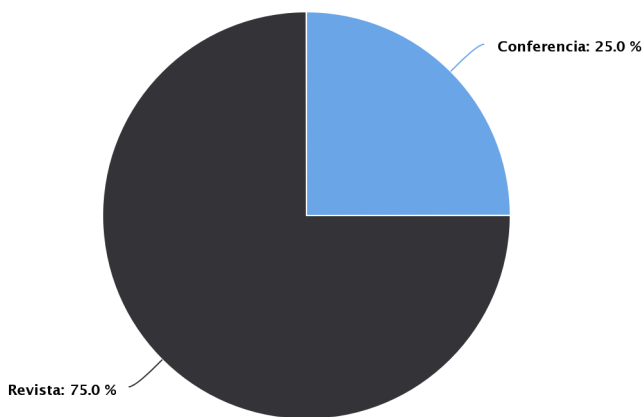


Figura 28: Proporción de medios de publicación.

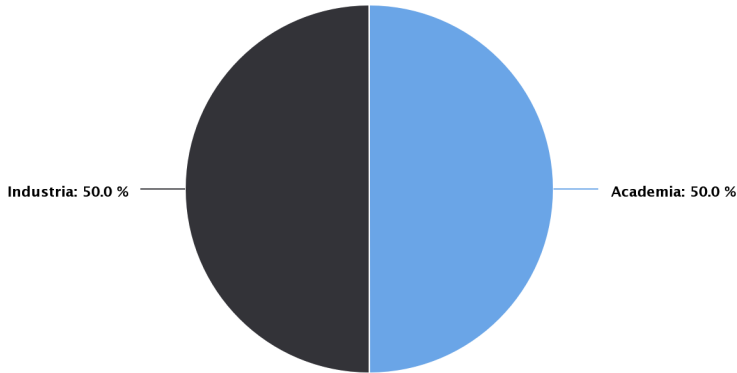


Figura 29: Proporción de contexto de publicación de los estudios incluidos.

<b>Institución</b>	<b>Estudios</b>
Abaxx Technology AG, Forststr. 7, 70174 Stuttgart, Germany	1
Advanced Information Services	1
Carnegie Mellon Univ, Pittsburgh, United States	1
Embry-Riddle Aeronautical University	1
Faculty of Computer Sciences, Preston University, Islamabad, Pakistan	1
Fakultät für Informatik, Universität Karlsruhe, D-76128 Karlsruhe, Germany	1
Motorola	1
Software Engineering Institute	1
Union Switch & Signal	1

Cuadro 15: Instituciones participantes en los estudios incluidos.

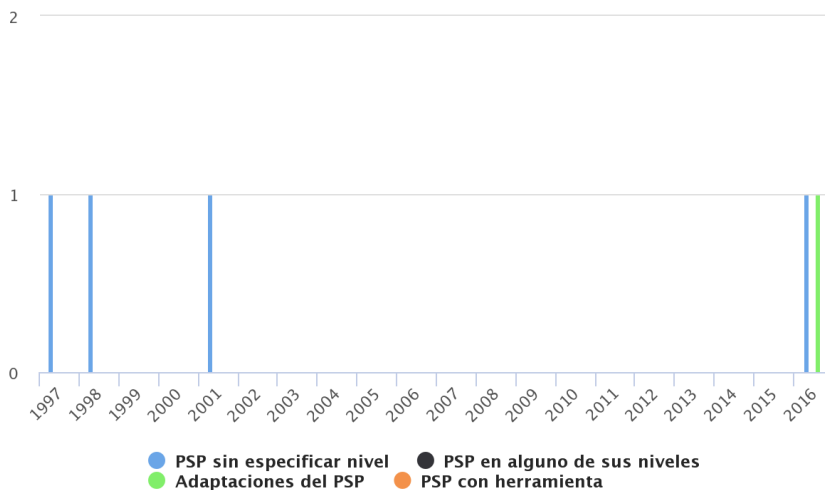


Figura 30: Cantidad de estudios agrupados por proceso utilizado por año.

Debido a que no fue necesario realizar modificaciones en las planillas utilizadas, se consideran validadas para la ejecución de la SLR.

Al responder la RQ.4 si bien los estudiantes sumaron la experiencia del primer piloto continuó habiendo dificultades. Estas dificultades surgieron a la hora de identificar sobre qué variables se aplicaron los tests estadísticos y se requirió validación por parte de los tutores.

Se consideró modificar la cadena de búsqueda para cerciorarse que los estudios arrojados presenten al menos una de las cuatro estrategias empíricas mencionadas, de la siguiente manera: *“Personal Software Process” AND (“Experiment” OR “Survey” OR “Action-Research” OR “Case Study”)*

*“Personal Software Process”* se aplicó sobre *título, resumen y palabras clave*, mientras que los demás términos se aplicaron sobre *todos los campos disponibles*. Sin embargo, al verificar si los 4 estudios incluidos en el piloto 1 se encontraban entre los resultados arrojados por dicha cadena, se comprobó que 2 no formaban parte de los resultados. Se contactó al servicio de asistencia de Scopus planteando dicha problemática y la respuesta obtenida fue que al aplicar búsquedas sobre *“todo los campos disponibles”* la cadena de búsqueda se aplica únicamente sobre todos los campos del texto que se encuentren disponibles en la página del autor. Debido a que en la mayoría de los casos no se encuentra disponible el texto completo en la página del autor, se decidió descartar esta cadena de búsqueda.

Al no existir un criterio claro para determinar que un protocolo se encuentra completamente refinado, en este trabajo se interpreta la ausencia de modificaciones como condición suficiente. Por lo tanto, se asume que la cadena de búsqueda es adecuada, los datos extraídos aportan información para responder las preguntas de investigación y su síntesis permite responder dichas preguntas. Por esto, se considera el protocolo 2 el protocolo final.

#### 6.3.6 *Modificaciones del protocolo 2*

El análisis de los estudios que utilizan caso de estudio y experimento como estrategia empírica no implicó modificaciones en el protocolo 2. Por otra parte, no se pudo refinar el proceso para investigación-acción ni para entrevistas ya que los estudios analizados no cumplieron los criterios de inclusión/exclusión.



## APLICACIÓN DEL MÉTODO DE INVESTIGACIÓN

---

En este capítulo se detalla la ejecución del proceso de selección, siguiendo el proceso definido en el protocolo 2. Por motivos de extensión de tiempo del proyecto de grado, no se ejecutó por completo la SLR, sino que se alcanzó hasta finalizar la ejecución del proceso de selección.

La cadena de búsqueda utilizada fue “*Personal Software Proces*”. La misma fue aplicada sobre *título, resumen y palabras clave* en el buscador Scopus el 7 de diciembre del 2016, arrojando 232 estudios.

### 7.1 PROCESO DE SELECCIÓN

Luego de aplicar la cadena de búsqueda se aplicaron los criterios de inclusión/exclusión. Los estudiantes siguieron un mismo listado de estudios de manera de respetar el orden de lectura. Sobre el total de 232 estudios, se realizó el filtrado previo obteniendo un total de 220 estudios. Luego, sobre este conjunto, se leyó título, resumen y conclusiones y se aplicaron los criterios de selección. Esto dió como resultado un total de 104 estudios incluidos. Para estos 104 estudios se realizó la lectura completa y, luego de aplicados los criterios nuevamente, se obtuvo un total de 38 estudios incluidos. Estos estudios son los que se considerarían en la etapa de extracción de datos. Dichos valores se pueden observar gráficamente en la figura 31.

