

# Assessment Method of Empirical Studies in Software Engineering

Dante Carrizo, Jacqueline Manriquez  
Departamento de Ingeniería Informática y Ciencias de la Computación  
Universidad de Atacama UDA  
Copiapó, Chile  
[dante.carrizo@uda.cl](mailto:dante.carrizo@uda.cl), [jacqueline.manriquezb@uda.cl](mailto:jacqueline.manriquezb@uda.cl)

## I. INDICE DE TABLAS

TABLA I.	VALIDEZ DE LA CONCLUSIÓN ESTADÍSTICA. ....	2
TABLA II.	VALIDEZ INTERNA. ....	3
TABLA III.	VALIDEZ DE CONSTRUCTO. ....	4
TABLA IV.	VALIDEZ EXTERNA SIN INTERACCIÓN. ....	4
TABLA V.	VALIDEZ EXTERNA CON INTERACCIÓN.....	4
TABLA VI.	VALIDEZ DE CREDIBILIDAD. ....	5
TABLA VII.	VALIDEZ DE CONFIABILIDAD. ....	5
TABLA VIII.	VALIDEZ DE CONFIRMACIÓN ..... 5	5
TABLA IX.	PAUTA DE LA VALIDEZ DE CONCLUSIÓN ESTADÍSTICA. ....	6
TABLA X.	PAUTA DE VALIDEZ INTERNA ..... 7	7
TABLA XI.	PAUTA DE LA VALIDEZ DE CONSTRUCTO ..... 7	7
TABLA XII.	PAUTA DE LA VALIDEZ EXTERNA SIN INTERACCIÓN ..... 7	7
TABLA XIII.	PAUTA DE LA VALIDEZ EXTERNA SIN INTERACCIÓN ..... 8	8
TABLA XIV.	PAUTA DE LA VALIDEZ DE CREDIBILIDAD..... 8	8
TABLA XV.	PAUTA DE LA VALIDEZ DE CONFIABILIDAD ..... 8	8
TABLA XVI.	PAUTA DE LA VALIDEZ DE CONFIRMACIÓN ..... 8	8

II. PREGUNTAS A FORMULAR PARA LOS CRITERIO DE LAS AMENAZA A LA VALIDEZ CON SUS RESPECTIVOS PPC Y PPV

TABLA I. VALIDEZ DE LA CONCLUSIÓN ESTADÍSTICA.

ID	AMENAZA	CONTROL	PREGUNTA	NOMBRE CRITERIO	PPC	PPV
1.1	Violación de los supuestos del modelo estadístico: Puede sobreestimar o infraestimar el tamaño y la significación del efecto del tratamiento.	Elegir adecuadamente la prueba estadística paramétrica o no paramétrica, dependiendo si los datos cumplen o no los supuestos del modelo paramétrico.	¿Todos los supuestos estadísticos son conocidos y comprobados?	Los supuestos estadísticos son conocidos y comprobados.	12.5	30%
1.2	Baja potencia estadística: Potencia: es la probabilidad de rechazar la hipótesis nula cuando la hipótesis alternativa es la cierta.	Elegir adecuadamente el nivel de significación estadística y el tamaño de la muestra.	¿Cuál es la magnitud de la Covariación, el Error de tipo II aumenta cuando el valor del $\alpha$ es bajo y la muestra es pequeña?	Análisis de rechazar la hipótesis nula cuando la hipótesis alternativa es la cierta	12.5	
1.3	Tasa de Error Tipo I	Con algunas pruebas estadísticas, la probabilidad de cometer error Tipo I se incrementa a medida que aumenta el número de pruebas ejecutadas.	¿Si se ajusta el número de contraste posible se incrementa el Error de tipo I?	Tasa Error de tipo I respecto al número de contraste.	12.5	
1.4	Imprecisión de las medidas	Cuando se utilizan medidas de poca fiabilidad y validez puede aumentar la varianza error, provocando conclusiones erróneas sobre la covariación.	¿Hay más errores que constituyen un problema en los estadísticos?	Errores que provocan imprecisión en las medidas	12.5	
1.5	Escasa fiabilidad en la aplicación de los tratamientos	Si los tratamientos no se aplican de forma homogénea a quienes participan o al mismo participante de una aplicación a otra, puede aumentar la varianza error, provocando conclusiones erróneas sobre la covariación.	¿Se administraron los mismo tratamiento a distintas personas, de un mismo lugar en un tiempo determinado?	Administración de tratamiento respecto a las personas de un mismo lugar en un tiempo determinado	12.5	
1.6	Restricción del rango de las variables	Cuando las variables presentan restricciones en su variación tanto en su rango inferior (efecto suelo), como en su rango superior (efecto techo), la potencia disminuye y la inferencia se debilita.	¿Existe un rango para cuando las variables presentan restricciones?	Las variables presentan restricciones en su variación respecto al rango	12.5	
1.7	Muestra muy heterogénea	Cuanto más heterogénea sea la muestra, mayor será la desviación típica y menor la probabilidad de detectar la posible covariación entre la VI y la VD	¿Las características se correlacionan con las variables dependientes?	Muestra muy heterogénea	12.5	
1.8	Estimación imprecisa de la magnitud del efecto	Algunos estadísticos infraestiman o sobreestiman la magnitud de un efecto.	¿Existe una infraestimación o una sobreestimación de la magnitud del efecto?	Existe una infraestimación o una sobreestimación de la magnitud del efecto	12.5	

