

Jaime Urquiza Fuentes
Carlos A. Lázaro Carrascosa
J. Ángel Velázquez Iturbide

Evaluación de la Usabilidad de R-Zoom, una Interfaz para la Construcción de Animaciones de Programas

Número 2008-01

Serie de Informes Técnicos DLSI1-URJC
ISSN 1988-8074
Departamento de Lenguajes y Sistemas Informáticos I
Universidad Rey Juan Carlos

Índice

1. Introducción	1
2. Diseño de la evaluación	2
2.1. Infraestructura	2
2.2. Participantes	2
2.3. Protocolo	2
3. Resultados de la evaluación	4
3.1 Validez de los datos para el estudio comparativo	5
3.2 Análisis comparativo de la eficacia	9
3.3 Análisis comparativo de la eficiencia	14
3.4 Análisis comparativo de la satisfacción de los usuarios	19
Referencias	20
Apéndice: Cuestionario de opinión sobre la evaluación de R-Zoom	21

Evaluación de la usabilidad de R-Zoom, una interfaz para la construcción de animaciones de programas

Jaime Urquiza Fuentes, Carlos A. Lázaro Carrascosa, and Ángel Velázquez Iturbide

Universidad Rey Juan Carlos, C/Tulipán s/n, 28933, Móstoles Madrid
{jaime.urquiza, carlos.lazaro, angel.velazquez}@urjc.es

Summary. Este trabajo presenta los detalles de la evaluación de la usabilidad de R-Zoom, una interfaz para la creación de animaciones de programas. R-Zoom es una interfaz que implementa una solución de tipo Focus+Context, y ha sido evaluada de forma comparativa con otra que implementaba una solución de tipo Overview+Detail. Los resultados muestran que los usuarios experimentados que usaron R-Zoom fueron más eficaces y eficientes, además R-Zoom produce mayor satisfacción de usuario.

1 Introducción

Este trabajo se enmarca dentro de un proyecto de desarrollo de un entorno de programación con capacidad para construir animaciones de programas [3]. El entorno produce automáticamente representaciones gráficas (visualizaciones) de cada uno de los pasos de ejecución del programa, y la tarea que debe realizar el usuario es seleccionar de entre todas las visualizaciones, aquellas que formarán parte de la animación. Para ello hemos diseñado *R-Zoom*, una técnica de visualización de información basada en los principios *Focus+Context*[2], que permite mezclar visualizaciones reducidas (miniaturas) y versiones detalladas de las mismas, de forma que el usuario no pierde el contexto de trabajo.

Nos hemos centrado en evaluar la usabilidad de *R-Zoom* siguiendo el estándar de usabilidad proporcionado por la norma ISO 9241-11 [1], analizaremos la eficacia, eficiencia y la satisfacción de los usuarios al realizar las tareas con ambas interfaces. Como el objetivo principal de esta técnica es seleccionar las visualizaciones que formarán parte de la animación, las tareas de la evaluación consistirán en la búsqueda de determinadas visualizaciones dentro de una colección de miniaturas.

Somos conscientes de las ventajas de las interfaces *Overview+Detail*[2], así que en esta evaluación compararemos *R-Zoom* con una solución *Overview+Detail*, en adelante *O+D*, a la que hemos añadido las mejoras relacionadas con la generación de las miniaturas.

2 Diseño de la evaluación

Esta evaluación consistió en una sesión experimental de dos horas de duración. A continuación detallamos la infraestructura, los participantes y el protocolo de la sesión.

2.1 Infraestructura

La infraestructura consistió en ordenadores tipo *PC* con el sistema operativo Microsoft Windows XP Professional y la siguiente configuración hardware: procesador Pentium III 933 MHz, 256MB de memoria RAM, monitor de 17" Hitachi CM620ET funcionando a una resolución de 1024x768 píxeles y tarjetas gráficas Intel 82815 AGP 32MB. Hemos utilizado dos aplicaciones¹ para realizar la evaluación: el software de monitorización que mide tanto el tiempo empleado por los usuarios en realizar las tareas como los errores cometidos por estos, y el entorno *WinHIPE* especialmente configurado para la utilización de las interfaces a comparar, *R-Zoom* y *O+D*. Configuramos ambas interfaces para que usaran todo el espacio disponible en la pantalla, así como aplicar un factor de reducción inicial a todas las visualizaciones de un 50%. Este factor de reducción representa una solución de compromiso entre el número de miniaturas visibles simultáneamente en pantalla y su grado de comprensión. Rensink [4] demuestra que los factores de reducción menores del 50% no tienen efecto en tareas de reconocimiento de formas y siluetas, característica fundamental en el reconocimiento de las visualizaciones generadas en *WinHIPE*.