Durante el proceso de selección se presentaron discrepancias sobre 16 estudios, por lo que se procedió a realizar la discusión grupal sobre cada uno de ellos. Luego de realizada la misma, no se alcanzó un acuerdo sobre 5 estudios. Por lo tanto, se debió consultar a los tutores sobre los siguientes estudios:

- **Estudio 13:** fue discutido con los tutores ya que fue incluido únicamente por el estudiante A y el estudiante B. Este estudio compartía el mismo diseño experimental presentado en el estudio 8 pero aplicado sobre una muestra diferente. Se consideró que se trataban de dos experiencias empíricas independientes por lo que se resolvió incluirlo.
- **Estudio 17:** fue discutido con los tutores ya que fue incluido únicamente por el estudiante A. A pesar de presentar la experiencia de un programador,

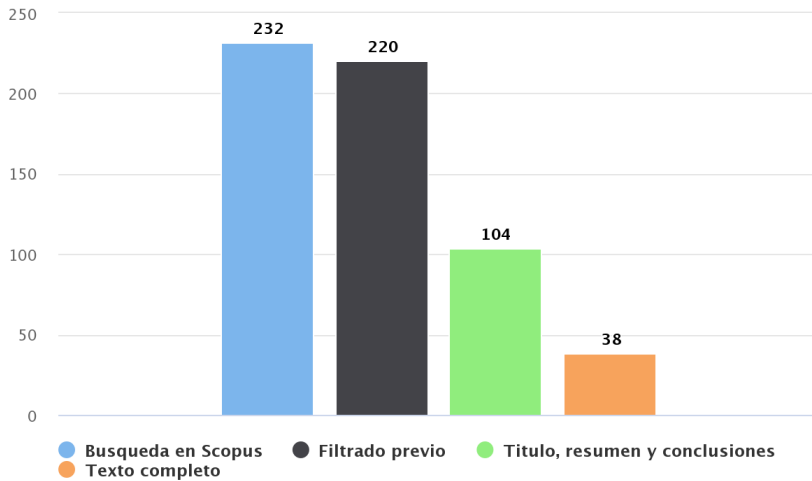


Figura 31: Estudios incluidos por etapa del proceso de selección.

se consideró que no calificaba como una experiencia empírica diseñada, por lo que se decidió excluirlo.

- **Estudio 59:** fue discutido con los tutores ya que fue incluido únicamente por el estudiante A. Se presenta una experiencia donde se aplica una metodología con múltiples iteraciones donde una de ellas es la aplicación de PSP. Por lo tanto, las métricas obtenidas reflejan el efecto de dicha metodología y no del PSP específicamente.
- **Estudio 103:** fue discutido con los tutores ya que fue incluido únicamente por el estudiante A. Los datos que analiza son obtenidos de otro estudio. Sin embargo, se considera un estudio válido ya que utiliza los datos crudos del otro estudio pero para realizar un análisis diferente.
- **Estudio 142:** fue discutido con los tutores ya que fue incluido únicamente por el estudiante A y el estudiante B. A pesar de presentar la experiencia de un programador, se consideró que no calificaba como una experiencia empírica diseñada, por lo que se decidió excluirlo.

Como resultado de la discusión grupal junto con el apoyo de los tutores, se incluyeron los estudios detallados en el cuadro 16.

<b>ID</b>	<b>Título</b>	<b>Autores</b>
8	Improving reliability of defects logging in MVC-PSP	Nachiengmai W., Ramingwong S.
9	A Case of Engineering Quality for Mobile Healthcare Applications Using Augmented Personal Software Process Improvement	Ghayyur S.A.K., Awan D., Sikander Hayat Khiyal M.
13	Implementing personal software process in undergraduate course to improve model-view-controller software construction	Nachiengmai W., Ramingwong S.
29	The development of a design tool for personal software process (PSP)	Chaiyo Y., Ramingwong S.
39	Improving PSP education by pairing: An empirical study	Rong G., Zhang H., Xie M., Shao D.
58	Software defect detection and process improvement using personal software process data	Gopichand M., Swetha V., Ananda Rao A.
76	Investigating the effect of pair programming and software size on software quality and programmer productivity	Sison R.
77	A modified spiral model using PSP, TSP, and six sigma process model for embedded systems	Shaout A., Chhaya T.
79	The impact of design and code reviews on software quality: An empirical study based on PSP data	Kemerer C.F., Paulk M.C.
84	Quality improvement applying design and code reviews for developing small programs	Lopez-Martin C.
86	Cost benefit analysis of personal software process training program	Taek L., Dookwon B., Hoh P.I.
89	Teaching disciplined software development	Rombach D., Münch J., Ocampo A., Humphrey W.S., Burton D.

103	Factors affecting personal software quality	Paulk M.C.
121	The impacts of quality and productivity perceptions on the use of software process improvement innovations	Green G.C., Hevner A.R., Collins R.W.
128	Reliability improvement of web-based software applications	Dávila-Nicanor L., Mejía-Alvarez P.
130	Are found defects an indicator of software correctness? An investigation in a controlled case study	Runeson P., Jönsson M.H., Scheja F.
149	The personal software process: Experiences from Denmark	Abrahamsson P., Kautz K.
153	Personal software process: Classroom experiences from finland	Abrahamsson P., Kautz K.
154	The personal software process in practice: Experience in two cases over five years	Grütter G., Ferber S.
157	Is prior knowledge of a programming language important for software quality?	Wohlin C.
162	An experiment measuring the effects of Personal Software Process (PSP) training	Prechelt L., Unger B.
167	Experiences from teaching PSP for freshmen	Runeson P.
168	Integrating pair programming into a software development process	Williams L.
171	Incorporating PSP into a traditional Software Engineering course: An experience report	Maletic J.I., Howald A., Marcus A.
174	Critical factors affecting personal software processes	Zhong X., Madhavji N.H., El Emam K.
178	Personal Software Process (PSP) concept applied to beginning Engineers	Anneberg L., Ferguson R.

179	Replicated empirical study of the impact of the methods in the PSP on individual engineers	Wesslén A.
195	Understanding software defect detection in the personal software process	Wohlin C., Wesslen A.
196	Applying the personal software process in CS1: An experiment	Hou L., Tomayko J.
197	Using a Personal Software Process to improve performance	Hayes, W.
204	Using the personal software process to motivate good programming practices	Grove R.F.
217	Results of applying the personal software process	Ferguson P., Humphrey W.S., Khajenoori S., Macke S., Matvya A.
218	Doing quality work: The role of software process definition in the computer science curriculum	Hilburn T.B., Towhidnejad M.
223	Using a defined and measured personal software process	Humphrey W.S.
227	Personal software process: An experiential report	Khajenoori S., Hirmanpour I.
228	Introducing the personal software process	Humphrey W.S.
229	Experiencing disciplined software engineering at the personal level	McAlpin John, Liu Jiang B.
233	Personal process in software engineering	Humphrey W. S.

Cuadro 16: Estudios incluidos luego del proceso de selección

## 7.2 ANÁLISIS CUANTITATIVO

A continuación se presenta un resumen cuantitativo del proceso de selección. En la figura 32 se representa el porcentaje de estudios incluidos tras el proceso de selección. En la figura 33 se representa una gráfica de la cantidad de estudios totales y seleccionados por año.

La figura 34 presenta el porcentaje de los medios de publicación de los estudios. De forma complementaria, en el Cuadro 17 se detallan las fuentes de publicación de los estudios. A partir del cuadro 18 se pueden observar las instituciones participantes. Es importante tener en cuenta que algunos estudios fueron realizados por varias instituciones en conjunto.

