

Evaluación del Uso de un Paquete de Laboratorio en una Replicación Experimental

Martín Solari¹, Cecilia Apa², Sira Vegas³

¹Universidad ORT Uruguay

²Universidad de la República, Uruguay

³Universidad Politécnica de Madrid, España

`martin.solari@ort.edu.uy, ceapa@fing.edu.uy, svegas@fi.upm.es`

Abstract. *Replication of experiments is a necessary step for evolving the knowledge about techniques and tools in software engineering. Lab packages are one of the instruments used for experimental knowledge transfer. This study presents an evaluation of a lab package proposal used in an experimental replication. The objective is to validate that the proposed lab package is a better alternative than unstructured ones. The evaluation results show a high degree of completeness and usability in the proposed lab package. Positive results have been observed in efficacy and efficiency of the replication, and in the satisfaction of the participating researcher.*

1. Introducción

La experimentación es uno de los métodos utilizados para avanzar en el conocimiento científico. Mediante los experimentos se pueden manipular las variables que afectan un fenómeno y descubrir relaciones causa-efecto [Campbell and Stanley 1963]. La experimentación es un enfoque de investigación aceptado por la comunidad de Ingeniería de Software (en adelante IS), sin embargo, todavía presenta desafíos metodológicos importantes [Shull et al. 2008].

El proceso de investigación experimental requiere de replications que permitan confirmar, refutar o refinar los resultados obtenidos. Las replications atenúan las amenazas a la validez introducidas por las restricciones del contexto y las desviaciones introducidas por el propio investigador. Para esto se requieren replications progresivamente diferenciadas, que permitan ampliar la validez de los resultados [Juristo and Vegas 2009]. La replicación es un proceso costoso, que supone esfuerzo y conocimientos específicos por parte del investigador que la realiza.

Con el objetivo de facilitar esta tarea, se utilizan instrumentos de transferencia de conocimiento [Shull et al. 2004], como los paquetes de laboratorio y la comunicación directa entre investigadores. Los paquetes de laboratorio (en adelante PL) son contenedores de las instrucciones y materiales necesarios para replicar un experimento. Los PL son documentos heterogéneos y complejos que contienen distintos niveles de conocimiento sobre el experimento. El objetivo principal de los PL es facilitar la realización de replications que puedan ser integradas al cuerpo de conocimiento experimental [Basili et al. 1996].

El objetivo de este estudio es la validación de una propuesta de estructura de PL para experimentos en IS [Solari 2010]. Se realiza una evaluación de una replicación que

utiliza un PL estructurado de acuerdo a la propuesta. Se pretende validar que el uso de este PL implica una mejora con respecto a otros PL no estructurados. En un paso anterior de validación de la propuesta se comprobó la viabilidad de la misma mediante su instanciación. Esta instanciación fue realizada para otro experimento y reportada de forma separada [Apa et al. 2011].

La replicación que se utiliza para la validación fue realizada en el Grupo de Ingeniería de Software (GrISE) de la Universidad Politécnica de Madrid (UPM). Evaluar el uso del PL en una replicación permite observarlo en su contexto habitual y vinculado con sus objetivos operativos. Este artículo no reporta los resultados experimentales, sino que realiza una evaluación del PL usado en la replicación desde un punto de vista metodológico. Se aplicó un proceso de evaluación definido previamente que permite realizar una comparación con PL usados en replications anteriores.

El resto del artículo se organiza de la siguiente forma. En la sección 2 se resume la estructura de la propuesta de PL. En la sección 3 se describe la replicación utilizada como caso de estudio. En la sección 4 se desarrolla la evaluación de la replicación. Por último, en la sección 5 se presentan las conclusiones del estudio.

2. Características de la Propuesta de Paquete de Laboratorio

El PL utilizado en la replicación es una instancia de una propuesta de estructura genérica. La propuesta está basada en el concepto de familia de experimentos. Se entiende por familia a un conjunto de experimentos con un objetivo relacionado y diseño similar [Basili et al. 1999]. Por este motivo, el PL puede incluir varios experimentos en su estructura.