2.2 Participantes

Los usuarios que participaron en la evaluación eran usuarios finales de los interfaces a evaluar, con visión normal o corregida, y su participación fue absolutamente voluntaria sin ningún tipo de incentivo. Dichos usuarios eran estudiantes de primer año de Ingeniería Técnica Informática en la Universidad Rey Juan Carlos. Los estudiantes, un total de 43, estaban divididos en dos turnos distintos (mañana y tarde), y para tratar de evitar posibles desequilibrios entre los integrantes de cada turno, dividimos cada uno en dos grupos: el grupo A con 26 estudiantes en total y el grupo B con 17; ambos grupos usaron los dos interfaces, pero en distinto orden.

2.3 Protocolo

Las tareas a realizar durante la evaluación incluyeron ampliar las miniaturas, así como navegar a través de las colecciones de estas buscando una en particular. Las visualizaciones a buscar en cada tarea son pistas en sí mismas, ya que al representar estados de ejecución de un algoritmo, el usuario puede intuir la posición relativa de la miniatura en la colección. Consideramos que una tarea se termina con éxito si el usuario introduce el número correcto asociado a la visualización objetivo en la caja de texto del software de monitorización y hace click en el botón "siguiente" (véase la figura 1). Inmediatamente después se le presenta al usuario una nueva tarea.

¹ Ambas desarrolladas en Borland Delphi 5

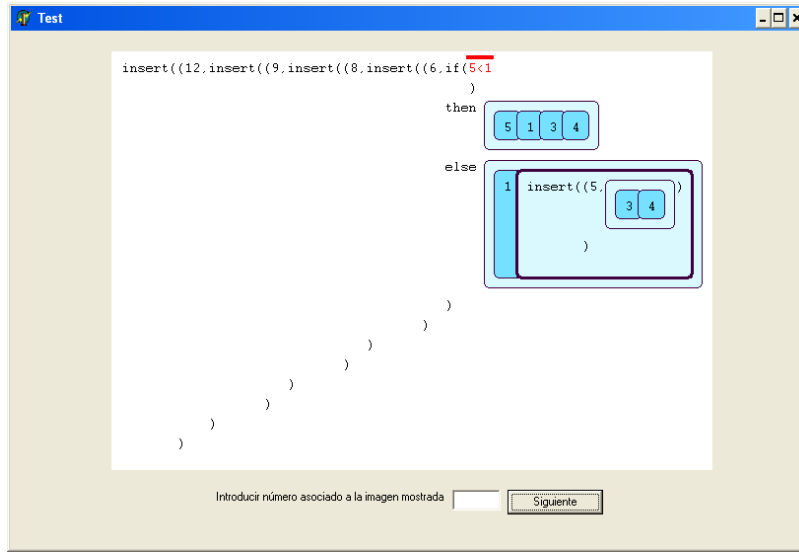


Fig. 1. Figura captura del sw de monitorización

Ambos grupos utilizaron los dos interfaces pero en distinto orden, así dispondremos de datos para comparar medidas entre usuarios con los dos interfaces y entre interfaces con el mismo usuario. El orden de ejecución de las tareas, del que mostramos un esquema en la figura 2, fue el siguiente. En primer lugar el instructor encargado de la evaluación explicó a cada grupo cómo se manejaba cada interfaz, *R-Zoom* al grupo A y *O+D* al grupo B. A continuación, cada grupo realizó una tarea de entrenamiento con la interfaz correspondiente, esta tarea era de las mismas características que aquellas a monitorizar. Las colecciones de miniaturas utilizadas tanto para las tareas de entrenamiento como para las monitorizadas eran resultado de la ejecución de diferentes algoritmos.

Grupo A	Grupo B
Explicación del interfaz correspondiente	
Tarea de entrenamiento con R-Zoom	Tarea de entrenamiento con OPD
9 tareas con R-Zoom	9 tareas con OPD
Cambio de interfaz + Explicación del siguiente interfaz	
9 tareas con OPD	9 tareas con R-Zoom
Responder al cuestionario de opinión	

Fig. 2. Esquema del protocolo seguido en la evaluación

Los usuarios realizaron un total de nueve tareas con cada interfaz: tres tareas diferentes (con el objetivo de búsqueda al principio, a la mitad y al final de la colección) con tres tipos diferentes de colecciones (ocupando 1 pantalla, pantalla y media, y tres pantallas). La figura 2 muestra un esquema del protocolo seguido en el experimento. El orden de presentación de las tareas estaba fijado previamente y los usuarios no sabían que las visualizaciones objetivos estaban en las zonas antes mencionadas. La colección y la imagen objetivo asociadas a cada tarea eran las mismas, independientemente del grupo al que pertenecía el usuario. Para evitar que los usuarios se quedaran atascados en una tarea, se les permitió un límite de 4 fallos por tarea; después del cuarto error, el software de monitorización comunicaba al usuario la respuesta correcta y continuaba con la siguiente tarea.