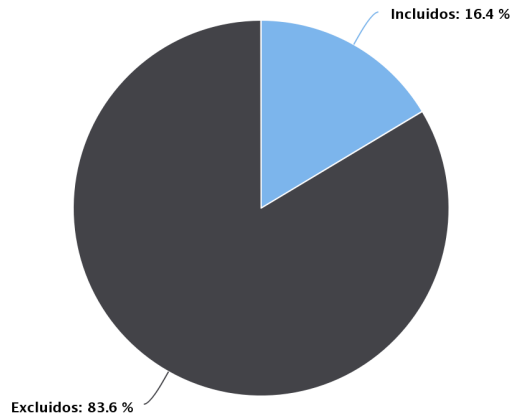


Figura 32: Proporción de estudios incluidos y excluidos.

Fuente de publicación	Estudios
Lecture Notes in Computer Science	3
Software Engineering Education Conference	3
SIGCSE Bulletin (Association for Computing Machinery, Special Interest Group on Computer Science Education)	3
International Multi-Conference on Engineering and Technological Innovation	2
IEEE Software	2
International Symposium on Software Reliability Engineering	2
IEEE Transactions on Software Engineering	2
International Conference on Information, Science and Security	1

Mobile Information Systems	1
Lecture Notes in Electrical Engineering	1
International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology	1
International Conference on Software Engineering	1
Asia-Pacific Software Engineering Conference, APSEC	1
International Arab Journal of Information Technology	1
Journal of Systems and Software	1
CrossTalk	1
Information and Software Technology	1
International Conference on Quality Software	1
Conference Proceedings of the EUROMICRO	1
Computers in Education Journal	1
Empirical Software Engineering	1
International Symposium on Empirical Software Engineering	1
International Software Metrics Symposium	1
Computer	1
Annals of Software Engineering	1
IEEE Pacific RIM Conference on Communications, Computers, and Signal Processing	1
International Conference on the Software Process	1

Cuadro 17: Fuente de publicación en los estudios incluidos.

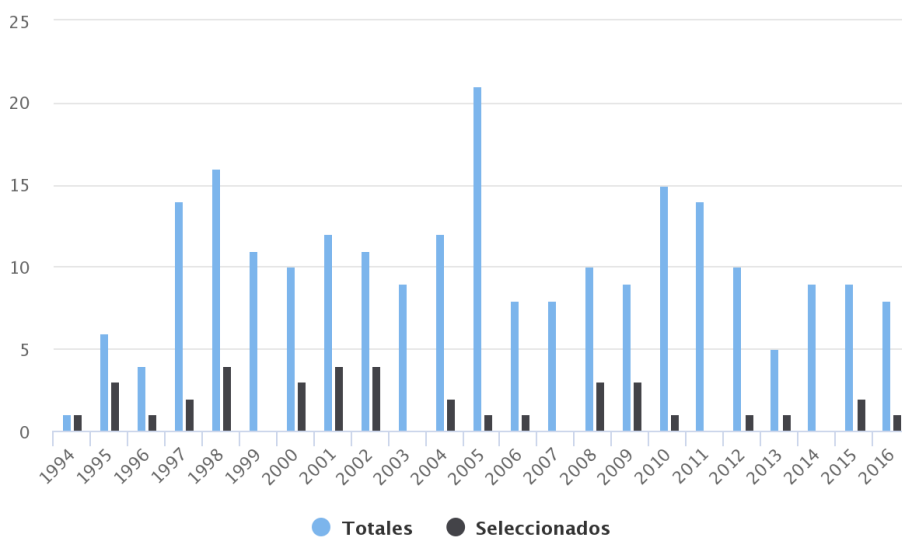


Figura 33: Cantidad de estudios incluidos y totales por año.

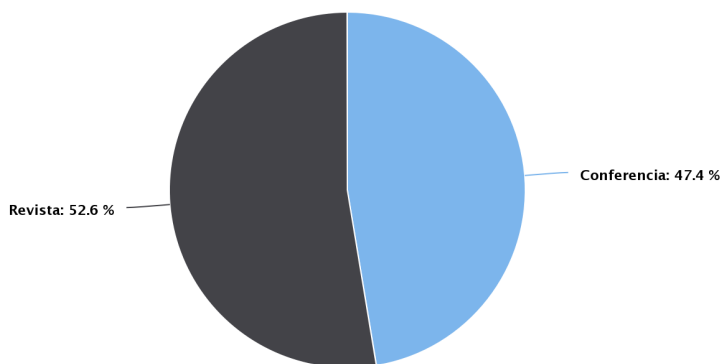


Figura 34: Proporción de medios de publicación.

<b>Institución</b>	<b>Estudios</b>
Carnegie Mellon University, Pittsburgh, United States	9
Computer Engineering Department, Faculty of Engineering, Chiang Mai University, Thailand	3
Dept. of Communication Systems, Software Engineering Research Group, Lund University, SE-221 00 Lund, Sweden	3
Department of Computer Science, Embry-Riddle Aeronautical University, Daytona Beach, FL 32114, United States	3
Bradley University, United States	2
Copenhagen Business School, Department of Informatics, Howitzvej 60, 3., DK-2000, Denmark	2
Technical Research Centre of Finland, VTT Electronics, P.O. Box 1100, Fin-90571 Oulu, Finland	2
University of Pittsburgh, 278A Mervis Hall, Pittsburgh, PA 15260, United States	2
Faculty of Computer Sciences, Preston University, Islamabad, Pakistan	1
State Key Laboratory for Novel Software Technology, Software Institute, Nanjing University, Nanjing, China	1
National ICT Australia	1
University of New South Wales, NSW, Australia	1
Dept. of CSE, BVRIT, Narasapur, Medak, India	1
Dept. of CSE, JNTUA, Anantapur, India	1
College of Computer Studies, De la Salle University, Manila, Philippines	1
University of Michigan-Dearborn, Michigan, ND, United States	1
Information Systems Departament, CUCEA, Guadalajara University, Jalisco, Mexico	1
Department of Computer Science and Engineering, Korea University	1

Fraunhofer Institute for Experimental Software Engineering, Fraunhofer-Platz 1, 67663 Kaiserslautern, Germany	1
Institute of Electrical and Electronics Engineers, American Society for Quality, India	1
Information Systems Department, Hankamer School of Business, Baylor University, P.O. Box 98005, Waco, TX 76798, United States	1
Info. Syst. and Decis. Sci. Dept., College of Business Administration, University of South Florida, Tampa, FL 33620, United States	1
CINVESTAV-IPN, Sección de Computación, Av. I.P.N. 2508, Zacatenco. México, DF. 07300, Mexico	1
Robert Bosch GmbH, Corporate Research and Development, Eschborner Landstraße 130-132, Frankfurt, Germany	1
Line Information GmbH, Garbershoff 4, Helmstorf, Germany	1
Department of Software Engineering and Computer Science, Blekinge Institute of Technology, Box 520, Ronneby, Sweden	1
Fakultät für Informatik, Universität Karlsruhe, D-76128 Karlsruhe, Germany	1
IEEE Germany	1
Abaxx Technology AG, Forststr. 7, 70174 Stuttgart, Germany	1
Department of Computer Science, North Carolina State University, Raleigh, NC 27695-7534, United States	1
Div. of Computer Science, Dept. of Math. Sciences, University of Memphis, Memphis, TN 38152-3240, United States	1
Department of Computer Science, University of Quebec, Montreal, Que., Canada	1
National Research Council, Ottawa, Ont., Canada	1

Department of Electrical Engineering, Lawrence Technological University, Southfield, MI 48075, United States	1
Computer Science and Information Systems, Grand Valley State University, Allendale, MI, United States	1
Indiana University of Pennsylvania, Indiana, PA 15705, United States	1
Advanced Information Services, United States	1
Motorola	1
Union Switch and Signal, United States	1
Advanced Information Services, Inc.	1
Chicago Software Proc. I., Heartland SPIN, United States	1
University of Chicago, United States	1
Illinois Institute of Technology, United States	1
University of Central Florida, United States	1
Lebanon Valley College, United States	1

Cuadro 18: Instituciones participantes en los estudios incluidos.