TABLA II. VALIDEZ INTERNA.

ID	AMENAZA	CONTROL	PREGUNTA	NOMBRE CRITERIO	PPC	PPV
2.1	Ambigüedad en la contigüidad temporal de la causa: Falta de claridad sobre cuál la causa y cuál el efecto por ambigüedad temporal de las variables objeto de estudio.	Diseños longitudinales y técnicas de análisis de datos (ecuaciones estructurales, análisis de senderos, etc.) Tec. Análisis de Datos Est. Longitudinales	¿Se comprueba repetidamente como actúa el tratamiento y la ausencia del mismo?	Comprobación del actuar del tratamiento y la ausencia del mismo.	12.5	20%
2.2	Historia: Acontecimientos ocurridos durante el ensayo (dentro o fuera) que puedan influir en la VD y confundir los resultados (factores ambientales, vida personal del sujeto, etc.)	Grupo control o técnicas de aleatorización, de constancia o de eliminación (mantener constantes las VE o eliminarlas de todos los grupos) Grupo de control Aleatorización Constancia y eliminación	¿Se realiza una selección aleatoria y una asignación aleatoria?	Se realiza una selección y una asignación aleatoria del grupo de control.	12.5	
2.3	Maduración: Cambios debidos al transcurso del tiempo en los participantes independientes del tratamiento.	Reducir el tiempo del experimento y añadir un grupo control para no tener conclusiones erróneas. Grupo de control Reducir tiempo estudio, Menor tiempo entre el tratamiento y la medición de la VD. Resultados: $GE^1 > GC^2 \Rightarrow$ Eficiente $GE = GC \Rightarrow$ Ineficiente $GE < GC \Rightarrow$ Inaceptable	¿Se realiza una comparación de los resultados al ajustar un experimento con otro en forma aleatoria?	Se realiza una comparación de los resultados al ajustar un experimento con otro en forma aleatoria.	12.5	
2.4	Administración de pruebas: La familiaridad con la prueba distorsiona el resultado de la siguiente. En diseños intersujetos si se utilizan medidas pretest o postest, y en intrasujetos (efecto de orden o error progresivo y efectos residuales o de arrastre)	Grupo de control Contrabalanceo	Intersujeto: ¿Se utilizan medidas de pretest y postest equilibradamente? Intrasujeto: ¿Existe neutralización de los posibles efectos de aprendizaje?	Administración de Prueba: Análisis Intersujeto y análisis Intrasujeto	12.5	
2.5	Instrumentación: Cambios en el tiempo en los instrumentos o procedimientos de medida de la VD.	Entrenamiento de observadores y utilización de instrumentos estandarizados, válidos y fiables. Cuando son errores aleatorios (aumentarían la varianza de error) que se controla utilizando grupos grandes de sujetos Inst. Válidos, sensibles... Entrenamiento observador	¿Existe fiabilidad y consistencia de las medidas de la percepción y decisión de los observadores? (cansancio, aburrimiento, aprendizaje)	Existe fiabilidad y consistencia de las medidas de la percepción y decisión de los observadores (cansancio, aburrimiento, aprendizaje).	12.5	
2.6	Selección diferencial: Relacionada con la formación no aleatoria de grupos, frecuente en los diseños cuasi-experimentales.	Diseño intersujeto: Selección aleatoria, bloques aleatorios, emparejamiento. Diseño intrasujeto: selección aleatoria de la población. Diseños no experimentales: muestras amplias y representativas Asignación aleatoria Bloques aleatorios Emparejamientos	¿Se realiza una selección aleatoria o asignación aleatoria de la muestra?	Selección Diferencial Creación de diseño para intersujeto y intrasujeto al realizar una selección aleatoria o asignación aleatoria de la muestra.	12.5	
2.7	Mortalidad experimental: Pérdida diferencial de participantes por tener un número pequeño de ellos.	Grupos numerosos	¿Existe pérdida diferencial de participantes durante el experimento?	Pérdida diferencial de participantes durante el experimento.	12.5	
2.8	Regresión estadística: Selección de sujetos con puntuaciones previas muy extremas que se acercan con posterioridad a sus valores medios. Falta de fiabilidad de los instrumentos de medida.	Realizar varias medidas antes de seleccionar a los sujetos (pero se podrían familiarizar con el test y aparecer la amenaza "administración de prueba") Instrumentos fiables Toma anterior de medidas	Al omitir las puntuaciones extremas. Volver a medir ¿los valores se acercan a los valores medios?	Realización medición antes y después al sujeto.	12.5	