Los usuarios del grupo A realizaron primero nueve tareas con la interfaz *R-Zoom*. Simultáneamente, los usuarios del grupo B realizaron las mismas tareas con la interfaz *O+D*. Una vez completadas las primeras nueve tareas, cada grupo cambió de interfaz y realizó otras nueve tareas, estas tareas comprendían diferentes imágenes, en las mismas posiciones – principio, mitad y final de la colección –, pero en diferente orden a las anteriores y con las mismas colecciones anteriormente usadas.²

Las variables dependientes de la evaluación fueron: el número de errores por cada tarea, el tiempo empleado en relizar las tareas y la satisfacción de los usuarios con los interfaces. El tiempo empleado en realizar la tarea era el tiempo transcurrido desde que el usuario hizo click por última vez en el botón “siguiente” (comienzo de la primera tarea, o fin de la tarea anterior), hasta que el usuario hace click en el botón “siguiente” con el número correcto asociado a la visualización objetivo.

Una vez terminadas todas la tareas, pedimos a los usuarios que hicieran comentarios sobre los interfaces evaluados, así medimos la satisfacción de los usuarios. Para ello les pedimos que contestaran a un cuestionario con preguntas sobre: impresión general, utilidad, facilidad de uso, necesidad de estos interfaces en la construcción de animaciones de programas, ventajas, e inconvenientes. En el apéndice 4 se puede encontrar una copia de dicho cuestionario.

3 Resultados de la evaluación

Hemos analizado los resultados de la evaluación de dos formas distintas: basado en grupo-interfaz, y basado en tarea. El análisis basado en grupo-interfaz (en adelante GIB) referencia los datos según el grupo (A o B) y la interfaz usada, RZ (*R-Zoom*) u OD (*O+D*). Así podremos diferenciar cuatro grupos distintos: ARZ, AOD, BRZ y BOD. Además, este análisis contempla la experiencia de los usuarios con las tareas de la evaluación. El grupo A comenzó con la interfaz *R-Zoom*, por lo que ARZ implica sólo experiencia de una tarea, la de entrenamiento, lo mismo ocurre con BOD. AOD y BRZ implican una experiencia de la tarea de entrenamiento junto con las nueve tareas con la primera interfaz utilizada, *R-Zoom* en el caso de AOD y *O+D* en el caso de BRZ.

En el análisis basado en tarea (en adelante TB), referenciamos los datos usando el análisis GIB junto con la información de la colección y la imagen objetivo. Así,

² En <http://www.escet.urjc.es/~jurquiza/rzoom/demos.htm> se encuentran un conjunto de vídeos de demostración que sirven como ejemplo a una sesión de la evaluación.

BOD12 identifica a la tarea donde los usuarios del grupo B, usando la interfaz *O+D*, deben encontrar la primera imagen de la segunda colección.

A continuación verificamos la validez de los datos medidos en cuanto a errores cometidos y tiempo empleado por los usuarios, para finalmente analizar de forma separada los resultados de eficacia, eficiencia, y satisfacción de los usuarios.

3.1 Validez de los datos para el estudio comparativo

Idealmente, los usuarios del mismo grupo habrían realizado las tareas bajo las mismas circunstancias; de esta forma, podemos analizar los datos independientemente de los diferentes turnos en los que se dividían los grupos. Hemos comprobado si los diferentes turnos pertenecientes al mismo grupo realmente pertenecen a la misma población. Los datos referentes a los errores cometidos no muestran diferencia alguna entre turnos de un mismo grupo (véanse las figuras 3 y 4), por lo que podrán ser tratados como dos grupos en el estudio comparativo.