### 7.3 CONCLUSIONES

Debido a que únicamente se ejecutó el proceso de selección de la SLR, las conclusiones presentadas en esta sección serán relacionadas a este proceso.

Se identificaron distintos patrones que permitieron descartar un gran número de estudios. Por ejemplo, aquellos estudios enfocados en evaluar la enseñanza del PSP y no la aplicación o el efecto del mismo en el software. Otro patrón detectado fue que varios estudios realizan experiencias empíricas donde aplican PSP pero sin ser el foco del estudio y sin presentar resultados relevantes para la investigación.

Durante la etapa de filtrado por título, resumen y conclusiones en varias ocasiones se accedió al resumen de los estudios mediante el link que provee Scopus y no se descargó el estudio completo. Esto fue así ya que en algunas oportunidades la lectura del título y resumen bastaba para tomar una decisión de inclusión o exclusión. Además, era significativamente más rápida esta opción que realizar la búsqueda del estudio mediante la web. La desventaja que esto trajo consigo fue

que algunos resúmenes se presentan en idioma inglés, pero al momento de realizar la lectura completa del estudio se comprobaba que el idioma era otro. Por lo tanto estos estudios fueron descartados en la etapa de lectura completa, cuando, en caso de haber sido descargados previamente, hubieran sido descartados durante la etapa de lectura de título, resumen y conclusiones. Esto no impacta directamente en la SLR ya que no varía el resultado final de estudios incluidos luego del proceso de selección, pero sí varía los resultados intermedios del mismo. Para el caso puntual de esta ejecución, esto sucedió únicamente para un estudio de los 220 considerados.

Respecto a la discusión grupal se destaca que se vio entorpecida ya que las justificaciones de exclusión realizadas a nivel individual por los estudiantes no eran tan detalladas como se necesitó para dicha instancia. Además que las decisiones de inclusión no fueron justificadas. Por lo tanto, durante la resolución de discrepancias fue necesario releer algunos estudios para tomar una decisión.

## CONCLUSIONES Y TRABAJO FUTURO

---

En este capítulo se presentan las consideraciones finales del proyecto y se dan pautas de trabajo futuro.

### 8.1 CONCLUSIONES

El objetivo principal de este trabajo fue recopilar los beneficios conocidos del PSP en el desarrollo de software. Para realizarlo, se definió y validó un protocolo de trabajo para la ejecución de una revisión sistemática de la literatura sobre los efectos del PSP en el desarrollo de software. Para llevar a cabo la SLR fue necesario realizar un estudio sobre revisiones sistemáticas de la literatura como método de investigación empírica.

Se elaboró una guía de trabajo, la cual resultó indispensable para adquirir conocimientos acerca de las SLR en general y para aplicarla al PSP en particular. La guía propuesta abarca conceptos como: identificar la necesidad de realizar una SLR, especificar las preguntas de investigación y desarrollar y validar el protocolo de trabajo. Durante la realización de esta SLR fue posible identificar conceptos clave de un gran número de estudios y extraer datos de los mismos de forma precisa y concisa. Se destaca también la experiencia obtenida a lo largo del proyecto de grado, logrando tener un primer acercamiento a la utilización del método científico experimental en la investigación en el área de ingeniería de software.

Con respecto a las SLR como método de investigación se considera que su principal ventaja es que, debido a que se trata de un proceso definido y riguroso, asegura que los resultados obtenidos son tanto válidos como replicables. Como punto débil se destaca que, a pesar de que una SLR puede estar correctamente definida, queda sujeta al investigador que la ejecute. Por ejemplo, el proceso de selección depende de la comprensión que se haga del estudio y si se considera que aplican los criterios de selección o no. Para mitigar esto, tanto el proceso de selección como la extracción de datos de todos los estudios fue realizado por los tres estudiantes de manera independiente.

A partir de la guía de trabajo se elaboró un protocolo de trabajo para realizar una SLR sobre los efectos del Personal Software Process en el desarrollo de software. Este protocolo fue luego modificado y validado mediante la ejecución

de dos experiencias piloto (sobre un total de 8 estudios), obteniendo el protocolo final de trabajo. Es importante destacar que el mismo puede ser utilizado por otros estudiantes o investigadores para replicar esta SLR o expandirla a otros buscadores.

Una vez obtenido el protocolo de trabajo final, se ejecutó el proceso de selección de la SLR sobre un total de 232 estudios. El resultado de dicho proceso fue un total de 38 estudios incluidos, siendo éstos el 16 % del total. De dichos estudios se desprende que la investigación sobre los efectos del PSP en el desarrollo de software comenzó en el año 1994 y se ha distribuido de manera homogénea, existiendo un promedio de entre una y dos publicaciones por año (exceptuando los años 1999, 2003, 2007 y 2011, donde no existieron publicaciones). Sin embargo, entendemos que no es una temática que se encuentre ganando relevancia con el paso de los años, ya que el máximo de publicaciones se alcanzó hace casi dos décadas, durante el período entre 1998 y 2002 con un total de 15 publicaciones. Desde entonces, han pasado 15 años, y el total de publicaciones alcanza tan solo un total de 16 estudios, superando en uno el valor alcanzado en 1998-2002.

Las investigaciones realizadas provienen de un total de más de 10 países, por lo que entendemos que, a lo largo del tiempo, el interés se presentó a nivel mundial y no centralizado en una región específica. Sin embargo, las publicaciones más recientes (los últimos 5 años) se encuentran en el continente asiático (particularmente Tailandia, Pakistán y China).

A pesar de que los datos obtenidos mediante las experiencias piloto no buscan responder a las preguntas de investigación, se presentan las respuestas a las preguntas de investigación como resultados preliminares de la SLR.

**RQ.1: ¿Qué efectos produce aplicar el PSP en el desarrollo de software?** Para responder esta pregunta se consideró únicamente aquellos estudios que utilizaron PSP en contraposición a aquellos que no lo hicieron, ya que se consideró que los efectos producidos al aplicar variaciones del PSP no aportan información relevante a esta pregunta. Por lo tanto, esta pregunta fue respondida únicamente a partir de la información extraída de dos estudios (E8 y C5 del piloto 2). Los mismos muestran que se mejoró significativamente la estimación de tiempo, el rendimiento y los defectos por KLOC. Además, disminuyó el tiempo de integración entre partes individuales. Por lo tanto, a partir de los datos extraídos de los dos estudios anteriormente mencionados, entendemos que la aplicación de PSP produce un efecto positivo en el desarrollo de software, en comparación con la ausencia de una metodología.

**RQ.2: ¿Qué medidas se utilizan a la hora de evaluar los efectos producidos al aplicar PSP?** Se identificó un total de 27 métricas utilizadas, las

cuales buscan medir distintos aspectos del desarrollo de software como calidad, costo, tamaño, tiempo y productividad. LOC/hora parece ser la métrica de mayor interés ya que se encuentra presente en la mayoría de los estudios incluidos (75 %). En segundo lugar figura el tiempo total de desarrollo, presente en un 63 % de los estudios y luego los defectos removidos cada KLOC, presente en un 50 % de los estudios. A partir de esto interpretamos que, a pesar de que el PSP se centra en la calidad del desarrollo de software, lo que más se busca medir al aplicar PSP es el efecto en la productividad. A esto le sigue el efecto en el tiempo y, en tercer lugar, en la calidad.