<sup>1</sup> GE: Grupo Experimental, se aplica el tratamiento o factor testeado.

<sup>2</sup> GC: Grupo de Control, no se aplica el factor testeado.

**TABLA III. VALIDEZ DE CONSTRUCTO.**

ID	AMENAZA	CONTROL	PREGUNTA	NOMBRE CRITERIO	PPC	PPV
3.1	Inadecuada comprensión del constructo teórico.	Desarrollar teoría adecuada al constructo que lo relacione con otros constructos y con otras operaciones. Técnica del simple y doble ciego.	¿Existe escasa definición de los constructos y sus niveles?	Definición de los constructos teóricos (VI) y sus niveles.	20	20%
3.2	Inadecuada medición de los constructos.		¿Existe la medida de solo una variable dependiente un solo método?	Medición de las variables dependiente y métodos.	20	
3.3	Reactividad de los dispositivos instrumentales.		¿Existe un punto de generalización de un estudio a otro?	Generalización.	20	
3.4	Recelo de Evaluación		¿Existe recelo del sujeto a la situación de evaluación?	Recelo de Evaluación.	20	
3.5	Adivinación de la hipótesis		¿Los sujetos intentar adivinar la hipótesis experimental como lo espera el investigador?	Adivinación de la hipótesis.	20	

**TABLA IV. VALIDEZ EXTERNA SIN INTERACCIÓN.**

ID	AMENAZA	CONTROL	PREGUNTA	NOMBRE CRITERIO	PPC	PPV
4.1	Interacción entre selección y tratamiento (validez poblacional - Selección X tratamiento) Carece de validez poblacional cuando los resultados no pueden generalizarse a la población de referencia ni a otras poblaciones debido a que la interacción hallada es específica de los participantes del estudio (no representan a la población).	Acotando las características de la población de referencia y seleccionando aleatoriamente la muestra (si no es posible hay que incluir muchos sujetos heterogéneos en las variables relacionadas y replicar el experimento en muestras diferentes. Aleatorización Muestras heterogéneas Replicación con muestras diferentes	¿Se puede generalizar los efectos y causas a personas que no pertenecen al grupo de estudio?	Interacción entre selección y tratamiento. Generalizar los efectos y causas a personas que no pertenecen al grupo de estudio.	16.7	15%
4.2	Interacción entre situación y tratamiento (validez ecológica - Situación X tratamiento) Disminuye la validez ecológica. Provocada por la artificialidad de las situaciones experimentales. Muy frecuente en el método experimental y poco frecuente en la metodología cualitativa	Replicar la investigación en situaciones diferentes y que la situación experimental sea lo más natural posible. Replicación en ambientes diferentes.	¿Se puede generalizar efecto y causa a situaciones más allá de las estudiadas?	Interacción entre situación y tratamiento. Generalizar efecto y causa a situaciones más allá de las estudiadas.	16.7	
4.3	Interacción entre historia y tratamiento (validez histórica - Historia X tratamiento) Déficit de la validez histórica. Si durante la investigación ocurren acontecimientos que interactúan con el tratamiento, los resultados se vincularán a ese momento temporal y no se podrá generalizar a momentos diferentes.	Replicar el experimento en momentos temporales distintos. Replicación en momentos temporales diferentes	¿Se puede generalizar efecto y causa a otras ocasiones temporales como pasado o futuro?	Interacción entre historia y tratamiento Generalizar efecto y causa a otras ocasiones temporales.	16.7	

