

# Estudio experimental de la efectividad de la entrevista abierta frente a la entrevista independiente de contexto

Dante Carrizo<sup>1</sup>, Oscar Dieste<sup>2\*</sup>, Natalia Juristo<sup>2</sup>, Marta López<sup>3</sup>

<sup>1</sup> Universidad de Atacama, Dpto. Ingeniería Informática y Ciencias de la Computación, Avda. Copayapu 485, Copiapó, Chile

<sup>2</sup> Universidad Politécnica de Madrid, Facultad de Informática, Campus de Montegancedo, 28660 Boadilla del Monte, Spain  
{odieste, natalia}@fi.upm.es

<sup>3</sup> Universidad Complutense de Madrid, Facultad de Informática, C/ Prof. José García Santesmases, s/n. 28040 Madrid, Spain

**Resumen. OBJETIVO:** Estudiar la efectividad relativa de las técnicas de entrevista abierta y entrevista independiente de contexto en la identificación de requisitos para sistemas software. **MÉTODO:** Experimento controlado ensayando dos factores: “tipo de entrevista” (abierta e independiente de contexto) y “tipo de problema” (identificación de requisitos para dos sistemas software distintos). Debido al reducido número de sujetos experimentales, se ha utilizado un diseño de medidas repetidas dentro de los sujetos. **RESULTADOS:** El poder estadístico del estudio es en general insuficiente, pero hemos podido constatar que el tipo de problema posee efectos significativos en la efectividad del proceso de relevamiento de requisitos. La experiencia del analista juega también un papel importante. El tipo de entrevista utilizada afecta en mucha menor medida. **CONCLUSIONES:** Es necesario comprender mejor qué es un problema de relevamiento de requisitos. Esto nos permitirá diseñar estrategias más eficaces para una identificación certera de los requisitos del software.

**Palabras clave:** Entrevista abierta, entrevista independiente de contexto, tipo de problema, experiencia del analista, relevamiento de requisitos, efectividad.

## 1 Introducción

El proceso de requisitos, independientemente del modelo teórico considerado (e.g.: [1-3]), tiene como primera fase el relevamiento de requisitos. Existen varias técnicas para realizar dicho relevamiento, tales como el análisis de protocolos [4], emparejamiento [5], grupos de trabajo [6], etc. No obstante, en la práctica la técnica de educación de requisitos más frecuentemente utilizada son las entrevistas [7].

Existen varios tipos de entrevistas. Los tipos más usuales son la “abierta” o “no estructurada”, en la que se trabaja con preguntas sin prefijar de antemano, y la “cerrada” o “estructurada”, basada en una serie de preguntas predeterminadas y en un

---

\* Enviar correspondencia a este autor.

orden concreto [8]. Ambos tipos son el *continuum* de un conjunto, situándose en algún lugar intermedio la entrevista independiente de contexto, caracterizada por el uso de preguntas abiertas y genéricas. Este tipo de entrevista fue inicialmente propuesta por [9] y es recomendada por varios autores [10,11] para la realización de las primeras sesiones de relevamiento.

Los distintos tipos de entrevistas han sido comparadas experimentalmente para estudiar su efectividad [12]. Los estudios realizados apuntan a la existencia de dos efectos bastante bien definidos. Por una parte, las entrevistas estructuradas (o parcialmente estructuradas, tales como las entrevistas independientes de contexto) parecen superiores que las entrevistas abiertas en términos de la información obtenida [7,13,14]. Por otra parte, esta superioridad se manifiesta en distintos dominios de forma razonablemente consistente; esto es, no parece que el tipo de problema influya en la efectividad de las entrevistas, aunque este particular no ha sido contrastado empíricamente sino mediante una revisión sistemática [12].

Algunos estudios también señalan que la experiencia de los analistas influye de modo relativamente reducido en la eficacia del proceso de relevamiento; esto es, los analistas experimentados no parece que obtengan más información ni sean más eficientes que los analistas noveles cuando utilizan entrevistas estructuradas o parcialmente estructuradas [7,13].

El objetivo de este trabajo es estudiar si dichos efectos ocurren realmente. Ello es de la mayor relevancia para el desarrollo de software, ya que cuanto más completo sea el resultado del proceso de requisitos, mayor calidad poseerá el software resultante.