La estructura fue elaborada para soportar todas las actividades del proceso de investigación experimental, desde el planteamiento del estudio, hasta la interpretación de los resultados. Dicha estructura divide al PL en módulos independientes y vinculados a propósitos específicos. Éstos pueden ser de dos tipos: *módulos experimento* y *módulos estudio*. Los módulos experimento están destinados a explicar el experimento en general, mientras que los de estudio describen cada una de las replications u otros estudios realizados en el marco de la familia de experimentos.

La propuesta de PL está compuesta por tres elementos: una descripción genérica de la estructura que debe tener un PL específico, una plantilla y una lista de verificación del contenido. De esta manera, la propuesta de PL puede utilizarse como una guía en el proceso de organización del conocimiento y los materiales de un experimento de IS. La descripción detallada de la propuesta que se está validando está disponible para ser consultada¹.

La estructura propuesta divide al PL en varios módulos inter-relacionados que se presentan esquemáticamente en la Figura 1. Cada uno de los módulos soporta alguna actividad del proceso de investigación experimental. En la estructuración del contenido se ha tenido en cuenta que el PL es un documento técnico que puede ser leído en forma no-lineal y usado para distintos propósitos.

El módulo de mayor nivel de abstracción es el de la **teoría** del experimento, que establece el marco teórico para la realización del experimento y la interpretación de los

¹<http://www.ort.edu.uy/fi/publicaciones/ingsoft/investigacion/publicados/DesarrolloPropuestaPL.pdf>

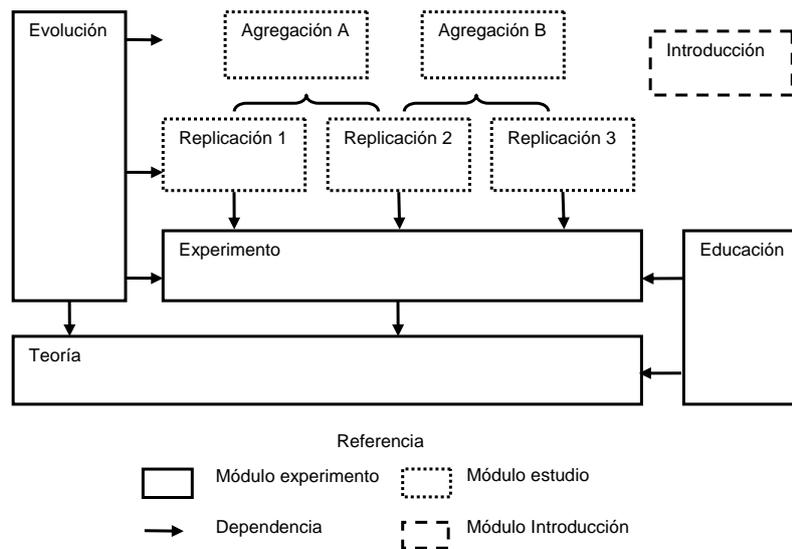


Figura 1. Módulos de la propuesta de PL.

resultados. En el módulo **educación** se recogen los materiales utilizados para el entrenamiento de los sujetos que van a participar del experimento. En el módulo **experimento** se definen los aspectos generales del mismo y las instrucciones para replicarlo. Este módulo está organizado según las actividades del proceso de replicación: planteamiento, diseño experimental, operación y análisis. Además de las instrucciones para cada actividad, se incluye el material operativo y las herramientas necesarias para replicar el experimento.

La estructura también considera la dinámica evolutiva de la investigación experimental. En el módulo **evolución** se realiza un resumen de todos los estudios realizados. Este módulo sirve para establecer la relación entre las versiones del experimento y sus replications, así como para ver la evolución de la familia de experimentos en conjunto. Los módulos **replicaciones** y **agregaciones** guardan la información de cada estudio de la familia de experimentos.

Por tratarse de un documento técnico y operativo, el PL debe cumplir ciertos criterios de calidad de contenido. Para lograr estos objetivos de calidad, en la propuesta se incluye una lista de verificación que es usada en las revisiones del PL. Los aspectos contemplados en la verificación son: orientación a la tarea, organización del material operativo, soporte del proceso experimental, formato y estilo de escritura.