**TABLA V. VALIDEZ EXTERNA CON INTERACCIÓN.**

ID	AMENAZA	CONTROL	PREGUNTA	NOMBRE CRITERIO	PPC	PPV
4.4	Interferencia de tratamientos múltiples: Varios tratamientos intrasujetos pueden condicionarse mutuamente.	Contrabalanceo	¿Se puede neutralizar en la medida de lo posible efectos de aprendizaje o arrastre en respuesta a un tratamiento respecto al anterior?	Neutralizar efectos de interferencia de un tratamiento respecto al anterior Intrasujeto.	16.7	5%
4.5	Efecto reactivo de las pruebas: Parecida a la anterior y también intersujetos con medidas pretest.	Igual que la amenaza de administración de pruebas	¿Se puede neutralizar en la medida de lo posible efectos de aprendizaje o arrastre en respuesta a un pretest sensibilizando a los sujetos en un tratamiento respecto al anterior?	Neutralizar efectos de interferencia de un tratamiento respecto al anterior Intersujeto.	16.6	
4.6	Efectos reactivos de los dispositivos experimentales: El sujeto responde lo que espera que responda el observador.	También puede que los sujetos no respondan al tratamiento sino a la novedad de la situación en la que el tratamiento se administra. Técnicas de Ciego y doble ciego	¿Se previene el efecto placebo o el sesgo del observador si conocimientos a que grupo se pertenece?	Efecto placebo.	16.6	

Tabla VI. VALIDEZ DE CREDIBILIDAD.

ID	AMENAZA	CONTROL	PREGUNTA	NOMBRE CRITERIO	PPC	PPV
5.1	Existe documentación que revele interacción durante la investigación.	Revisión de existencia documentada.	¿Se describe la existencia de las notas de campo que surgieron de las acciones y de las interacciones durante la investigación?	Notas de campo.	14.3	5%
5.2	Transcripciones textuales por medio de algún método empírico	Revisión de existencia documentada.	¿Se usó de transcripciones textuales por medio de algún método empírico para respaldar los significados e interpretaciones presentadas en los resultados del estudio?	Método empírico.	14.3	
5.3	Efecto de la presencia del investigador	Revisión de existencia documentada.	¿Se consideró el efecto de la presencia del investigador sobre la naturaleza de los datos?	Efectos de la presencia del investigador.	14.3	
5.4	Discusión de los comportamientos y experiencias	Revisión de existencia documentada.	¿Existe una discusión de los comportamientos y experiencias en relación con la experiencia de los informantes?	Discusión de comportamiento.	14.3	
5.5	Usó de la triangulación en la recolección de datos	Revisión de existencia documentada.	¿Se usó de la triangulación en la recolección de datos para determinar la congruencia entre los resultados	Triangulación de recolección de datos.	14.3	
5.6	Discusión de las interpretaciones con otros investigadores	Revisión de existencia documentada.	¿Existe una discusión de las interpretaciones con otros investigadores?	Discusión de interpretaciones.	14.3	
5.7	Hallazgos significantes y aplicables en el propio contexto	Revisión de existencia documentada.	¿Aparecían hallazgos significantes y aplicables en el propio contexto del lector?	Hallazgos significativos.	14.2	

TABLA VII. VALIDEZ DE CONFIABILIDAD.