Para realizar dicho estudio, hemos planificado y ejecutado un estudio experimental [15]. La experimentación ha sido ampliamente utilizada en ámbitos como la Medicina u otras Ingenierías como una forma legítima de obtener información y favorecer el avance de las áreas de conocimiento. En la Ingeniería del Software se comienza a utilizar, si bien su aplicación es todavía bastante reducida; aún así se han realizado un buen número de experimentos en esta área [16]. Mediante el experimento se construye una realidad simplificada que permite mantener bajo control diversas variables (o aspectos del problema bajo estudio) y analizar sus efectos. En nuestro caso, hemos planteado un experimento controlado aplicando dos tipos de entrevista (abierta e independiente de contexto) a dos problemas de relevamiento de requisitos distintos (que se describirán inmediatamente a continuación). La influencia de la experiencia en el relevamiento de requisitos no ha podido comprobarse experimentalmente, ya que los sujetos experimentales de los que disponíamos eran una muestra de conveniencia y el factor experiencia no podía ni aleatorizarse ni bloquearse.

No obstante, sí hemos podido estudiar la relación entre la experiencia de los sujetos y los datos de eficacia mediante correlación estadística entre los datos experimentales y observacionales. Para la presentación de este trabajo se ha estructurado el artículo de la siguiente forma: en la sección 2 se detallan las hipótesis de experimentación; en la sección 3 se describe el experimento, cuyo análisis y resultados se presentan en la sección 4; en la sección 5 se listan las amenazas a la validez de este trabajo y, finalmente, se muestran las conclusiones en la sección 6.

## 2 Hipótesis de trabajo

Las hipótesis de trabajo del presente experimento son las siguientes:

$H_{a0}$ : No existen diferencias en efectividad entre la entrevista abierta y la entrevista independiente de contexto.

$H_{a1}$ : La entrevista independiente de contexto posee una mayor efectividad que la entrevista abierta

$H_{b0}$ : El tipo de problema no afecta a la efectividad de las entrevistas.

$H_{b1}$ : El tipo de problema afecta a la efectividad de las entrevistas.

## 3 Descripción del experimento

### 3.1 Factores

El experimento posee dos factores: el tipo de entrevista y el problema a estudiar, en consonancia con las hipótesis de trabajo planteadas.

El factor “tipo de entrevista” posee dos niveles: “Entrevista abierta” y “Entrevista independiente de contexto”. Por “Entrevista abierta”, entendemos aquella entrevista cuyas preguntas no se han prefijado de antemano, sino que se realizan a medida que transcurre el relevamiento a voluntad del analista. Hemos asegurado que aquellos sujetos que han realizado la entrevista abierta no pudieran prepararla de antemano, manteniendo en secreto el problema a estudiar hasta la misma realización de la entrevista.

Por “Entrevista independiente de contexto”, entendemos una entrevista preparada utilizando un conjunto de preguntas predefinidas de tipo genérico, esto es, aplicable a distintos tipos de problemas. Dado que las preguntas son de tipo genérico, no existe una única plantilla, sino que depende del autor. En este experimento, hemos utilizado una plantilla basada en [10] y que se incluye en el anexo A.

El factor “problema a estudiar” posee asimismo dos niveles. El primero, al que nos referiremos con el nombre de “Comisiones”, trata de la identificación de los requisitos para un sistema de gestión de las comisiones de un hipotético departamento universitario. El segundo, al que denominaremos “Reciclado”, versa sobre un sistema de control de máquinas de reciclado de pilas. Ambos problemas han sido definidos en detalle, incluyendo: objetivos, requisitos, procesos y funciones, conceptos, atributos, relaciones y restricciones.

Existen dos motivos que justifican los niveles utilizados. En primer lugar, no deseábamos que los problemas fueran bien conocidos para los sujetos, de modo que los requisitos fueran inferidos (lo que ocurriría, por ejemplo, con un sistema de información clásico) en lugar de extraídos. Por ese motivo, el dominio de ambos problemas es poco habitual. En segundo lugar, deseábamos que los tipos de sistemas a desarrollar fueran muy diferentes para que, en caso de que posean un efecto sobre las entrevistas, dichos efectos fueran fácilmente visibles incluso con pocos sujetos experimentales.

### 3.2 Variables respuesta

Hemos estudiado varios aspectos de la información extraída, como por ejemplo el número de procesos, conceptos, etc. Sin embargo, el análisis no está concluido todavía, por lo que en este artículo nos centraremos únicamente en una variable respuesta: el número de requisitos identificados por los sujetos experimentales.

Hemos medido la variable respuesta a partir de los reportes realizados por los sujetos experimentales. Una semana después de la sesión de relevamiento, los sujetos experimentales entregaron una lista conteniendo los requisitos del software. Dichos requisitos se compararon con una lista patrón predefinida para cada problema (disponible también bajo petición), lo cual nos permitió determinar el número de requisitos identificados por cada sujeto.