Test Statistics ^b						
	ENTREN	ARZ11	ARZ12	ARZ13	ARZ21	ARZ22
Mann-Whitney U	49,000	54,500	52,500	57,000	58,000	55,500
Wilcoxon W	77,000	244,500	80,500	247,000	248,000	245,500
Z	-1,472	-1,109	-1,294	-1,648	-,719	-1,014
Asymp. Sig. (2-tailed)	,141	,267	,196	,099	,472	,311
Exact Sig. [2*(1-tailed Sig.)]	,334 ^a	,497 ^a	,427 ^a	,611 ^a	,651 ^a	,534 ^a

Test Statistics ^b						
	ARZ23	ARZ31	ARZ32	ARZ33	AOD11	AOD12
Mann-Whitney U	59,500	63,000	66,500	59,500	66,500	60,500
Wilcoxon W	87,500	91,000	94,500	87,500	94,500	250,500
Z	-,876	-,607	,000	-,876	,000	-,751
Asymp. Sig. (2-tailed)	,381	,544	1,000	,381	1,000	,453
Exact Sig. [2*(1-tailed Sig.)]	,692 ^a	,866 ^a	1,000 ^a	,692 ^a	1,000 ^a	,735 ^a

Test Statistics ^b					
	AOD13	AOD21	AOD22	AOD23	AOD31
Mann-Whitney U	66,500	59,500	48,500	61,000	60,500
Wilcoxon W	94,500	87,500	238,500	251,000	250,500
Z	,000	-,876	-1,425	-,688	-,751
Asymp. Sig. (2-tailed)	1,000	,381	,154	,492	,453
Exact Sig. [2*(1-tailed Sig.)]	1,000 ^a	,692 ^a	,306 ^a	,778 ^a	,735 ^a

Test Statistics ^b		
	AOD32	AOD33
Mann-Whitney U	56,000	56,000
Wilcoxon W	84,000	84,000
Z	-1,096	-1,094
Asymp. Sig. (2-tailed)	,273	,274
Exact Sig. [2*(1-tailed Sig.)]	,572 ^a	,572 ^a

a. Not corrected for ties.

Fig. 3. Grupo A, datos de errores. Estos datos no siguen una distribución normal, por lo que hemos utilizado el test de *Mann-Whitney*. Todas las probabilidades (*Asymp Sig.* / *Exact Sig.*) son mayores que .05, por lo que no hay diferencias significativas

Test Statistics ^b						
	ENTREN.	BOD11	BOD12	BOD13	BOD21	BOD22
Mann-Whitney U	36,000	31,500	28,000	31,500	36,000	31,500
Wilcoxon W	81,000	76,500	64,000	76,500	81,000	76,500
Z	,000	-1,061	-1,377	-1,061	,000	-1,061
Asymp. Sig. (2-tailed)	1,000	,289	,168	,289	1,000	,289
Exact Sig. [2*(1-tailed Sig.)]	1,000 ^a	,673 ^a	,481 ^a	,673 ^a	1,000 ^a	,673 ^a

Test Statistics ^b						
	BOD23	BOD31	BOD32	BOD33	BRZ11	BRZ12
Mann-Whitney U	32,000	28,000	28,000	35,500	36,000	31,500
Wilcoxon W	68,000	64,000	64,000	80,500	81,000	76,500
Z	-,943	-1,377	-1,374	-,086	,000	-1,061
Asymp. Sig. (2-tailed)	,346	,168	,169	,931	1,000	,289
Exact Sig. [2*(1-tailed Sig.)]	,743 ^a	,481 ^a	,481 ^a	,963 ^a	1,000 ^a	,673 ^a

Test Statistics ^b					
	BRZ13	BRZ21	BRZ22	BRZ23	BRZ31
Mann-Whitney U	31,500	36,000	36,000	36,000	32,000
Wilcoxon W	76,500	81,000	81,000	81,000	68,000
Z	-1,061	,000	,000	,000	-,943
Asymp. Sig. (2-tailed)	,289	1,000	1,000	1,000	,346
Exact Sig. [2*(1-tailed Sig.)]	,673 ^a	1,000 ^a	1,000 ^a	1,000 ^a	,743 ^a

Test Statistics ^b		
	BRZ32	BRZ33
Mann-Whitney U	31,500	36,000
Wilcoxon W	76,500	81,000
Z	-1,061	,000
Asymp. Sig. (2-tailed)	,289	1,000
Exact Sig. [2*(1-tailed Sig.)]	,673 ^a	1,000 ^a

a. Not corrected for ties.

Fig. 4. Grupo B, datos de errores. Estos datos no siguen una distribución normal, por lo que hemos utilizado el test de *Mann-Whitney*. Todas las probabilidades (*Asymp Sig.* / *Exact Sig.*) son mayores que .05, por lo que no hay diferencias significativas

Sin embargo, sí hemos encontrado diferencias en los datos sobre el tiempo empleado por los usuarios para realizar las tareas. En la tabla 1 mostramos los detalles estadísticos de dichas tareas. El análisis completo se puede ver en las figuras 5 y 6.