**RQ.3: ¿Qué estrategias empíricas se utilizan, y cómo se describen, para evaluar los efectos producidos al aplicar PSP?** Las estrategias empíricas utilizadas a la hora de evaluar el PSP son los experimentos y casos de estudio, ya que no se incluyeron estudios que utilizaran ningún otro tipo de estrategias. Entendemos que esto es así ya que para evaluar los efectos del PSP en la industria, la estrategia empírica que mejor se adapta por su definición son los casos de estudio. Por otra parte, para medir los efectos en un contexto académico, se utilizan experimentos con estudiantes en universidades.

**RQ.4: ¿Qué métodos estadísticos se utilizan para analizar los resultados?** En cuanto a los métodos estadísticos, no se encuentran muy presentes en la experimentación sobre PSP ya que menos de la mitad de los estudios incluidos los utilizaron. Los métodos presentes fueron regresión cuadrática, F-test, ANOVA, prueba de rangos con signo de Wilcoxon y Bootstrap resampling test. Una de las dificultades detectadas durante este trabajo fue el escaso conocimiento en esta área como estudiantes de grado. Esto, sumado a que los estudios publicados no presentan demasiada información sobre la aplicación de los métodos estadísticos, significó un desafío al momento de responder esta pregunta de investigación.

Como ya fue mencionado, del protocolo final se alcanzó a ejecutar el proceso de selección. La extracción y el análisis de datos no fueron realizados. El motivo de ésta decisión se debió a la extensión de tiempo del proyecto de grado. La fase de planificación, durante la cual se definió y validó el protocolo de trabajo, insumió más tiempo del planificado debido a que no existe una condición para determinar cuándo un protocolo se encuentra lo suficientemente validado. Por lo tanto, a pesar de que se había definido que el protocolo de trabajo sería validado mediante la ejecución de una experiencia piloto, al finalizar el piloto 1 se consideró necesario realizar un nuevo piloto (piloto 2) para asegurar que fueran contempladas la mayor cantidad de estrategias empíricas posibles (encuestas, casos de estudio, investigación-acción y experimentos).

Haber realizado el proceso de selección, y haber obtenido 38 estudios incluidos, permite que otros estudiantes o investigadores ejecuten directamente la extracción y el análisis de datos, sin necesidad de realizar el proceso de selección.

## 8.2 TRABAJO FUTURO

En cuanto a posible trabajo futuro, es clara la posibilidad de ejecutar el protocolo de trabajo por completo (es decir, realizar las etapas de extracción y análisis de datos). Esto permitiría responder las preguntas de investigación e identificar nuevas líneas de trabajo.

Por otro lado, se podría extender la búsqueda realizada de forma de cubrir diferentes librerías digitales como ACM Digital, IEEE Xplore, entre otras. De esta manera, se podría trabajar sobre un conjunto de estudios más amplio y obtener resultados más significativos.

De forma complementaria, se podrían definir criterios de calidad de estudios para filtrar resultados que se consideran dudosos (ya sea por la definición del experimento, el tamaño de la muestra, la motivación, etc.).

## BIBLIOGRAFÍA

---

- IEEE Xplore Digital Library. URL <http://ieeexplore.ieee.org>.
- ACM Digital Library. URL <http://dl.acm.org>.
- CiteSeerX. URL <http://citeseer.ist.psu.edu>.
- Engineering Village. URL <https://www.engineeringvillage.com>.
- The Institution of Engineering and Technology. URL <http://www.theiet.org>.
- Google Académico. URL <https://scholar.google.com>.
- ScienceDirect. URL <http://www.sciencedirect.com>.
- Scopus. URL <https://www.scopus.com>.
- SpringerLink. URL <https://link.springer.com>.
- Systematic review toolbox. URL <http://systematicreviewtools.com>.
- Acerenza, F. ¿Qué sabemos de las revisiones individuales de software fuera del Personal Software Process? Seminario #6. Grupo de Ingeniería de Software, Facultad de Ingeniería Universidad de la República, Montevideo, Uruguay, 2015.
- Apa, C.; Robaina, R.; de León, S., and Vallespir, D. Conceptos de Ingeniería de Software Empírica. Technical Report RT 10-02, Instituto de Computación, Facultad de Ingeniería Universidad de la República, Montevideo, Uruguay, 2010.
- Brereton, P.; Kitchenham, B.; Budgen, D.; Turner, M, and Khalil, M. Lessons from applying the systematic literature review process within the software engineering domain. *Journal of Systems and Software*, 80:571–583, 2007.
- Fava, V.; Mañana, M, and Villaverde, J. Planillas de soporte a SLR, 2017. URL <https://www.fing.edu.uy/node/29158>.
- French, W and Bell, C. *Desarrollo organizacional. Aportaciones de las ciencias de la conducta para el mejoramiento de la organización*. Prentice-Hall, 1996.
- Genero, M.; Cruz-Lemus, J., and Plattini, V. *Métodos de investigación de ingeniería del software*. RA-MA, 2014.

Humphrey, W. *A Discipline for Software Engineering*. Addison-Wesley, 1995.

Humphrey, W. The Personal Software Process. Technical Report CMU/SEI-2000-TR-022 ESC-TR-2000-022, Software Engineering Institute, Carnegie Mellon University, Pittsburgh, PA, USA, 2000.

Khan, K.; ter Riet, G.; Glanville, J.; Sowden, A.; , and Kleijnen, J. Undertaking Systematic Review of Research on Effectiveness. crd's Guidance for those Carrying Out or Commissioning Reviews. Technical Report IBSN 1 900640 20 1, NHS Centre for Reviews and Dissemination, University of York, England, 2001.

Kitchenham, B. and Charters, S. Guidelines for performing Systematic Literature Reviews in Software Engineering. Technical Report EBSE 2007-001, Keele University and Durham University Joint Report, 2007.

Marshall, C.; Brereton, P., and Kitchenham, B. Tools to Support Systematic Reviews in Software Engineering: A Feature Analysis. In *18th International Conference on Evaluation and Assessment in Software Engineering*, 2014.

Wohlin, C; Runeson, P; Host, M; Ohlsson, M; Regnell, B, and Wesslen, A. *Experimentation in Software Engineering*. Springer, 2012.



## ANÁLISIS DE HERRAMIENTAS

---

En este capítulo se presenta un análisis de un conjunto de herramientas que brindan soporte a la ejecución de una revisión sistemática de la literatura. Debido a que el proceso a seguir durante una SLR es complejo y riguroso, contar con una herramienta que dé soporte al mismo puede ayudar a obtener resultados de mejor calidad y evitar potenciales errores.

### A.1 CARACTERÍSTICAS DE LAS HERRAMIENTAS

Se define un conjunto de características consideradas necesarias en una herramienta para ser de utilidad en el contexto de esta investigación.

- C.1. Importación de estudios completos en formato PDF, sin necesidad de ser agregados manualmente por parte de los estudiantes, de manera automatizada.
- C.2. Eliminación de potenciales estudios duplicados.
- C.3. Uso colaborativo, soportando múltiples usuarios trabajando de forma independiente y conjunta, cuando se desee.
- C.4. Extracción de metadatos <sup>9</sup> y presentación de forma gráfica de manera automatizada.
- C.5. Filtrado de estudios por metadatos.
- C.6. Inclusión de comentarios a estudios.

### A.2 HERRAMIENTAS ANALIZADAS

Las herramientas a considerar fueron *RevMan*, *Mendeley*, *ReadCube*, *Zotero*, *EndNote*, *Parsifal*, *SESRA*, *SRDB.PRO*, *SLuRp*, *StArt*, *SLRTOOL* y *SLRTOOL*.