3. Descripción de la Replicación UPM 2011

Para la validación de la propuesta de PL se realizó la replicación de un experimento que estudia la efectividad y eficiencia en detección de defectos de distintas técnicas. El experimento es parte de una línea de investigación que ha evolucionado por más de 10 años [Juristo and Vegas 2003]. La replicación fue realizada en noviembre y diciembre de 2010, aunque por motivos prácticos será denominada UPM 2011. Ésta se enmarca en el curso de Verificación y Validación de Software para estudiantes de pre-grado y post-grado realizado en la Facultad de Informática de la UPM. Se omitirán los detalles de la replicación que no son de interés para la validación del PL.

Durante el transcurso de la investigación experimental, el material operativo e

instrucciones han ido cambiando, evolucionando y adaptándose a los distintos contextos de cada replicación. La replicación evaluada es la primera que usó el PL estructurado en el contexto de esta línea de investigación. La replicación fue realizada por una investigadora que no tenía conocimiento previo del experimento. Esta condición es necesaria para poder evaluar el efecto que produce el PL en la transferencia de conocimiento.

Las actividades que se siguieron en la replicación fueron las definidas en el PL con algunas salvedades: la actividad de análisis aún no se ha realizado y por ende tampoco la interpretación ni agregación de los resultados. Resulta interesante comparar el esfuerzo realizado en la replicación UPM 2011 con el estimado que se brinda como guía en el PL, las horas reales y estimadas se detallan en la Tabla 1.

Cuadro 1. Esfuerzo real y estimado de replicación (en horas-persona).

Nº	Actividad	Estimado PL	Real UPM 2011
1	Reunión preparatoria	3	2.0
2	Estudio del paquete y plan inicial	8	10.0
3	Planteamiento y diseño adaptado	8	8.7
4	Validación del diseño y consulta de dudas	5	15.8
5	Preparación del material	8	34.4
6	Sesiones del experimento	6	21.5
7	Corrección de formularios	21	46.6
Total		59	139.0

En algunas actividades, el esfuerzo invertido en la replicación supera al estimado por el PL. El esfuerzo real en las actividades 1 a 3 (que refieren al planteamiento y preparación del experimento) no tiene casi diferencia con el estimado. Sin embargo, en las actividades 4 a 7 (que refieren al diseño y operativa del experimento) sí presentan variación. La diferencia entre el esfuerzo real y el estimado está directamente relacionada con la cantidad de sujetos participantes, el contexto experimental y los cambios realizados al diseño base del experimento.

El diseño de la replicación UPM 2011 fue más complejo que el diseño base de la familia de experimentos. Resultaba necesario balancear variables como la preferencia de lenguaje de programación y formación previa del estudiante. Por tanto, fueron necesarias más instancias de validación del diseño. En la replicación participaron 63 sujetos y se realizaron dos rondas del ejercicio: una sobre programas en lenguaje C y otra sobre programas en lenguaje Java. Esto supone un esfuerzo de preparación y corrección que fue mayor al habitual para el experimento (que supone 35 sujetos y una única ronda del ejercicio).

Otra de las causas del esfuerzo adicional fue que los sujetos participantes pudieron elegir si realizar el ejercicio en idioma inglés o en idioma español. Esto se debió a que provenían de distintos tipos de programas de estudio, siendo algunos de éstos internacionales. Esto implicó versiones en ambos idiomas de formularios de recogida de datos, especificación de los programas, código fuente, entre otros.

4. Evaluación del Uso del Paquete de Laboratorio

La evaluación del PL usado en la replicación se realizó a partir de un proceso definido, que se ilustra en la Figura 2. Este proceso fue aplicado tanto a esta replicación,

como a las replications anteriores de la misma familia de experimentos que no utilizaron este PL. El proceso de evaluación está basado en métodos cualitativos para identificar, analizar y sintetizar la información relevante [Seaman 1999]. El resultado final consiste en la valoración de atributos que pueden ser comparados con los obtenidos en evaluaciones anteriores.