La métrica de la variable respuesta es el porcentaje de requisitos identificados por un sujeto. Utilizamos el porcentaje ya que los problemas a estudiar poseían un número distinto de requisitos (14 en un caso y 18 en el otro), lo que implica que la comparación de los valores absolutos no posee sentido alguno.

### 3.3 Sujetos experimentales

Los sujetos experimentales fueron 13 alumnos del Máster de Ingeniería del Software de la Facultad de Informática de la Universidad Politécnica de Madrid. Se trata, por lo tanto, de una muestra de conveniencia, lo que explica que no realizaremos una estimación *ex ante* del poder estadístico y cálculo del tamaño muestral.

Todos los alumnos fueron titulados superiores en informática procedentes de varios países latinoamericanos (Argentina, Brasil, Colombia, Ecuador, España y Perú) con experiencia en desarrollo de software y, en menor medida, también en análisis de requisitos. La Tabla 1 muestra los datos descriptivos relativos a la experiencia de los 11 alumnos que aportaron esta información. Los sujetos identificados como C y E en la Tabla del anexo B no respondieron el cuestionario demográfico desarrollado para recopilar este tipo de datos.

La *saliencia* del experimento, esto es, el interés de los sujetos en ejecutar en forma debida la tarea experimental, estaba asegurada ya que el experimento era una de las prácticas evaluables del curso de Requisitos que dichos estudiantes cursaban en aquel momento en el Máster de Ingeniería del Software. En esta práctica debían obtener la lista de requisitos funcionales y no funcionales de los sistemas analizados.

**Tabla 1.** Datos descriptivos de los sujetos experimentales.

Estadísticos descriptivos					
	N	Mínimo	Máximo	Media	Desv. típ.
Experiencia total en años	11	3	16	7,55	3,588
Experiencia como analista en años	11	0	8	3,00	2,449

### 3.4 Diseño

El diseño experimental utilizado ha sido un factorial 2 x 2. Sin embargo, debido al bajo número de sujetos experimentales, hemos decidido realizar medidas repetidas

para aumentar la potencia estadística. En concreto, cada sujeto realiza dos sesiones de relevamiento, uno para cada problema, utilizando el mismo tipo de entrevista.

Un diseño más efectivo a priori para evitar efectos de *carry-over* (por ejemplo, aprendizaje), consistiría en alternar no sólo el problema sino también el tipo de entrevista dentro de los sujetos. Sin embargo, no hemos utilizado esta estrategia debido a que los tipos de entrevistas no son tan distintos como para que los efectos de aprendizaje no existan y, en contrapartida, los sujetos podrían contaminar la segunda entrevista, usando estrategias aprendidas en la 1ª entrevista durante la 2ª entrevista. Dado que la existencia de efectos de *carry-over* no puede descartarse, la sesión de relevamiento (1ª ó 2ª) se introducirá como factor durante el análisis de los resultados.

### **3.5 Preparación del experimento**

A cada sujeto se le asignó el tipo de entrevista a aplicar (abierta o independiente de contexto) mediante insaculación y, por lo tanto, de modo aleatorio. La asignación fue ciega para los experimentadores hasta el análisis de los datos. En concreto, la medición del número de requisitos identificados por cada sujeto se hizo de forma ciega respecto al tipo de entrevista.

Dos de los experimentadores desempeñamos el papel de clientes, uno para cada problema. Para ello, estudiamos en profundidad los casos hasta que pudimos responder de forma natural a preguntas sobre los mismos.

Cada sesión de relevamiento estaba limitada a 30 minutos. Los 5 primeros minutos de la sesión los dedicamos a explicar a los sujetos los objetivos de la sesión y el tipo de problema. A continuación y por un tiempo máximo de 20 minutos, se realizó la entrevista propiamente dicha. Durante los últimos 5 minutos, los experimentadores en el rol de clientes descansamos y nos preparamos para la siguiente entrevista.

Las sesiones de relevamiento de requisitos se planificaron en 3 días distintos. El relevamiento acerca del problema de reciclado se realizó los días 1 y 2, mientras que el relevamiento acerca del problema de las comisiones se realizó los días 2 y 3. Esta configuración se consideró la más adecuada ya que generaba un carga de trabajo asumible tanto para los experimentadores como para los sujetos.

La asignación de sujetos a sesiones no pudo aleatorizarse debido a las restricciones de índole laboral de los sujetos experimentales y, por tanto, no se pudo asegurar por diseño un completo balance entre grupos. No obstante, no hay ninguna razón para suponer que la falta de aleatorización haya producido sesgo alguno y, finalmente, los grupos experimentales mostraron un carácter balanceado al finalizar el experimento.