Dichas herramientas se obtuvieron del buscador *Systematic Review Tools* [sys] y del estudio *Tools to Support Systematic Reviews in Software Engineering: A*

<sup>9</sup> Metadatos: título, año de publicación, autor y fuente de la publicación.

*Feature Analysis* [Marshall et al., 2014]. Por otra parte, también fueron consideradas herramientas ya conocidas por los estudiantes. Las herramientas analizadas en el estudio se encontraban presentes en el buscador.

*SLR-TOOL*, *SLRTOOL* y *StArt* fueron descartadas previo a su análisis. La razón por la cual se descartaron *SLR-TOOL* y *SLRTOOL* fue su pobre calificación en el estudio anteriormente mencionado debido a la falta de funcionalidades clave. En dicho estudio se establece que *SLR-TOOL* no permite el uso de múltiples usuarios. Además no provee un método automatizado para importar estudios, siendo necesario realizar esta tarea manualmente. Por otra parte, se destaca que las funcionalidades de *SLRTOOL* no se encuentran implementadas eficientemente y su uso colaborativo es pobre. En relación a *StArt*, el estudio no recomienda su uso ya que no permite un uso colaborativo y no presenta ningún aporte a nivel funcional.

A continuación se realiza una breve descripción de las herramientas analizadas.

- **RevMan:** Aplicación para planificar y ejecutar SLR. *RevMan* presenta, para cada etapa de la revisión, la sección que es necesario completar antes de poder avanzar a la siguiente. Esta característica favorece la realización de la revisión ya que actúa como guía, pero no permite ningún tipo de flexibilidad durante su ejecución ni la posibilidad de no realizar ciertas etapas (en caso que no se desee o no apliquen). La interfaz y el uso de la aplicación se consideró poco amigable para el usuario. Por otra parte, no permite un uso colaborativo. Por último, permite la realización de gráficas y tablas, pero no de manera automatizada.
- **Mendeley:** Gestor de referencias bibliográficas y documentos. Esta aplicación permite almacenar documentos en formato PDF. Sin embargo, estos deben ser incluidos de forma manual. Permite un uso colaborativo y realizar comentarios sobre los documentos almacenados.
- **ReadCube:** Gestor de referencias bibliográficas y documentos. Permite almacenar documentos en formato PDF de forma manual. No permite un uso colaborativo. Permite insertar comentarios en los documentos y realizar búsquedas sobre todos los documentos a la vez.
- **Zotero:** Herramienta de investigación para gestionar documentos. Permite almacenar documentos de forma manual, pero no de manera automatizada. Provee un uso colaborativo entre múltiples usuarios. Aunque no se permite adjuntar comentarios a los estudios, es posible insertar etiquetas a los mismos. Además, es posible realizar un filtrado por etiqueta sobre los documentos.

- **EndNote:** Esta aplicación, al igual que las anteriores, únicamente permite la importación de documentos de forma manual. Permite agregar notas a los documentos y realizar búsquedas sobre todos los documentos. A su vez permite un uso colaborativo. Esta herramienta es paga, oscilando entre los U\$S 110 y U\$S 250.
- **Parsifal:** El objetivo de esta herramienta web es apoyar la ejecución de revisiones sistemáticas de literatura en el contexto de la ingeniería de software. *Parsifal* actúa como guía a lo largo del proceso de planificación ayudando en la definición de las diferentes etapas. Durante la ejecución de la SLR en sí (proceso de selección, extracción y síntesis de datos) no aporta valor significativo. Permite la detección de duplicados.
- **SESRA:** Esta herramienta no fue analizada en su totalidad debido a que se detectaron múltiples errores durante su uso que terminaban su ejecución. Por lo tanto, no se la consideró una aplicación sólida y no se tuvo en cuenta durante el análisis.
- **SRDB.PRO:** Al igual que *Parsifal*, esta herramienta guía el proceso de planificación de la SLR pero no aporta ninguna funcionalidad considerable durante la ejecución de la misma. Permite la eliminación de duplicados. Por otra parte, es una herramienta paga (con un costo mensual de aproximadamente U\$S 160).
- **SLuRp:** A pesar de ser la herramienta recomendada en el estudio *Tools to Support Systematic Reviews in Software Engineering: A Feature Analysis*, la misma no pudo considerarse en el análisis de herramientas. Esto se debe a que el proceso de instalación no pudo completarse con éxito y los autores de la herramienta no fueron capaces de brindar una solución cuando se les consultó. Por lo tanto debió ser descartada.

### A.3 EVALUACIÓN DE LAS HERRAMIENTAS

El cuadro 19 presenta una comparación de las herramientas analizadas, marcando con una **X** las características cubiertas por cada herramienta.

Como se puede observar en el cuadro 19, ninguna de las herramientas satisface todas las características deseables. *Mendeley* y *Zotero* resultan ser las herramientas que cubren más características, soportando únicamente 2 de ellas.

Herramienta	C.1	C.2	C.3	C.4	C.5	C.6
RevMan						
Mendeley			X			X
ReadCube						X
Zotero			X		X <sup>1</sup>	
EndNote			X			
Parsifal		X				
SRDB.PRO		X				

<sup>1</sup> A pesar que no permite el filtrado por metadatos directamente, se podrían agregar las etiquetas correspondientes y así cumplir con la característica C.5.

Cuadro 19: Tabla comparativa de herramientas.

#### A.4 CONCLUSIONES

Debido a que no se encontró una herramienta que realmente brinde un apoyo significativo durante todo el proceso, se decidió no utilizar ninguna herramienta durante la ejecución de la SLR. Por lo tanto, para dar soporte a la SLR se construyó un conjunto de planillas para documentar el proceso de selección de estudios y extracción de datos.

## EJERCICIOS ESTÁNDARES DEL PSP

El curso introductorio de PSP consta de la realización de 10 ejercicios siguiendo todas las versiones de PSP. Dichos ejercicios son los propuestos en [Humphrey, 1995] y se puede observar en el cuadro 20.

<b>Ejercicio</b>	<b>Proceso</b>
Desviación estándar	PSP 0
Contador de líneas	PSP 0.1
Contador de líneas y objetos	PSP 0.1
Regresión lineal	PSP 1
Integración numérica	PSP 1.1
Intervalo de confianza	PSP 1.1
Correlación	PSP 2
Clasificación	PSP 2.1
Test Chi-cuadrado	PSP 2.1
Regresión múltiple	PSP 3

Cuadro 20: Ejercicios propuestos para el curso de PSP.



## ANTECEDENTES DE SLR SOBRE PSP

---

La búsqueda realizada en Scopus utilizó la siguiente cadena de búsqueda:  
( *TITLE-ABS-KEY ( "Personal Software Process" ) AND TITLE-ABS-KEY ( "systematic literature review" )* ).

La búsqueda no arrojó ningún resultado.

Para el caso de ScienceDirect se realizó exactamente la misma búsqueda que para Scopus y la salida fue nula.

En cuanto a IEEE Xplore, la cadena de búsqueda utilizada fue la anteriormente mencionada pero la misma se aplicó sobre resumen, términos indexados e información de las citas bibliográficas (título del estudio, autor, título de la publicación, etc.). La misma no arrojó resultados.

La cadena de búsqueda utilizada en SpringerLink fue:  
*"Personal Software Process" & "Systematic Literature Review"*.

Los resultados arrojados por el buscador fueron los siguientes:

- Tauterat, T.. *Development of a Method for the Economic Evaluation of Predictive Maintenance.*
- Valverde, M. C., Vallespir, D., Marotta, A., Panach, J. I.. *Applying a Data Quality Model to Experiments in Software Engineering.*
- da Silva, F. Q. B., Suassuna, M., França, A. C. C., Grubb, A. M., Gouveia, T. B., Monteiro, C. V. F., Ebrahim dos Santos, I.. *Replication of empirical studies in software engineering research: a systematic mapping study.*

Tras analizar los tres estudios, ninguno presentaba una revisión sistemática de la literatura sobre el Personal Software Process ni se relacionaba con la temática.