ID	AMENAZA	CONTROL	PREGUNTA	NOMBRE CRITERIO	PPC	PPV
6.1	Error constante	Revisión de existencia documentada.	Error constante (sistemático), que se produce cuando las mediciones que se obtienen con una escala son sistemáticamente mayores o menores que lo que realmente deben ser.	Error constante (sistemático)	50	5%
6.2	Error causal	Revisión de existencia documentada.	Error causal (al azar o no sistemático), que se produce cuando las medidas son alternativamente mayores o menores de lo que realmente deben ser.	Error causal (al azar o no sistemático)	50	

TABLA VIII. VALIDEZ DE CONFIRMACIÓN

ID	AMENAZA	CONTROL	PREGUNTA	NOMBRE CRITERIO	PPC	PPV
7.1	Descripción de las características de los informantes y proceso de selección	Revisión de existencia documentada.	¿Existe una descripción de las características de los informantes y su proceso de selección?	Características de los informantes.	33.3	5%
7.2	Mecanismo de grabación de recolección de datos	Revisión de existencia documentada.	¿Existen uso de mecanismos de grabación?	Mecanismo de grabación	33.3	
7.3	Describe los contextos físicos, interpersonales y sociales en el informe	Revisión de existencia documentada.	¿Se describir los contextos físicos, interpersonales y sociales en el informe?	Contextos físicos, interpersonales	33.4	

III. APLICACIÓN DE LA PAUTA DE EVALUACIÓN PROPUESTA A CASO DE ESTUDIO [39]

TABLA IX. PAUTA DE LA VALIDEZ DE CONCLUSIÓN ESTADÍSTICA.

CRITERIO		EVALUACIÓN										
ID: 1.1	Los supuestos estadísticos son conocidos y comprobados.	<b>Elevada</b>										
<b>Justificación:</b> Se eligió satisfactoriamente la prueba estadística no paramétrica, pues se realizó análisis de la varianza (ANOVA), pero cuando ANOVA no es aplicable debido a la falta de normalidad, se utilizó un análisis no paramétrico, la prueba de Kruskal-Wallis. {663,04}												
<b>Table 22</b> Kruskal-Wallis analysis of the relationships between experimental conditions and the underselection of elicitation techniques.												
Experimental conditions (independent variables)		Kruskal-Wallis <i>H</i> value										
		Chi-squared approximation										
		Degrees of freedom <i>d.f.</i>										
		<i>p</i> -Value										
Selection methods	23.956	2										
Case statements	1.229	2										
ID: 1.2	Análisis de rechazar la hipótesis nula cuando la hipótesis alternativa es la cierta.	<b>Elevada</b>										
<b>Justificación:</b> Los valores obtenidos satisfacen el rezado de la Hipótesis Nula. Esto significa que al menos uno de los métodos de selección (propia, aprendidas, framework) conduce a la selección de alguna técnicas. ( $H = 1.229$ , $X^2 = 1.229$ , $d.f. = 2$ , $p\text{-value} = 0.541 > 0.05$ ). {664,01}												
ID: 1.3	Tasa Error de tipo I respecto al número de contraste.	<b>Elevada</b>										
<b>Justificación:</b> El número de sujetos era 11, sobre esa base, los métodos no paramétricos apropiados se aplicaron en algunos casos en los que las condiciones de normalidad y homocedasticidad eran difíciles de cumplir. Se aplicó una corrección de Bonferroni (que ajusta el criterio de significancia de $0,05 / 3 = 0,017$ ) para evitar el problema de error de tipo I. {663,07}												
ID: 1.4	Errores que provocan imprecisión en las medidas.	<b>Elevada</b>										
<b>Justificación:</b> El no presenta errores que constituyan un problema en los estadísticos, pues no aumenta la varianza error y no provoca conclusiones erróneas sobre la covariación. {663,08}												
ID: 1.5	Administración de tratamiento respecto a las personas de un mismo lugar en un tiempo determinado.	<b>Elevada</b>										
<b>Justificación:</b> Los tratamientos se aplicaron tres veces diferentes a los mismos participantes, no se presentó cambio en el resultado de la varianza. {663,07}												
ID: 1.6	Las variables presentan restricciones en su variación respecto al rango.	<b>Pequeña</b>										
<b>Justificación:</b> No se presentan variables con restricciones en su variación respecto al rango, es decir, no se presentan resultados respecto a la potencia ni la inferencia dentro del estudio. {663,08}												
ID: 1.7	Muestra muy heterogénea.	<b>Elevada</b>										
<b>Justificación:</b> Existe una muestra heterogénea donde la desviación típica es más o menos mayor, pues al realizar la prueba de Levene no fue significativa. {666,04}												
<b>Table 28</b> Comparison of the meaHomogeneity of variances test (Levene)n number of concepts, relationships and requirements acquired by subjects during session 2 using the elicitation techniques recommended by the <i>own</i> and <i>framework</i> methods.												
Response variable	Method	Means	Normality test (Kolmogorov-Smirnov)			Homogeneity of variances test (Levene)		Equality of means test (Students's <i>t</i> for equal variances)			Equality of means test (Mann-Whitney <i>U</i> )	
			Value	<i>d.f.</i>	<i>p</i> -Value	Value	<i>p</i> -Value	<i>t</i> Value	<i>d.f.</i>	<i>p</i> -Value	<i>U</i> value	<i>p</i> -Value
Number of concepts	Own	1.60	.367	5	.026	3.2	.105	-	-	-	4.5	.036
	Framework	3.00	.285	6	.138							
Number of relationships	Own	2.20	.231	5	.200	4.63	.060	2.31	9	.046	-	-
	Framework	4.50	.265	6	.200							
Number of requirements	Own	6.20	.221	5	.200	2.1	.184	2.50	9	.034	-	-
	Framework	9.83	.189	6	.200							
ID: 1.8	Existe una infraestimación o una sobreestimación de la magnitud del efecto.	<b>Moderada</b>										
<b>Justificación:</b> Dentro del estudio se realizó la aplicación de estadísticos pero no se especifica algún efecto parcial en la magnitud de un efecto. {666,03}												