Tarea	Valores <i>t</i> y <i>p</i>
AOD12	$t(24) = -2.261, p = 0.033$
BOD22	$t(15) = 2.722, p = 0.016$
BRZ11	$t(15) = 2.486, p = 0.025$
BRZ12	$t(15) = 2.409, p = 0.029$

Table 1. Detalle del análisis estadístico de igualdad de medias para las tareas con diferencias significativas entre turnos

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
ENTREN.	Equal variances assumed	5,332	,030	1,665	24	,109	59,21805	35,57208	-14,19912	132,63521
	Equal variances not assumed			2,444	23,993					
ARZ11	Equal variances assumed	6,343	,019	1,347	24	,191	49,23308	36,56366	-26,23061	124,69678
	Equal variances not assumed			2,042	23,673					
ARZ12	Equal variances assumed	,458	,505	,837	24	,411	10,36090	12,37366	-15,17707	35,89887
	Equal variances not assumed			1,107	20,622					
ARZ13	Equal variances assumed	2,128	,158	-,968	24	,343	-5,66165	5,84645	-17,72814	6,40483
	Equal variances not assumed			-,789	7,898					
ARZ21	Equal variances assumed	11,522	,002	-1,525	24	,140	-25,81203	16,93131	-60,75654	9,13248
	Equal variances not assumed			-1,071	6,810					
ARZ22	Equal variances assumed	1,132	,298	1,265	24	,218	17,81203	14,07699	-11,24146	46,86552
	Equal variances not assumed			1,769	23,006					
ARZ23	Equal variances assumed	,606	,444	1,564	24	,131	16,90226	10,81004	-5,40858	39,21309
	Equal variances not assumed			1,793	14,458					
ARZ31	Equal variances assumed	1,717	,202	,559	24	,581	24,32331	43,50500	-65,46660	114,11322
	Equal variances not assumed			,434	7,479					
ARZ32	Equal variances assumed	,059	,810	1,596	24	,124	22,10526	13,85114	-6,48209	50,69262
	Equal variances not assumed			1,489	9,521					
ARZ33	Equal variances assumed	7,747	,010	,289	24	,775	6,59398	22,83609	-40,53738	53,72535
	Equal variances not assumed			,386	21,198					
AOD11	Equal variances assumed	,411	,528	-1,008	24	,323	-81,26316	80,61569	-247,84577	85,11945
	Equal variances not assumed			-1,230	16,889					
AOD12	Equal variances assumed	,365	,551	-2,261	24	,033	-21,34586	9,44235	-40,83391	-1,85782
	Equal variances not assumed			-2,245	10,612					
AOD13	Equal variances assumed	,472	,499	-1,329	24	,196	-20,29323	15,27043	-51,80985	11,22338
	Equal variances not assumed			-1,257	9,745					
AOD21	Equal variances assumed	,070	,794	-,808	24	,427	-18,14286	22,44870	-64,47471	28,18899
	Equal variances not assumed			-,903	13,590					
AOD22	Equal variances assumed	1,550	,225	-,430	24	,671	-5,74436	13,36390	-33,32609	21,83737
	Equal variances not assumed			-,528	17,151					
AOD23	Equal variances assumed	,237	,630	-,159	24	,875	-1,62406	10,21107	-22,69868	19,45056
	Equal variances not assumed			-,153	9,987					
AOD31	Equal variances assumed	,132	,719	-1,180	24	,250	-24,09774	20,43009	-66,26337	18,06788
	Equal variances not assumed			-1,282	12,745					
AOD32	Equal variances assumed	6,259	,020	-1,681	24	,106	-39,03759	23,22445	-86,97050	8,89531
	Equal variances not assumed			-1,254	7,185					
AOD33	Equal variances assumed	,116	,737	,564	24	,578	11,41353	20,23937	-30,35848	53,18555
	Equal variances not assumed			,550	10,266					

Fig. 5. Grupo A, datos de tiempo empleado en realizar las tareas. Estos datos siguen una distribución normal, por lo que hemos utilizado el test *t-student para igualdad de medias*. En este caso, sí observamos probabilidades menores que .05