### **3.6 Ejecución**

La ejecución del experimento se realizó en gran medida de acuerdo al plan previsto. No obstante, debido a que algunos sujetos no respetaron la planificación se tuvo que reprogramar algunas sesiones. Esto provocó que la carga de trabajo del experimentador que representó el papel de cliente del sistema de las comisiones fuese mayor en el día 3, mientras que la carga de trabajo del experimentador del sistema de reciclado fue mayor en el día 1. Esta distribución desigual del cansancio de los experimentadores ha quedado reflejada en las gráficas analíticas que se muestran en la sección 4.

Todas las entrevistas fueron grabadas para realizar posteriores análisis de contenidos. Cada sujeto recibió su grabación. Ello no implica riesgo alguno, ya que el experimento tenía como objetivo analizar la efectividad de las técnicas y no los problemas de comprensión o retención de información por parte del analista. Es más, dado que los sujetos no reportaron sus hallazgos hasta días más tarde de la realización del experimento, las grabaciones probablemente permitieron una identificación más certera de los requisitos.

Los experimentadores en el rol de clientes intentaron contestar lo más completa y certeramente posible a las preguntas planteadas por los sujetos experimentales, sin retener información de ninguna clase.

Al finalizar el experimento, los sujetos respondieron a un cuestionario demográfico que nos permitió obtener información acerca de su experiencia en desarrollo de software y análisis de requisitos, así como sus opiniones acerca de los problemas analizados. Los resultados obtenidos tras la aplicación del cuestionario nos permitieron estudiar la posible influencia de las variables contextuales, especialmente la experiencia de los sujetos, en los resultados experimentales.

## 4 Resultados

Los resultados del experimento se muestran en el anexo B. El análisis estadístico fue realizado por uno de los experimentadores utilizando SPSS v. 18.

El anexo B muestra que existe un *drop-out* apreciable, aunque no importante. De las 26 sesiones planificadas (13 sujetos x 2 sesiones cada uno), 25 llegaron a realizarse, pero sólo 20 de los 26 reportes de requisitos esperados fueron entregados (76%). No existe una razón especial que justifique el *drop-out* mas allá de la motivación que, como alumnos, poseían los sujetos.

La Tabla 2 muestra el análisis de varianza de todos los factores, incluyendo la sesión (1ª o 2ª) en la que cada sujeto realizó el relevamiento de requisitos. Ninguno de los factores posee un efecto significativo, por lo cual podríamos concluir (anticipamos que equivocadamente) que:

- No podemos rechazar la hipótesis  $H_{a_0}$ . Por lo tanto, las dos técnicas de entrevista ensayadas (Abierta e Independiente de contexto) poseen una eficacia similar.
- No podemos rechazar la hipótesis  $H_{b_0}$ . Por lo tanto, el tipo de problema no afecta a la efectividad de las entrevistas.

Adicionalmente, la sesión en la que cada sujeto realizó el relevamiento de requisitos (1ª o 2ª) no afecta a los resultados.

No obstante, las conclusiones anteriores son al menos discutibles, ya que la potencia de los test estadísticos utilizados es muy baja. La mayor potencia ocurre en el test respecto al factor Problema y sólo alcanza el 50.7%, lejos de la recomendación habitual del 80%. En los restantes factores e interacciones, la potencia es netamente menor. En consecuencia, era previsible que ninguno de los factores alcanzara la significación estadística. De hecho, esta baja potencia fue la que motivó que se utilizara un diseño de medidas repetidas con el fin de aumentar el número de sujetos

experimentales por tratamiento y, con ello, el poder global, tal y como se ha indicado en la sección 3.

Por lo tanto, en lugar de estudiar si un factor es o no significativo al nivel  $\alpha=0.05$ , creemos preferible estudiar las tendencias e intentar obtener alguna conclusión, aunque sea contingente, de las mismas.

**Tabla 2.** Análisis de varianza incluyendo todos los factores.

Pruebas de los efectos inter-sujetos								
Variable dependiente: Porcentaje requisitos								
Origen	Suma de cuadrados tipo III	gl	Media cuadrática	F	Sig.	Eta al cuadrado parcial	Parámetro de no centralidad	Potencia observada <sup>b</sup>
Modelo corregido	5995,217 <sup>a</sup>	7	856,460	1,425	,281	,454	9,975	,379
Intersección	38592,086	1	38592,086	64,208	,000	,843	64,208	1,000
Tecnica	12,572	1	12,572	,021	,887	,002	,021	,052
Sesion	658,752	1	658,752	1,096	,316	,084	1,096	,162
Problema	2780,572	1	2780,572	4,626	,053	,278	4,626	,507
Tecnica *	244,826	1	244,826	,407	,535	,033	,407	,091
Sesion								
Tecnica *	130,371	1	130,371	,217	,650	,018	,217	,071
Problema								
Sesion *	476,446	1	476,446	,793	,391	,062	,793	,130
Problema								
Tecnica *	354,752	1	354,752	,590	,457	,047	,590	,109
Sesion *								
Problema								
Error	7212,533	12	601,044					
Total	62709,000	20						
Total corregida	13207,750	19						