RUBRICA DE CONCLUSION ESTADISTICA

TABLA X. PAUTA DE VALIDEZ INTERNA

CRITERIO		EVALUACIÓN
ID:2.1	Comprobación del actuar del tratamiento y la ausencia del mismo.	<b>Moderada</b>
<b>Justificación:</b> Dentro del estudio no se especifica un análisis longitudinal, pero se expresa que los sujetos aplicaron los métodos en orden, empezando por su propio método y terminando por el framework propuesto, a horas fijas durante un periodo semestral sin ellos tener el conocimiento del experimento, y se hizo de esa forma para evitar inferencia con tratamientos múltiples. {667,03}		
ID:2.2	Se realiza una selección y una asignación aleatoria del grupo de control.	<b>Elevada</b>
<b>Justificación:</b> Dentro del estudio se especifica la creación de grupos de control elegidos al azar. {667,02}		
ID:2.3	Se realiza una comparación de los resultados al ajustar un experimento con otro en forma aleatoria.	<b>Pequeña</b>
<b>Justificación:</b> Dentro del estudio no se especifica el análisis de resultados entre los grupos experimentales y los grupos de control para ver si: {660,03} GE>GC => Eficiente GE=GC => Ineficiente GE<GC => Inaceptable		
ID:2.4	Administración de Prueba: Análisis Intersujeto y análisis Intrasujeto	<b>Pequeña</b>
<b>Justificación:</b> Análisis Intersujeto: No se especifica la aplicación de contrabalanceo para medición de pretest y postest. Pero se informa la aplicación de 3 test de 30 minutos. {667,02} Análisis Intrasujeto: No se especifica la aplicación de contrabalanceo para la medición de la neutralización. Pero si se informa que se realizó una encuesta a los sujetos donde se especificó que si no hubo efectos de orden o error progresivo y efectos residuales. {667,02}		
ID:2.5	Existe fiabilidad y consistencia de las medidas de la percepción y decisión de los observadores (cansancio, aburrimiento, aprendizaje).	<b>Elevada</b>
<b>Justificación:</b> Se especifican diferentes medidas para que existiera fiabilidad y consistencia, ejemplo 30 min para evitar la fatiga, se aplicaron diferentes situaciones contextuales para evitar el aprendizaje. {667,02}		
ID:2.6	Selección Diferencial Creación de diseño para intersujeto y intrasujeto al realizar una selección aleatoria o asignación aleatoria de la muestra.	<b>Elevada</b>
<b>Justificación:</b> Diseño Intersujeto: Se especifica la selección aleatoria de los grupos de emparejamiento. {660,03} Diseño Intrasujeto: Se especifica que se realizó una asignación aleatoria de la población de estudiantes de Master de Ingeniería de Software con experiencia. {665,01}		
ID:2.7	Perdida diferencial de participantes durante el experimento. {663,03}{664,02}{666,12}	<b>Moderada</b>
<b>Justificación:</b> Se menciona que el grupo de control es pequeño pero no existe la pérdida de participantes. {666,12}		
ID:2.8	Realización medición antes y después al sujeto.	<b>Moderada</b>
<b>Justificación:</b> Se realizó una encuesta al finalizar el experimento en la cual los sujetos encontraron la solución propuesta: {666,11} <ul style="list-style-type: none"> <li>• Fácil Uso (100%)</li> <li>• Aplicable (70%)</li> <li>• Útil (60%)</li> <li>• Correcta (60%)</li> </ul>		