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
ENTREN	Equal variances assumed	9,525	,008	-1,506	15	,153	-34,30556	22,78017	-82,86035	14,24924
	Equal variances not assumed			-1,586	9,632	,145	-34,30556	21,62583	-82,74183	14,13072
BOD11	Equal variances assumed	6,193	,025	,684	15	,504	23,86111	34,87810	-50,47981	98,20203
	Equal variances not assumed			,708	12,552	,492	23,86111	33,69764	-49,20318	96,92540
BOD12	Equal variances assumed	,006	,940	,972	15	,347	16,45833	16,93733	-19,64272	52,55939
	Equal variances not assumed			,959	13,444	,355	16,45833	17,17024	-20,51163	53,42830
BOD13	Equal variances assumed	,607	,448	,406	15	,690	3,65278	8,99348	-15,51637	22,82193
	Equal variances not assumed			,413	14,734	,686	3,65278	8,85003	-15,24030	22,54586
BOD21	Equal variances assumed	1,073	,317	,021	15	,983	,52778	24,77775	-52,28475	53,34031
	Equal variances not assumed			,021	13,304	,984	,52778	25,14578	-53,67043	54,72599
BOD22	Equal variances assumed	3,023	,103	2,722	15	,016	20,54167	7,54743	4,45469	36,62864
	Equal variances not assumed			2,675	12,969	,019	20,54167	7,67880	3,94855	37,13478
BOD23	Equal variances assumed	,613	,446	,882	15	,391	9,84722	11,15938	-13,93843	33,63288
	Equal variances not assumed			,893	14,945	,386	9,84722	11,03096	-13,67224	33,36668
BOD31	Equal variances assumed	,002	,968	,672	15	,512	12,65278	18,82368	-27,46894	52,77450
	Equal variances not assumed			,666	13,970	,516	12,65278	18,99847	-28,10314	53,40869
BOD32	Equal variances assumed	,074	,789	,247	15	,808	5,87500	23,79645	-44,84594	56,59594
	Equal variances not assumed			,246	14,379	,809	5,87500	23,92008	-45,30192	57,05192
BOD33	Equal variances assumed	,893	,360	1,404	15	,181	25,27778	18,00963	-13,10884	63,66439
	Equal variances not assumed			1,376	12,632	,193	25,27778	18,36766	-14,52113	65,07668
BRZ11	Equal variances assumed	1,247	,282	2,486	15	,025	183,80556	73,93282	26,22147	341,38964
	Equal variances not assumed			2,457	13,681	,028	183,80556	74,80595	23,01127	344,59984
BRZ12	Equal variances assumed	3,401	,085	2,409	15	,029	19,34722	8,03196	2,22751	36,46693
	Equal variances not assumed			2,341	11,316	,039	19,34722	8,26625	1,21520	37,47925
BRZ13	Equal variances assumed	,312	,585	,932	15	,366	5,88889	6,31673	-7,57489	19,35267
	Equal variances not assumed			,934	14,875	,365	5,88889	6,30322	-7,55594	19,33372
BRZ21	Equal variances assumed	,051	,825	,336	15	,742	4,48611	13,35966	-23,98933	32,96155
	Equal variances not assumed			,335	14,624	,742	4,48611	13,38881	-24,11552	33,08774
BRZ22	Equal variances assumed	,850	,371	,388	15	,704	2,98611	7,70113	-13,42846	19,40068
	Equal variances not assumed			,393	14,818	,700	2,98611	7,58917	-13,20713	19,17935
BRZ23	Equal variances assumed	1,641	,220	,865	15	,401	10,13889	11,72543	-14,85328	35,13106
	Equal variances not assumed			,825	8,811	,431	10,13889	12,28818	-17,74984	38,02762
BRZ31	Equal variances assumed	,255	,621	,227	15	,824	8,68056	38,24416	-72,83493	90,19604
	Equal variances not assumed			,228	14,957	,823	8,68056	38,07145	-72,48698	89,84809
BRZ32	Equal variances assumed	1,020	,329	,291	15	,775	3,30556	11,37058	-20,93026	27,54137
	Equal variances not assumed			,301	12,527	,768	3,30556	10,98403	-20,51531	27,12642
BRZ33	Equal variances assumed	2,389	,143	-,643	15	,530	-18,06944	28,11586	-77,99698	41,85809
	Equal variances not assumed			-,660	13,781	,520	-18,06944	27,39686	-76,91772	40,77883

Fig. 6. Grupo B, datos de tiempo empleado en realizar las tareas. Estos datos siguen una distribución normal, por lo que hemos utilizado el test *t-student para igualdad de medias*. En este caso, sí observamos probabilidades menores que .05

Por lo tanto, los datos sobre las tareas AOD12, BOD22, BRZ11 y BRZ12, no se utilizarán para el análisis comparativo, ya que los distintos turnos de cada grupo no se pueden tratar como uno solo.