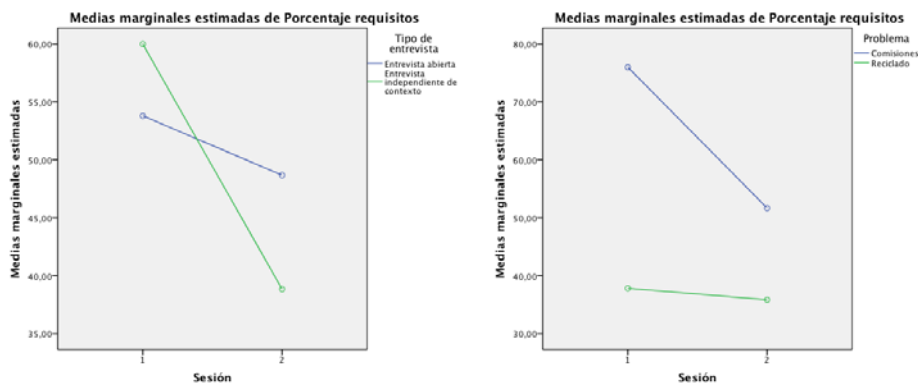
a. R cuadrado = ,454 (R cuadrado corregida = ,135)

b. Calculado con alfa = ,05

El primer aspecto que creemos relevante es el papel desempeñado por la variable “sesión”. Esta variable representa el hecho de que se realizan dos sesiones de educación, en dos momentos distintos, por los mismos sujetos (aunque referidos a distintos problemas), lo que implica la posible existencia de efectos de *carry-over*. La ANOVA no ha identificado tal efecto, pero un examen de las medias marginales apunta a su existencia. Tal y como se observa en la Fig 1, la efectividad disminuye tanto respecto al factor “tipo de entrevista” como al factor “problema a estudiar” entre la 1ª y la 2ª sesión.

Este efecto de *carry-over* es contrario al señalado por el análisis teórico, el cual apunta a un incremento de la efectividad motivado por un efecto de aprendizaje. No obstante, creemos que una explicación razonable es que la disminución en efectividad viene motivada por el cansancio de los experimentadores que desempeñaban el papel de clientes. Nótese que la ejecución del experimento implicó la realización de 13 entrevistas, cada una de ½ hora de duración. Obviamente, a medida que progresaba la realización de entrevistas, el cansancio de los experimentadores aumentó y eso se reflejó en una menor cantidad de información suministrada a los sujetos experimentales.

Como prueba adicional de que el cansancio puede explicar el efecto de *carry-over*, obsérvese la disminución relativa de las medias marginales referidas al “problema a estudiar” (mostradas en el 2º gráfico de la Fig. 1). La caída es mucho más acusada en el problema de las comisiones que en el problema del reciclado. Esto no hace más que reflejar el mayor cansancio del experimentador que asumía el papel de cliente del sistema de las comisiones respecto al otro experimentador, a cargo del problema del reciclado. En este caso, el cansancio estuvo provocado tanto por la mayor carga de trabajo en la segunda sesión de este experimentador, como por la menor experiencia en *role-playing*.



**Fig. 1.** Influencia del factor sesión en el número de requisitos obtenidos

La existencia de un efecto de *carry-over* complica sobremanera el análisis de los resultados experimentales. Es, por lo tanto, preferible descartar la 2ª sesión y re-analizar utilizando los datos de la 1ª sesión únicamente [17]. Naturalmente, esto elimina la necesidad de una variable “sesión” y, aunque se pierde tamaño muestral, la ausencia de interacciones pueden compensar la disminución de poder estadístico. Dicho análisis de varianza restringido a la 1ª sesión se muestra en la Tabla 3.

Los resultados son bastante similares a los obtenidos anteriormente (ver Tabla 2), con la diferencia de que el factor problema es ahora significativo al nivel  $\alpha=0.05$ . Por lo tanto, no podemos mantener  $H_{b0}$  y debemos afirmar  $H_{b1}$ : el tipo de problema afecta a la efectividad del proceso de relevamiento de requisitos.