RUBRICA DE VALIDEZ INTERNA

TABLA XI. PAUTA DE LA VALIDEZ DE CONSTRUCTO

CRITERIO		EVALUACIÓN
ID:3.1	Definición de los constructos teóricos (VI) y sus niveles.	<b>Elevada</b>
<b>Justificación:</b> Se especifica la definición teórica de la aplicación de los tratamientos y la representación con precisión abstracta, informando la utilización de diferentes variables de respuestas para confirmar la relación causa-efecto, para el experimento 2. {665,02}		
ID:3.2	Medición de las variables dependiente y métodos.	<b>Elevada</b>
<b>Justificación:</b> Se especifica la utilización de la técnica de simple ciego para los informantes. {667,04}		
ID:3.3	Generalización.	<b>Elevada</b>
<b>Justificación:</b> Se aplica técnica de simple ciego, pero la explicación deja entre ver que se puede realizar generalización del instrumento en el área de las técnicas y expertos. {667,04}		
ID:3.4	Recelo de Evaluación.	<b>Elevada</b>
<b>Justificación:</b> En ninguna instancia se menciona que exista una desconfianza del sujeto a la situación de evaluación. {667,05}		
ID:3.5	Adivinación de la hipótesis.	<b>Elevada</b>
<b>Justificación:</b> Se aclara que los resultados experimentales pueden ser asumidos como válidos pues para los desarrolladores novatos la utilización de la técnica de entrevista abierta es más factible, mientras que aquellos con ligera experiencia utilizaran técnicas más avanzadas. {667,06}		

RUBRICA DE VALIDEZ DE CONSTRUCTO

TABLA XII. PAUTA DE LA VALIDEZ EXTERNA SIN INTERACCIÓN

CRITERIO		EVALUACIÓN
ID:4.1	Interacción entre selección y tratamiento. Generalizar los efectos y causas a personas que no pertenecen al grupo de estudio.	<b>Moderada</b>
<b>Justificación:</b> Se aplica la técnica de simple ciego, pero la explicación deja entre ver que se necesita más estudios para poder inferir una generalización poblacional de la misma área. {667,04}		
ID:4.2	Interacción entre situación y tratamiento. Generalizar efecto y causa a situaciones más allá de las estudiadas.	<b>Pequeña</b>
<b>Justificación:</b> No se explica que este estudio pueda ser utilizado en otros ambientes. {667,13}		
ID:4.3	Interacción entre historia y tratamiento Generalizar efecto y causa a otras ocasiones temporales.	<b>Moderada</b>
<b>Justificación:</b> Se menciona la diferencia entre desarrolladores novatos y expertos, pero se afirma que este framework sería aplicables por ingenieros de requisitos con más experiencia si se realizan más estudios. {667,06}		