Además detectamos otra anomalía en las tareas AOD11 y BRZ11, la primera tarea realizada después del cambio de interfaz en cada grupo. El tiempo invertido por los usuarios en esta tarea, era aproximadamente diez veces el tiempo invertido en las otras dos tareas de la misma colección (véase la tabla 2). Este dato nos dice que deberíamos haber incluido una tarea de entrenamiento en el cambio de interfaz.

	xxx11	xxx12	xxx13
AOD	482.61 s	34.11 s	42.88 s
BRZ	356.94 s	33.88 s	34.88 s

Table 2. Comparación de tiempos de la primera tarea después del cambio de interfaz (AOD11 y BRZ11) con las dos tareas siguientes en cada grupo

En conclusión, para el resto del análisis comparativo, decidimos ignorar los datos de tiempo referentes a las tareas involucradas: AOD11, AOD12, BOD22, BRZ11 y BRZ12. Pero existen más implicaciones, el análisis GIB estudia los datos independientemente de la tarea en concreto; por lo tanto al realizar las comparaciones entre los cuatro grupos habrá que comprobar la influencia de haber ignorado los datos de las tareas anteriormente referidas. Hemos descubierto que el tiempo empleado por los usuarios para completar las tareas y, tanto la colección como la imagen involucrada en dichas tareas, no son independientes. Para comprobar dicha influencia hemos verificado la dependencia entre el tiempo y, la colección y la imagen de cada tarea. Si no son independientes, entonces todo análisis comparativo deberá tener en cuenta los datos ignorados, y por lo tanto ignorar los datos homólogos. Utilizaremos los test de correlación $\tau - b$ de Kendall y ρ de Spearman. En la figura 7 mostramos el análisis de dependencias para ambos grupos. Los resultados muestran que los tiempos dependen tanto de la colección utilizada como de la imagen objetivo.

Por lo tanto, si ignoramos las tareas AOD11 y AOD12, todas las comparaciones con AOD deben ignorar sus respectivas tareas XXX11 y XXX12, lo mismo pasará con BOD22, BRZ11 y BRZ12.

3.2 Análisis comparativo de la eficacia.

Hemos caracterizado la eficacia de las interfaces como el número de errores cometidos por los usuarios al realizar las tareas. De las 774 tareas realizadas (43 usuarios x 2 interfaces x 3 colecciones x 3 visualizaciones), sólo detectamos errores en 59: 55 con un error y 4 con dos errores. A primera vista se puede decir que ambos interfaces son bastante eficaces.

Aquí detallamos el análisis de diferencias significativas según el análisis por grupo-interfaz (GIB) y por tarea (TB). Los test utilizados son el test de *Mann-Whitney* para comparar resultados entre los grupos A y B, y el test de *Wilcoxon* para comparar resultados dentro del mismo grupo pero con diferentes interfaces.

Correlations

			PHASE	COLECTION	IMAGE	TIME
Kendall's tau_b	TIME	Correlation Coefficient	-,019	,282**	-,448**	1,000
		Sig. (2-tailed)	,631	,000	,000	.
		N	416	416	416	416
Spearman's rho	TIME	Correlation Coefficient	-,024	,354**	-,578**	1,000
		Sig. (2-tailed)	,631	,000	,000	.
		N	416	416	416	416

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

Análisis de correlación en el grupo A

Correlations

			PHASE	COLECTION	IMAGE	TIME
Kendall's tau_b	TIME	Correlation Coefficient	-,230**	,195**	-,485**	1,000
		Sig. (2-tailed)	,000	,000	,000	.
		N	255	255	255	255
Spearman's rho	TIME	Correlation Coefficient	-,280**	,247**	-,623**	1,000
		Sig. (2-tailed)	,000	,000	,000	.
		N	255	255	255	255

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

Análisis de correlación en el grupo B

Fig. 7. Análisis de correlación entre el tiempo invertido en las tareas; y la fase, la colección y la imagen de dichas tareas

Análisis GIB

Tras realizar el análisis GIB, encontramos diferencias significativas entre algunas de las parejas *grupo-interfaz*. La tabla 3 muestra el resumen de los datos sobre errores cometidos. Así, para cada pareja grupo-interfaz proporcionamos el número total de errores cometidos, independientemente de si son en una misma tarea o en diferentes; el promedio de errores por tarea, que es la división del número de errores cometidos entre el número de tareas realizadas; y la tasa de error, que es el número de errores cometidos dividido entre el número máximo de errores que podrían haberse cometido (máximo de cuatro errores por tarea).

	Número de errores	Promedio de errores por tarea	Tasa de error
BOD	13	0.085	0.021
ARZ	24	0.102	0.025
BRZ	4	0.026	0.007
AOD	22	0.094	0.024

Table 3. Medidas GIB de la eficacia

Como podemos observar en las figuras 8 y 9, las tres comparaciones que se hacen con BRZ dan resultados significativos.