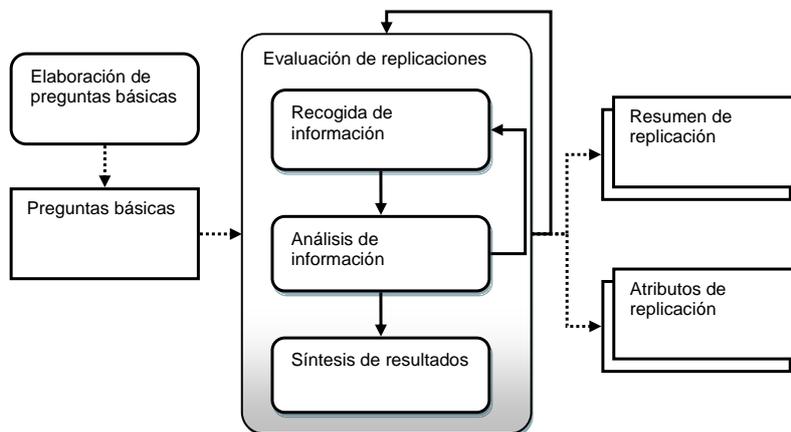


Figura 2. Proceso de Validación.

En la primera actividad del proceso de evaluación se realiza la **recogida de información**. En esta actividad se identifican las fuentes relevantes para evaluar la replicación. Estas fuentes incluyen los documentos y comunicaciones que se hayan usado, junto con el propio PL. Se utiliza como guía para la recogida de información una *lista de preguntas básicas*, que se elaboró antes de iniciar el proceso de evaluación de la replicación. De esta lista de preguntas básicas se deriva un cuestionario planteado al replicador responsable. En el mismo se recogen sus impresiones sobre la realización de la replicación. Si quedaran huecos en la información, se pueden realizar consultas posteriores.

El siguiente paso consiste en el **análisis de la información** recolectada. En esta actividad se analizan las fuentes mencionadas y se obtiene un *resumen de la replicación*. Cada una de las respuestas se codifica para obtener un lenguaje uniforme para describir la replicación. El resumen se utiliza para analizar los incidentes de la replicación. Se entiende por incidente cualquier suceso ocurrido durante la replicación que genere dudas u obstáculos para la realización de la misma.

La última actividad del proceso consiste en la **síntesis de la información**. Esta actividad tiene en cuenta los incidentes de la replicación junto con la información de contexto para valorar *atributos de evaluación*. La mayoría de los atributos son valorados con escalas cualitativas. Para minimizar la subjetividad de la valoración, debe asegurarse que los niveles de valoración estén definidos previamente y sean claros para el evaluador.

4.1. Valoración de Atributos

Los atributos de evaluación se agrupan en distintos aspectos vinculados con los objetivos de los PL. Como instrumentos de transferencia de conocimiento, los PL deben cubrir todas las actividades del proceso de investigación experimental. Esto se evalúa mediante el aspecto **completitud**. Por otro lado debe evaluarse la **usabilidad** del documento para transmitir este conocimiento.

Además de la transferencia de conocimiento, los PL tienen un objetivo operativo: la realización de replicaciones. Esto implica obtener replicaciones útiles, que puedan ser integradas al cuerpo de conocimientos de cierta área. Esto se evalúa mediante el aspecto **eficacia**. Por otro lado, debe considerarse que una replicación es una actividad que consume recursos, en particular el tiempo del replicador. Esto se evalúa en el aspecto **eficiencia**. Otro aspecto evaluado es la **satisfacción** del replicador, para el cual se tiene en cuenta la perspectiva del principal investigador participante.

El aspecto de evaluación **completitud** se divide en cinco atributos más específicos. Cada uno de estos atributos evalúa el comportamiento del PL frente a distintas necesidades del proceso de replicación experimental. Mediante el atributo *alcance* se determina el grado en que el PL cubre las actividades del proceso. El atributo *instrucción* se utiliza para determinar la profundidad de cada sección del PL. La *adaptabilidad* evalúa el grado en que el diseño experimental puede ser aplicado a las restricciones específicas de la replicación. El atributo *escalabilidad* analiza la capacidad del PL para contener los reportes de las distintas replicaciones y soportar la agregación de resultados. Por último, el *versionado* es la capacidad del PL de contener varias versiones del experimento. Los valores de los atributos de evaluación para la completitud de la replicación UPM 2011 pueden verse en la Tabla 2.

Cuadro 2. Valoración de la completitud.