RUBRICA DE VALIDEZ EXTERNA SIN INTERACCIÓN

Tabla XIII. Pauta de la Validez Externa Sin Interacción

RUBRICA DE VALIDEZ EXTERNA SIN INTERACCIÓN	CRITERIO		EVALUACIÓN
	ID:4.4	Neutralizar efectos de interferencia de un tratamiento respecto al anterior Intrasujeto.	
	<b>Justificación:</b> No se usa técnica de contrabalanceo para que los sujetos no realicen un aprendizaje y se provoque una inferencia de efectos de los siguientes. {667,02}		
ID:4.5	Neutralizar efectos de interferencia de un tratamiento respecto al anterior Intersujeto.		<b>Pequeña</b>
	<b>Justificación:</b> No se usa la técnica de contrabalanceo para que los sujetos no realicen un aprendizaje o arrastre en respuesta a un pretest. {667,02}		
ID:4.6	Efecto placebo.		<b>Moderada</b>
	<b>Justificación:</b> Se realiza técnica de simple ciego en el experimento 2. Pero no se informa que los sujetos experimenten una influencia del observador. {667,02}		

TABLA XIV. PAUTA DE LA VALIDEZ DE CREDIBILIDAD

PAUTA DE LA VALIDEZ DE CREDIBILIDAD	CRITERIO		EVALUACIÓN
	ID:5.1	Notas de campo.	
	<b>Justificación:</b> Se describen parcialmente notas relacionadas con las acciones y de interacción, en especial enunciando que los sujetos no tenían conocimiento del experimento. {667,02}		
ID:5.2	Método empírico.		<b>Elevada</b>
	<b>Justificación:</b> En todo momento se expresa la realización del experimento para validar el estudio. {667,02}		
ID:5.3	Efectos de la presencia del investigador.		<b>Elevada</b>
	<b>Justificación:</b> Se enuncia en 2 oportunidades que el investigador para mejorar los resultados estadísticos realizó alguna acción. {667,02}		
ID:5.4	Discusión de comportamiento.		<b>Elevada</b>
	<b>Justificación:</b> En todo momento se explica el comportamiento de las experiencias de los sujetos, bajo las reglas establecidas en la validez interna. {667,06}		
ID:5.5	Triangulación de recolección de datos.		<b>Moderada</b>
	<b>Justificación:</b> No se utiliza triangulación literalmente pero si se menciona los resultados esperados de ella. {667,06}		
ID:5.6	Discusión de interpretaciones.		<b>Elevada</b>
	<b>Justificación:</b> Se realiza en diversas áreas una discusión exhaustiva con buena base en la interpretación de resultados con otros investigadores y expertos. {667, 07} {667,12}		
ID:5.7	Hallazgos significativos.		<b>Pequeña</b>
	<b>Justificación:</b> No se mencionan hallazgos significativos novedosos que hayan nacido paralelamente a la investigación sin haberse tenido en cuenta. {667,13}		

TABLA XV. PAUTA DE LA VALIDEZ DE CONFIABILIDAD

RUBRICA DE VALIDEZ DE CONFIABILIDAD	CRITERIO		EVALUACIÓN
	ID:6.1	Error constante (sistemático).	
	<b>Justificación:</b> En la mayoría de los resultados expresados están bajo tipos estadísticos. {667, 04}		
ID:6.2	Error causal (al azar o no sistemático)		<b>Moderada</b>
	<b>Justificación:</b> Al utilizar un método empírico se obtienen resultados no sistemáticos, pero no se encuentran contrastados con valores reales. {667,12}		

TABLA XVI. PAUTA DE LA VALIDEZ DE CONFIRMACIÓN

RUBRICA DE VALIDEZ DE CONFIRMACIÓN	CRITERIO		EVALUACIÓN
	ID:7.1	Características de los informantes.	
	<b>Justificación:</b> Se especifica detalladamente en número de sujetos, su nivel de experiencia, su forma de elección y la descripción de sus grupos de control. {666,12}		
ID:7.2	Mecanismo de grabación.		<b>Pequeña</b>
	<b>Justificación:</b> No se explica la existencia de una documentación grabada de la realización de los experimentos. {667,02}		
ID:7.3	Contextos físicos, interpersonales.		<b>Elevada</b>
	<b>Justificación:</b> Se menciona específicamente hacia que área apunta dicha investigación. {667,08}		