El factor “tipo de entrevista” está muy lejos de alcanzar la significación estadística, por lo que  $H_{a0}$  se mantiene. Sin embargo, no puede negarse que la entrevista independiente de contexto ejerce una influencia positiva en el proceso de relevamiento de requisitos. Tal y como muestra la Fig. 2, la efectividad media de la entrevista independiente de contexto es mayor que la entrevista abierta para ambos tipos de problemas. La falta de significación estadística viene motivada por la gran dispersión de la efectividad de la entrevista independiente de contexto entre sujetos, así como al bajo tamaño muestral. No obstante, independientemente de la significación estadística, el tamaño de efecto asociado al tipo de entrevista es muy reducido ( $\eta^2 = 0.022$ ) por lo que su influencia sobre el proceso de relevamiento de requisitos es, en todo caso, baja.



**Tabla 3.** Análisis de varianza restringida a la 1ª sesión.

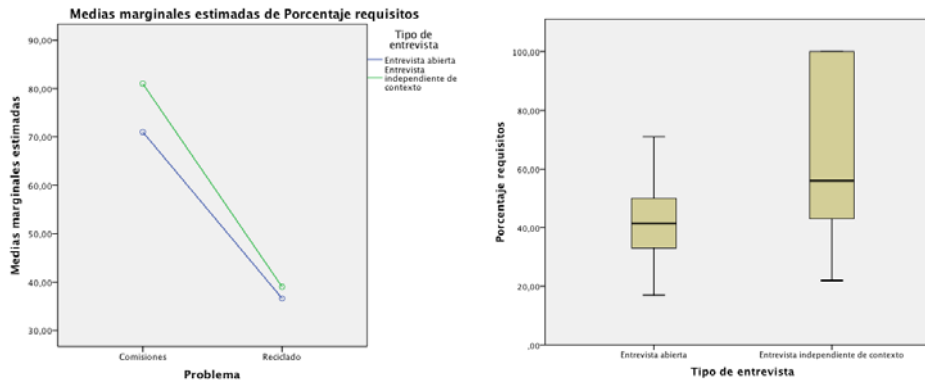
**Pruebas de los efectos inter-sujetos**

Variable dependiente: Porcentaje requisitos

Origen	Suma de cuadrados tipo III	gl	Media cuadrática	F	Sig.	Eta al cuadrado parcial	Parámetro de no centralidad	Potencia observada <sup>b</sup>
Modelo corregido	4406,982 <sup>a</sup>	3	1468,994	3,041	,102	,566	9,124	,465
Intersección Técnica	25476,275	1	25476,275	52,743	,000	,883	52,743	1,000
Problema Técnica *	75,620	1	75,620	,157	,704	,022	,157	,064
Problema	2870,636	1	2870,636	5,943	,045	,459	5,943	,555
Problema	28,407	1	28,407	,059	,815	,008	,059	,055
Error	3381,200	7	483,029					
Total	37845,000	11						
Total corregida	7788,182	10						

a. R cuadrado = ,566 (R cuadrado corregida = ,380)

b. Calculado con alfa = ,05



**Fig. 2.** Efecto del tipo de entrevista

Finalmente, hemos calculado las correlaciones entre algunos datos demográficos y la efectividad de los sujetos, tal y como se indica en la Tabla 4.

**Tabla 4.** Correlación entre los datos demográficos y la efectividad de los sujetos.

**Coefficientes<sup>a</sup>**

Modelo	Coefficients no estandarizados		Coefficients tipificados	t	Sig.
	B	Error típ.	Beta		
1 (Constante)	86,215	51,729		1,667	,171
Experiencia total en años	-,921	4,336	-,078	-,212	,842
Experiencia como analista en años	7,976	2,919	,814	2,732	,052
Dummy_Conoc_Problema	-13,435	15,998	-,276	-,840	,448
Dummy_Dific_Problema	-18,066	16,199	-,352	-1,115	,327

a. Variable dependiente: Porcentaje requisitos

No esperábamos obtener conclusión alguna debido al reducido número de sujetos experimentales, pero los resultados contradijeron nuestras presunciones. En concreto, la experiencia del analista correla muy fuertemente ( $\beta=0.814$ ) con la efectividad del proceso de relevamiento, siendo el efecto casi significativo ( $p=0.52$ ). En otras palabras, los analistas experimentados realizan un relevamiento más eficaz que los analistas noveles, al contrario de lo indicado por [7].

El resto de las variables demográficas no muestran correlaciones ni importantes ni fácilmente interpretables (por ejemplo, la dificultad del problema correla negativamente con la efectividad de las entrevistas, lo cual es incongruente a primera vista).