La librería digital CiteSeerX no arrojó ningún resultados tras aplicar la siguiente cadena de búsqueda sobre título, resumen y palabras clave:

*title:( "Personal Software Process" OR "Systematic Literature Review" ) AND abstract:( "Personal Software Process" AND "Systematic Literature Review" ) AND keyword:( "Personal Software Process" OR "Systematic Literature Review" )"*

Finalmente, para el caso de ACM se aplicó la siguiente cadena de búsqueda sobre el resumen:

*(Abstract:Personal and Abstract:Software and Abstract:Process and Abstract:Systematic and Abstract:Literature and Abstract:Review).*

Al igual que la mayoría de las librerías consultadas, no se obtuvo ningún resultado.

Este protocolo sigue las pautas definidas en la guía de trabajo elaborada por los estudiantes.

#### D.1 IDENTIFICAR LA NECESIDAD DE LA REVISIÓN

Existe un interés particular por investigar y recopilar la información existente respecto al Personal Software Process y sus efectos en el desarrollo de software. Debido a esto, se realizó una búsqueda para constatar si existían, de hecho, estudios sistemáticos sobre esta temática. Para esto, se utilizaron las siguientes librerías digitales: Scopus, ScienceDirect, IEEE Xplore, SpringerLink, CiteSeerX y ACM.

Para el caso de Scopus, ScienceDirect, IEEE Xplore, CiteSeerX y ACM se aplica la siguiente cadena de búsqueda “*Personal Software Process AND Systematic literature review*” sobre el título, resumen y palabras clave de los estudios. Para el caso particular de SpringerLink se realizó la búsqueda sobre texto completo (utilizando la misma cadena de búsqueda que en los demás buscadores), ya que es la única opción que ofrece la librería.

Ninguna de las librerías arrojó resultados luego de la búsqueda, a excepción de SpringerLink. Éste último arrojó 3 resultados, ninguno de los cuales se relacionaba con la temática establecida. Por lo tanto, se considera necesario la necesidad de realizar una Revisión Sistemática de Literatura sobre los efectos del Personal Software Process en el desarrollo de software.

Las librerías digitales anteriormente mencionadas se escogieron siguiendo las pautas de la guía de trabajo desarrollada, por lo tanto se considera justificación suficiente para determinar que no se tiene conocimiento de ninguna revisión sistemática sobre la temática a estudiar.

#### D.2 ESPECIFICACIÓN DE LAS PREGUNTAS DE LA SLR

##### D.2.1 Preguntas de investigación (PI)

RQ.1. ¿Qué efectos produce aplicar el PSP en el desarrollo de software?

RQ.2. ¿Qué medidas se utilizan a la hora de evaluar los efectos producidos al aplicar PSP?

RQ.3. ¿Qué estrategias empíricas se utilizan, y cómo se describen, para evaluar los efectos producidos al aplicar PSP?

RQ.4. ¿Qué métodos estadísticos se utilizan para analizar los resultados?

#### D.2.2 *Definición del alcance*

- **Población:** toda aquella persona sobre las cuales se realizó un estudio empírico, ya sea estudiantes o profesionales que utilizaron PSP.
- **Intervención:** utilización del Personal Software Process, en cualquiera de sus niveles o formas de definición.
- **Comparación:** utilización de niveles o formas de definición diferentes a la presentada en la intervención.
- **Salida:** resumen cuantitativo de los resultados obtenidos tras utilizar el proceso.
- **Contexto:** industria y academia.
- **Diseño experimental:** experiencias empíricas.

### D.3 ESTRATEGIA DE BÚSQUEDA

#### D.3.1 *Cadena de búsqueda*

La cadena de búsqueda se aplica sobre el texto completo de los estudios indexados en los recursos de búsqueda. A continuación se presenta dicha cadena:

*“Personal Software Process”*.

La razón por la cual se optó por aplicarla sobre el texto completo (en lugar de únicamente sobre título, resumen y palabras clave) se debe a que podrían existir estudios que, en lugar de incluir la frase *Personal Software Process* en dichas secciones, utilicen simplemente la sigla PSP. Por otra parte, si se incluía dicha sigla como alternativa (es decir, que la cadena de búsqueda resultante fuese *“Personal Software Process OR PSP”*), se admitiría estudios que incluyan la sigla PSP pero que hiciera referencia a algo diferente que al *Personal Software Process*.

A su vez, si se exigía como condición necesaria que tanto la frase como la sigla estuvieran presentes (en cuyo caso la cadena de búsqueda sería “*Personal Software Process AND PSP*”), se perderían estudios que nunca utilizan la abreviación, pero sí la expresión *Personal Software Process*, o viceversa.

### D.3.2 *Recursos de búsqueda*

La librería digital utilizada para realizar la SLR es Scopus. La razón por la cual se opta por ésta se debe a que dicho buscador indexa las demás librerías mencionadas. Sin embargo, han existido casos puntuales en los que otras librerías han devuelto estudios que no se encontraban presentes en Scopus.

### D.3.3 *Criterios de selección*

Se incluyen aquellos estudios en los cuales:

IC1. El estudio trata sobre la aplicación de PSP y sus efectos en el desarrollo de software, presenta métricas para evaluar los mismos y/o métodos estadísticos para analizar los resultados obtenidos.

Se excluyen aquellos estudios que:

EC1. No traten sobre aplicar PSP.

EC2. No presenten información sobre el impacto en el software al aplicar PSP, métricas utilizadas ni métodos estadísticos aplicados.

EC3. No constan de un carácter empírico.

EC4. Ya hayan sido seleccionados previamente.

EC5. No abordan un estudio primario.

EC6. No han sido publicados <sup>10</sup>.

EC7. Los resultados solo están disponibles en formato de resumen o presentación PowerPoint.

EC8. El idioma no es inglés ni español.

---

<sup>10</sup> En caso de poder contactarse con el autor del estudio o conseguirlo mediante una fuente externa se incluiría el estudio (en caso que no violase ningún otro criterio).

#### D.3.4 Proceso de selección

El proceso de selección consta de 3 etapas, como se muestra en la Figura 35, en las que, dependiendo en la que se encuentre, se descartan estudios mediante ciertas consideraciones.

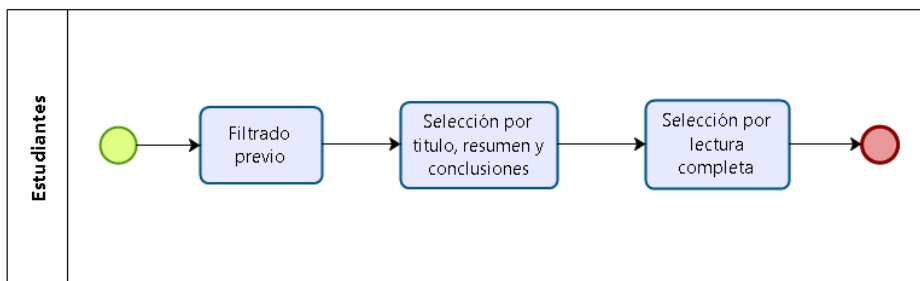


Figura 35: Etapas del proceso de selección

El proceso comienza con los estudios resultantes de aplicar la cadena de búsqueda sobre los recursos mencionados previamente. La primera etapa del proceso de selección es *filtrado previo*, que consiste en la lectura y posible eliminación de aquellos estudios no relacionados con la Ingeniería de Software. También se eliminan actas de conferencias o congresos donde se presenten estudios (que estén relacionados con la Ingeniería de Software o no) debido a que dichas actas proveen un resumen de la charla y no el estudio realizado. Los estudios a filtrar son asignados de forma equitativa a los tres estudiantes. Para realizar dicho filtrado, los estudiantes leen título, resumen y conclusiones de los estudios que le fueron asignados y deciden si descartarlos o no. En caso de que se presenten dudas sobre si descartar un estudio, se lo incluye. De esta manera se lo puede analizar en mayor profundidad en las siguientes etapas.