Atributo	Componente	Valor
<i>Alcance / Instrucción</i>	Planificación	Presente / Fundamentado
	Introducción	Presente / Fundamentado
	Diseño	Presente / Fundamentado
	Entrenamiento	Presente / Fundamentado
	Preparación del material	Presente / Detallado
	Operativa	Presente / Fundamentado
	Corrección	Presente / Básico
	Análisis	Presente / Fundamentado
	Reporte	Presente / Detallado
<i>Adaptabilidad</i>		Adaptable parcial
<i>Escalabilidad</i>	Reporte de replicación	Si
	Agregación de resultados	[Pendiente]
<i>Versionado</i>		Versiones recuperables

El atributo alcance se puede valorar con dos niveles: Falta o Presente. En esta replicación se ha utilizado un PL cuyo alcance abarca todas las actividades de replicación. Para el atributo instrucción, los valores posibles en orden creciente de profundidad son: Básico, Detallado y Fundamentado. Las descripciones de cada sección están en los niveles más altos, en la mayoría de los casos. Sin embargo, se han observado oportunidades de mejora para las secciones de preparación del material, corrección y reporte. Estas lecciones aprendidas serán tratadas en la sección 4.2. El diseño experimental ha sido adaptado exitosamente para realizar la replicación, sin embargo, se han utilizado elementos externos al PL. Aunque todavía está pendiente la agregación de resultados, se ha podido integrar al PL un reporte de la replicación realizada en el formato propuesto. El PL usado contiene varias versiones del experimento, aunque existe la posibilidad de mejorar la trazabilidad de los cambios.

La **usabilidad** es un aspecto que se evalúa teniendo en cuenta la perspectiva del

investigador que participó como responsable de la replicación. Para hacer más detallada la evaluación de este aspecto, la usabilidad se descompone en tres atributos: facilidad de *entendimiento*, *aplicación* y *búsqueda* de la información. A su vez, estos atributos se descomponen en componentes más concretos. Cada uno de ellos es evaluado por el replicador usando una escala de cinco valores, siendo 5 el mejor caso y 1 el peor caso. El cuestionario también ofrece la posibilidad al replicador de hacer comentarios directos sobre el PL. Los resultados de la evaluación de la usabilidad para la replicación pueden verse en la Tabla 3.

Cuadro 3. Valoración de la usabilidad.

Atributo	Componente	Valor
<i>Fácil de aplicar</i>	Orientación a la tarea	4
	Precisión	4
	Compleitud	4
<i>Fácil de entender</i>	Claridad	5
	Concreto	5
	Estilo	4
<i>Fácil de buscar</i>	Organización	5
	Capacidad de recuperación	3
	Efectividad visual	4

La evaluación subjetiva recogida del replicador muestra que el PL tiene un grado bueno o muy bueno de usabilidad en casi todos los atributos evaluados. La principal posibilidad de mejora observada está en la facilidad de recuperación. Profundizando en los comentarios sobre este atributo, se observó que el material operativo no era fácil de recuperar a partir del documento principal. La estructura del repositorio de material operativo adjunto al PL puede ser mejorada.

La eficiencia, eficacia y satisfacción son los aspectos de evaluación de más alto nivel. Éstos no solo contemplan el PL usado, sino también el contexto más amplio en que la replicación se realizó. Una replicación puede ser exitosa más allá de los instrumentos aplicados, aunque se puede considerar un indicador indirecto de la calidad del PL que se haya realizado la replicación en forma eficaz. La valoración realizada para estos aspectos puede verse en la Tabla 4.

Cuadro 4. Valoración de la eficacia, eficiencia y satisfacción.

Aspecto	Atributo	Valor
Eficiencia	<i>Esfuerzo real</i>	4 semanas - persona
	<i>Esfuerzo subjetivo</i>	Bajo
	<i>Estimación previa</i>	Aproximada
Eficacia	<i>Resolución de dudas</i>	Alta
	<i>Reproducción del ambiente</i>	Completa
	<i>Severidad media de errores</i>	Leve
Satisfacción	<i>Satisfacción del replicador</i>	Alta
	<i>Lista de beneficios</i>	Aprender experimentación
		Practicar experimentación
		Red de investigación

El aspecto **eficiencia** se divide en tres atributos de evaluación. El *esfuerzo real* es la medición directa o aproximada del esfuerzo que consumió para el investigador la

realización de la replicación. El *esfuerzo subjetivo* es complementario al anterior, analiza en qué grado la tarea resultó ardua para el replicador. También se tiene en cuenta la posibilidad que tuvo el replicador de *estimar en forma previa* el esfuerzo de la actividad. En cuanto al esfuerzo real, esta replicación demandó un tiempo aproximado de 4 semanas personas. Este valor es mayor a otras replicaciones evaluadas para la misma familia de experimento. Sin embargo, el PL permitió estimar previamente el esfuerzo y que el replicador perciba el esfuerzo subjetivo como bajo.