## 5 Amenazas a la validez

Existen cuatro aspectos que amenazan a la validez de las conclusiones alcanzadas:

- Se ha confundido el problema a estudiar y el experimentador en el papel de cliente. Esto implica que todos los efectos que se adscriben al tipo de problema pueden deberse en realidad al experimentador.
- No existe ningún modo de asegurar que los sujetos experimentales emplean adecuadamente la entrevista independiente de contexto. Dado que se permiten desviaciones respecto a la plantilla para indagar acerca de aspectos particulares del problema que surgen durante la entrevista, es previsible que algunos sujetos realicen entrevistas abiertas en realidad. No obstante, nótese que este hecho no es necesariamente negativo, en la medida en que refleja probablemente la práctica real de los analistas y por lo tanto aumenta la validez externa del experimento.
- Los experimentadores podrían identificar, durante la sesión de relevamiento, el tipo de entrevista utilizada y sesgar la información suministrada a los sujetos. Ello es posible debido a que durante la sesión los experimentadores pueden identificar el tipo de entrevista simplemente escuchando las preguntas. No obstante, los experimentadores no están a favor de uno u otro tipo por lo que creemos que este sesgo no debería ocurrir. Es más, en las conclusiones no existe una mejora evidente de la utilización de la entrevista abierta frente a la independiente de contexto, así que concluimos que este sesgo no existe en este estudio experimental.
- Los problemas planteados poseían una complejidad reducida, de modo que los sujetos pudieran obtener en 20 minutos la mayor parte de la información relevante. Esto puede explicar la ausencia de diferencias notables entre los tipos de entrevista, ya que independientemente del tipo de entrevista utilizado siempre fue posible obtener toda la información relevante.

## 6 Conclusiones

El experimento realizado complementa en gran medida las investigaciones realizadas hasta el momento. Si bien la entrevista independiente de contexto parece más efectiva que la entrevista abierta, tal y como sugieren [7,13,18], los efectos

observados distan de ser significativos. Es notable que ninguno de estos autores hayan estudiado el efecto del problema bajo estudio, ya que éste se ha demostrado ser el aspecto de mayor influencia en la efectividad del relevamiento de requisitos.

Adicionalmente, la experiencia del analista parece ejercer también una acusada influencia en la efectividad del relevamiento. Este es un hallazgo importante ya que, si bien se apoya en el sentido común, contradice los hallazgos realizados en estudios experimentales previos por [7,13].

Finalmente, nuestra experiencia en la realización de experimentos acerca del relevamiento de requisitos ha aumentado considerablemente. No sólo comprendemos mejor la dinámica del experimento, sino que somos también conscientes de aspectos que afectan a los resultados experimentales, como la aleatorización de clientes o los efectos del cansancio.

## Referencias

- [1] Swebok: Software Engineering Body of Knowledge, <http://www.swebok.org> (2005)
- [2] Davis, A.: Software Requirements: Objects, Functions and States. Englewood Cliffs, N.J. Prentice-Hall (1993)
- [3] Loucopoulos, P., Karakostas V.: System Requirements Engineering. McGraw-Hill (1995)
- [4] Ericsson, K.A., Simon H.A.: Protocol Analysis: Verbal Reports as Data. Cambridge, MA, USA, MIT Press (1996)
- [5] LaFrance, M.: The Knowledge Acquisition Grid: A Method for Training Knowledge Engineers. Int. Journal of Man-Machine Studies, vol 26 (2), pp. 245-255 (1987)
- [6] Wood, J., Silver, D.: Joint Application Design: How to Design Quality Systems in 40% Less Time. New York, Wiley (1989)
- [7] Agarwal, R., Tanniru, M. R.: Knowledge Acquisition Using Structured Interviewing: An Empirical Investigation. Journal of Manag. Inf. Systems, vol. 7 (1), pp. 123-141 (1990)
- [8] Shadbolt, N. R., Burton, A. M.: Knowledge Elicitation. In: Wilson, J. R., Corlett, E. N. (eds) Evaluation of Human Work: A Practical Ergonomics Methodology. London, UK, Taylor & Francis, pp. 321-345 (1990)
- [9] Gause, D. C., Weinberg, G. M.: Exploring Requirements: Quality Before Design. New York, Dorset House Publishing (1989)
- [10] Kotonya, G., Sommerville, I.: Requirements Engineering. Wiley (1998)
- [11] McConnell, S.: Context-free Requirement Questions. <http://www.construx.com/Page.aspx?hid=1742>
- [12] Dieste, O., Juristo, N.: Systematic Review and Aggregation of Empirical Studies on Elicitation Techniques. IEEE Transactions on Software Engineering, vol. 99 (2010)
- [13] Marakas, G. M., Elam, J. J.: Semantic Structuring in Analyst and Representation of Facts in Requirements Analysis. Information Systems Research, vol. 9, pp. 37-63 (1998)
- [14] Browne, G. J., Rogich, M. B.: An Empirical Investigation of User Requirements Elicitation: Comparing the Effectiveness of Prompting Techniques. Journal of Management Information Systems, vol. 17, pp. 223-249 (2001)
- [15] Juristo, N., Moreno, A.M.: Basics on Software Engineering Experimentation. Kluwer Academic Publishers (2001)
- [16] Sjoberg, D., Hannay, J., Hansen, O., Kampenes, V., Karahasanovic, A., Liborg, N., Rekdal, A.: A Survey of Controlled Experiments in Software Engineering. IEEE Transactions on Software Engineering, vol 31 (9), pp. 733 - 753 (2005)
- [17] Brown, B.W.: Crossover Experiment for Clinical Trials. Biometrics, vol. 36, pp. 69-79 (1980)
- [18] Browne, G. J. et al.: Evoking Information in Probability Assessment: Knowledge Maps and Reasoning-based Directed Questions. Management Science, vol. 43, pp. 1-14 (1997)