La siguiente etapa, *selección por título, resumen y conclusiones*, se puede observar en la Figura 36. Esta etapa tiene como entrada los estudios resultantes del *filtrado previo*. Cada estudiante, de manera paralela e independiente, realiza una lectura del título, resumen y conclusiones de dichos estudios, aplicando los criterios de inclusión/exclusión. Luego de su lectura se registra en la planilla de inclusión/exclusión si dicho estudio es incluido o excluido. En caso de ser excluido, se registra en la planilla su justificación. Si un estudio es incluido por al menos un estudiante, éste será considerado para la etapa final del proceso.

La etapa final, *selección por lectura completa*, se puede observar en la Figura 37. Esta etapa tiene como entrada los estudios incluidos en la *selección por título*,

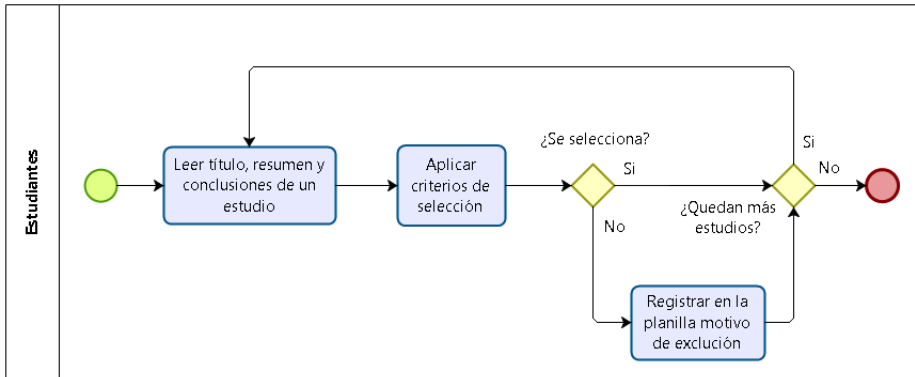


Figura 36: Detalle de la etapa de selección por título, resumen y conclusiones.

*resumen y conclusiones*. Cada estudiante, de manera paralela e independiente, realiza la lectura completa de dichos estudios, aplicando los criterios de inclusión/exclusión. Luego de su lectura se registra en la planilla de inclusión/exclusión si dicho estudio es incluido o excluido. En caso de ser excluido, se registra en la planilla su justificación. Aquellos estudios para los que existe acuerdo en cuanto a su inclusión o exclusión, son registrados en la planilla correspondiente. Por otra parte, se debe llegar a un acuerdo para aquellos en los que existe discrepancia. En caso de no alcanzar una resolución por parte de los 3 estudiantes, la decisión final queda a cargo de los tutores.

Finalmente, como resultado de la tercer etapa, se obtiene una colección de estudios finales, con los cuales se procede a realizar la extracción de datos.

#### D.3.5 Extracción de datos

A continuación se lista la información extraída de los estudios obtenidos.

- Fuente de publicación.
- Año de publicación.
- Autor y organización a la cual pertenece.
- Contexto (Academia/Industria).
- Estrategia empírica (experimento, caso de estudio, encuesta, etc.).
- Descripción del estudio.

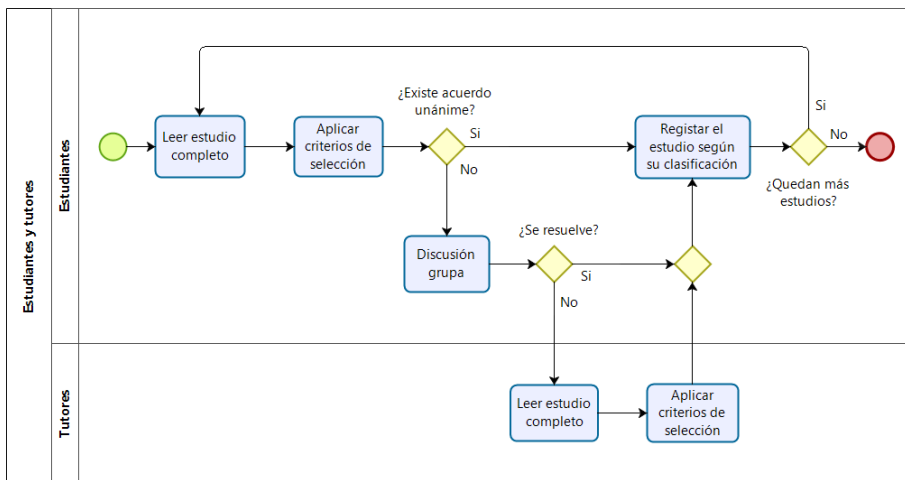


Figura 37: Detalle de la etapa de selección por lectura completa.

- Conclusiones del estudio.
- Niveles de PSP empleados.
- Métricas utilizadas para evaluar los efectos en el desarrollo de software.
- Métodos estadísticos utilizados para analizar los resultados.
- Tamaño de la muestra.

A su vez, para cada estudio, se registra la siguiente información:

- Nombre del responsable de la extracción de los datos
- Fecha de extracción
- Espacio para notas adicionales

#### D.3.6 Síntesis de datos extraídos

Para responder la primera pregunta de investigación (**RQ.1**) se presenta, para cada métrica obtenida, su media antes y después de utilizar el proceso en estudio (especificando el nivel de PSP que se utilizó), su intervalo de confianza y el tamaño de la muestra.

Para el caso de la **RQ.2** se tabulan las diferentes métricas utilizadas en los diferentes estudios.

En cuanto a la **RQ.3**, se describen las diferentes estrategias utilizadas para evaluar el PSP.

Finalmente, para la **RQ.4** se listan los métodos estadísticos utilizados, señalando la estrategia empírica utilizada y la métrica evaluada. En el cuadro 21 se muestra la tabla que se utiliza para presentar dicha información.

Estrategia empírica	
Métrica 1	Método estadístico empleado
Métrica 2	Método estadístico empleado

Cuadro 21: Ejemplo de presentación de datos sobre métodos estadísticos.

A continuación se detallan otros datos extraídos que se presentan en forma gráfica de forma complementaria, además de los anteriormente mencionados.

- Por año
- Cantidad de estudios resultantes tras realizar cada paso del proceso de selección
- Por contexto
- Por estrategia empírica
- Por niveles de PSP involucrados en el estudio

#### D.3.7 *Cronograma de la revisión*

1. Experiencia piloto
  - 1.1. Selección de estudios primarios sobre buscador Scopus
  - 1.2. Extracción de datos
  - 1.3. Síntesis de datos
  - 1.4. Diseminación de datos
  - 1.5. Conclusiones parciales
  - 1.6. Reporte de experiencia piloto
  - 1.7. Correcciones del protocolo

2. Aplicación del protocolo final
  - 2.1. Selección de estudios primarios sobre todos los recursos de búsqueda
  - 2.2. Extracción de datos
  - 2.3. Síntesis de datos
  - 2.4. Diseminación de datos
  - 2.5. Conclusiones
  - 2.6. Reporte final
3. Formateo del reporte final
  - 3.1. Marco teórico
  - 3.2. Resumen