La **eficacia** está dada por la posibilidad de obtener una replicación que se pueda integrar al cuerpo de conocimientos del área. Este aspecto se divide en tres atributos de evaluación. Uno de ellos observa la capacidad de *resolver dudas* en la preparación de la replicación. También se considera una replicación eficaz aquella que puede *reproducir el ambiente* experimental planificado. Por último, se evalúa como atributo la *severidad media de errores* experimentales ocurridos. Los valores de los atributos de eficacia para la replicación evaluada tienen los valores más altos de la escala. Se considera que se ha obtenido una replicación que ha cumplido con los objetivos propuestos. Aunque el análisis de resultados está pendiente, no se registraron incidentes durante la replicación que afectaran la operativa o la corrección. El PL utilizado, junto con otros instrumentos de comunicación, permitió responder las dudas del replicador.

La **satisfacción** es un aspecto de naturaleza subjetiva, que considera el grado en que el investigador participante está conforme con la experiencia. Para realizar una valoración más detallada se recoge como atributo, además de la *satisfacción general*, los *beneficios mencionados* por el replicador. En la replicación evaluada el grado de satisfacción fue alto. El replicador menciona beneficios en tres categorías distintas: aprendizaje de experimentación, práctica del proceso de replicación y oportunidad de participación activa en una red de experimentación.

4.2. Lecciones Aprendidas

Además de la valoración de atributos, la evaluación del uso del PL permitió obtener una lista de lecciones aprendidas. Las mismas abarcan mejoras sobre el PL, como también cambios en el propio experimento replicado. Recoger estas lecciones aprendidas permite tener caminos de mejora concretos para la evolución futura del PL y del experimento. En esta sección se presentan la mejoras propuestas relativas al PL. Las mismas se dividen en dos grupos: aquellas que son generales para todo el PL y las que se refieren a secciones particulares.

Una de las mejoras generales se refiere a un mayor vínculo entre el documento que contiene las instrucciones y el material operativo del experimento. En la versión actual del PL estos elementos se encuentran en ficheros distintos. Una solución propuesta para este problema es incluir hiper-vínculos y descripciones breves de cada elemento en el documento principal.

Otras lecciones aprendidas se recogieron para la plantilla de reporte de la replicación. En la misma se observó la falta de algunos campos para describir el diseño experimental y el contexto de replicación. Estas faltas se pueden corregir considerando los campos agregados para el reporte concreto de esta replicación.

Las mejoras propuestas para secciones particulares del PL son diversas. Estas se refieren a elementos que pueden ser agregados o clarificados en el documento. Los cam-

bios propuestos para cada sección están resumidos en la Tabla 5. Como era esperado, el análisis de estos cambios muestra que el PL todavía puede ser mejorado aún más para facilitar el proceso de replicación.

Cuadro 5. Modificaciones propuestas al PL

Sección	Modificación Propuesta
Planificación	Agregar historial de esfuerzo real, pedir al replicador que registre el suyo. Incluir fórmulas para afinar la estimación según el número de programas, tratamientos y sujetos.
Diseño	Explicar más en concreto el experimento actual y las versiones más importantes.
Material Operativo	Indicar formularios, instrucciones y objetos utilizados. Dar un vínculo y breve introducción a cada versión del experimento. Agregar una guía para agregar al PL el nuevo material generado.
Análisis	Mejorar los criterios de corrección, integrar las guías disponibles. Explicar el uso de la plantilla vacía de datos brutos. Los métodos de análisis no son concretos, explicar una aplicación más específica al experimento.
Reporte	Incluir el reporte y empaquetamiento como actividad del proceso de replicación.

Generalmente los materiales del experimento se actualizan, completan y corrigen de una replicación a otra. Sin embargo, por restricciones de cada replicación, no siempre la última versión de los materiales es la más adecuada. Aunque en el estado actual del PL están recuperables las versiones anteriores, no se dispone de un mecanismo formal y automatizado para hacerlo. Por ese motivo, fue sugerida la incorporación de un mecanismo formal de versionado y de trazabilidad de cambios para el PL.