## Anexo A: Entrevista independiente de contexto

1	¿Qué objetivos persigue con el desarrollo del Sistema?
2	¿En qué tipos de tareas/actividades de la organización debe ayudar el sistema?
3	¿Qué beneficios espera obtener?
4	¿Qué características desea que posea/tareas debe realizar el nuevo sistema?
5	¿Qué usuarios deberán utilizar el nuevo sistema?
6	¿Qué individuos, además de los usuarios, se verán afectados por la implantación del sistema?
7	¿Debe el nuevo sistema interactuar con otros sistemas software preexistentes/que se implementarán en el futuro?
8	¿Debe el nuevo sistema interactuar con otros sistemas hardware, Bases de datos, etc. preexistentes/que se implementarán en el futuro?
9	¿Existe alguna restricción temporal para el desarrollo del sistema?
10	¿Existe alguna restricción (estándar, regulación, etc.) que afecte al desarrollo del sistema?

A continuación, se podrían formular de nuevo las preguntas anteriores (donde sea aplicable o simplemente interesante) para cada una de las características del sistema que se han identificado anteriormente (simplemente, sustituir la palabra "SISTEMA" por cada característica de interés). También es posible salirse del guión previsto si ello fuera deseable o prometedor.

## Anexo B: Datos obtenidos en el experimento

Datos experimentales

	Nombre Sujeto	Problema	Tipo de entrevista	Sesión	Experiencia total en años	Experiencia como analista en años	Conocimiento del problema del sujeto	Dificultad del problema para el sujeto	Porcentaje requisitos
1	A	Comisiones	Entrevista IC	1	8	6	Familiar	Baja	100
2	B	Comisiones	Entrevista IC	2	9	0	Conocido	Media	64
3	C	Comisiones	Entrevista IC	1	.	.	.	.	100
4	D	Comisiones	Entrevista IC	2	8	4	Familiar	Baja	14
5	F	Comisiones	Entrevista A	2	16	2	Familiar	Alta	.
6	G	Comisiones	Entrevista IC	2	10	5	Familiar	Baja	.
7	H	Comisiones	Entrevista A	2	5	2	Familiar	Baja	100
8	I	Comisiones	Entrevista A	2	5	2	Familiar	Baja	29
9	J	Comisiones	Entrevista A	1	9	8	Familiar	Media	71
10	K	Comisiones	Entrevista IC	1	5	1	Desconocido	Media	43
11	L	Comisiones	Entrevista A	2	3	2	Familiar	Media	64
12	M	Comisiones	Entrevista A	.	5	1	Desconocido	Media	.
13	A	Reciclado	Entrevista IC	2	8	6	Familiar	Baja	39
14	B	Reciclado	Entrevista IC	1	9	0	Familiar	Media	22
15	C	Reciclado	Entrevista IC	2	.	.	.	.	33
16	D	Reciclado	Entrevista IC	1	8	4	Desconocido	Media	56
17	E	Reciclado	Entrevista A	1	.	.	.	.	44
18	F	Reciclado	Entrevista A	2	16	2	Conocido	Alta	.
19	G	Reciclado	Entrevista IC	.	10	5	Familiar	Baja	.
20	H	Reciclado	Entrevista A	1	5	2	Desconocido	Media	33
21	I	Reciclado	Entrevista A	1	5	2	Familiar	Media	17
22	J	Reciclado	Entrevista A	2	9	8	Desconocido	Media	33
23	K	Reciclado	Entrevista IC	2	5	1	Desconocido	Media	44
24	L	Reciclado	Entrevista A	1	3	2	Desconocido	Alta	50
25	M	Reciclado	Entrevista A	1	5	1	Desconocido	Media	39
Total	N	25	25	23	22	22	25	25	20