La incorporación de mecanismos de versionado y trazabilidad no es una tarea fácil. Los materiales que contiene un PL son muy heterogéneos y cuál sería la mejor forma de control de cambios está aún en discusión. El uso de un repositorio común facilitó en gran medida la replicación UPM 2011, pero aún así se notaron carencias. Estructurar el PL definiendo un proceso de control de cambios y gestionarlo a través de una herramienta está en discusión actualmente, intentando no generar una estructura más compleja y de difícil acceso.

5. Conclusiones y Trabajos a Futuro

La evaluación de la replicación presentada en este reporte ha sido utilizada como punto de validación para una propuesta de PL. Se replicó un experimento concreto usando un PL instanciado de acuerdo a la propuesta. En general, esta evaluación muestra como alternativa válida la propuesta de PL para apoyar la replicación de experimentos en IS.

Se ha podido comprobar que el PL utilizado es completo con respecto al proceso de investigación experimental. El mismo cubre las actividades de replicación en mayor grado que los PL no estructurados. También se ha comprobado que el documento tiene un grado de usabilidad aceptable desde el punto de vista del replicador responsable.

La evaluación de la replicación muestra que la misma se ha realizado en forma más eficaz que otras anteriores. Esto se ha valorado analizando los incidentes de replicación y dudas surgidas. En cuanto a la eficiencia, la replicación evaluada fue más compleja que

otras anteriores y demandó mayor esfuerzo. Sin embargo, no se han observado efectos negativos por el uso del PL. La evaluación ha contemplado la satisfacción del replicador con el PL y el experimento realizado, en donde también ha resultado positiva en este aspecto.

Además de la valoración de los atributos de evaluación, se recogieron lecciones aprendidas sobre el experimento y el PL. Las lecciones aprendidas permiten establecer un camino de mejora para el instrumento. Las mismas fueron clasificadas como mejoras generales y mejoras aplicables a secciones específicas del PL.

No se espera que la propuesta realizada sea una estructura definitiva para los PL, sino que tenga posibilidad de evolucionar en cada evaluación. El paso natural para esta investigación es impactar estos cambios en el PL y realizar evaluaciones en otros contextos de replicación, con diferencias progresivas al presentado.

Referencias

- Apa, C., Solari, M., Vallespir, D., and Vegas, S. (2011). Construcción de un paquete de laboratorio para un experimento en ingeniería de software. In *XIV Ibero-american Conference on Software Engineering (pendiente de publicación)*.
- Basili, V., Green, S., Laitenberger, O., Lanubile, F., Shull, F., and Sorumgaard, S. (1996). Packaging researcher experience to assist replication of experiments. Technical report, ISERN.
- Basili, V. R., Shull, F., and Lanubile, F. (1999). Building knowledge through families of experiments. *IEEE Transactions on Software Engineering*, 25(4):456–473.
- Campbell, D. T. and Stanley, J. (1963). *Experimental and Quasi-Experimental Designs for Research*. Wadsworth Publishing.
- Juristo, N. and Vegas, S. (2003). *Empirical Methods and Studies in Software Engineering: Experiences from ESERNET*, chapter Functional Testing, Structural Testing and Code Reading: What Fault Type Do They Each Detect? Springer.
- Juristo, N. and Vegas, S. (2009). Using differences among replications of software engineering experiments to gain knowledge. *Empirical Software Engineering and Measurement*, pages 356–366.
- Seaman, C. (1999). Qualitative methods in empirical studies of software engineering. *IEEE Transactions on Software Engineering*, 25(4):557 – 572.
- Shull, F., Mendonça, M. G., Basili, V., Carver, J., Maldonado, J. C., Fabbri, S., Travassos, G. H., and Ferreira, M. C. (2004). Knowledge-sharing issues in experimental software engineering. *Empirical Software Engineering*, 9(1):111.
- Shull, F., Singer, J., and Sjoberg, D. I. K., editors (2008). *Guide to Advanced Empirical Software Engineering*. Springer, London.
- Solari, M. (2010). Propuesta de paquete de laboratorio. Borrador en proceso de validación, Universidad ORT Uruguay.