Ranks

	Group	N	Mean Rank	Sum of Ranks
BODvsAOD	0	153	193,16	29554,00
	1	234	194,55	45524,00
	Total	387		
BODvsARZ	0	153	191,71	29332,00
	1	234	195,50	45746,00
	Total	387		
BRZvsAOD	0	153	187,03	28616,00
	1	234	198,56	46462,00
	Total	387		
BRZvsARZ	0	153	185,55	28388,50
	1	234	199,53	46689,50
	Total	387		

Test Statistics^a

	BODvsAOD	BODvsARZ	BRZvsAOD	BRZvsARZ
Mann-Whitney U	17773,000	17551,000	16835,000	16607,500
Wilcoxon W	29554,000	29332,000	28616,000	28388,500
Z	-,249	-,655	-2,371	-2,724
Asymp. Sig. (2-tailed)	,803	,513	,018	,006

a. Grouping Variable: Group

Fig. 8. Test de *Mann-Whitney* para diferencias significativas (GIB) entre datos sobre errores cometidos por usuarios de diferentes grupos

Ranks

		N	Mean Rank	Sum of Ranks	
BRZ - BOD	Negative Ranks	12 ^a	8,67	104,00	a. BRZ < BOD
	Positive Ranks	4 ^b	8,00	32,00	b. BRZ > BOD
	Ties	137 ^c			c. BRZ = BOD
	Total	153			
AOD - ARZ	Negative Ranks	21 ^d	18,50	388,50	d. AOD < ARZ
	Positive Ranks	17 ^e	20,74	352,50	e. AOD > ARZ
	Ties	196 ^f			f. AOD = ARZ
	Total	234			

Test Statistics^b

	BRZ - BOD	AOD - ARZ	
Z	-2,065 ^a	-,293 ^a	a. Based on positive ranks.
Asymp. Sig. (2-tailed)	,039	,770	b. Wilcoxon Signed Ranks Test

Fig. 9. Test de *Wilcoxon* para diferencias significativas (GIB) entre datos sobre errores cometidos por usuarios del mismo grupo usando diferentes interfaces

Análisis TB

También hemos encontrado diferencias significativas en el análisis TB, véanse las figuras 10, 11 y 12. Sin embargo estas tienen un carácter más bien anecdótico ya que, de 41 comparaciones posibles, sólo dos han dado resultados significativos y no podemos sacar conclusiones de ellas.

Test Statistics^a

	BOD-ARZ-11	BOD-ARZ-12	BOD-ARZ-13	BOD-ARZ-21	BOD-ARZ-22
Mann-Whitney U	200,000	213,000	216,500	178,500	199,500
Wilcoxon W	353,000	366,000	567,500	331,500	352,500
Z	-,939	-,331	-,306	-1,901	-,960
Asymp. Sig. (2-tailed)	,348	,741	,759	,057	,337

Test Statistics^a

	BOD-ARZ-23	BOD-ARZ-31	BOD-ARZ-32	BOD-ARZ-33	BOD-AOD-11
Mann-Whitney U	217,000	203,500	195,000	212,000	208,000
Wilcoxon W	370,000	554,500	546,000	563,000	559,000
Z	-,225	-,985	-1,770	-,444	-1,237
Asymp. Sig. (2-tailed)	,822	,325	,077	,657	,216

Test Statistics^a

	BOD-AOD-12	BOD-AOD-13	BOD-AOD-21	BOD-AOD-22	BOD-AOD-23
Mann-Whitney U	212,000	208,000	204,000	183,000	216,500
Wilcoxon W	563,000	559,000	357,000	336,000	369,500
Z	-,444	-1,237	-1,157	-1,476	-,253
Asymp. Sig. (2-tailed)	,657	,216	,247	,140	,800

Test Statistics^a

	BOD-AOD-31	BOD-AOD-32	BOD-AOD-33	BRZ-ARZ-11	BRZ-ARZ-12
Mann-Whitney U	212,000	219,000	220,500	187,000	200,000
Wilcoxon W	563,000	570,000	373,500	340,000	353,000
Z	-,444	-,089	-,022	-1,678	-,939
Asymp. Sig. (2-tailed)	,657	,929	,982	,093	,348

Test Statistics^a

	BRZ-ARZ-13	BRZ-ARZ-21	BRZ-ARZ-22	BRZ-ARZ-23	BRZ-ARZ-31
Mann-Whitney U	216,500	178,500	187,000	204,000	216,500
Wilcoxon W	567,500	331,500	340,000	357,000	567,500
Z	-,306	-1,901	-1,677	-1,157	-,306
Asymp. Sig. (2-tailed)	,759	,057	,094	,247	,759

Fig. 10. Test de *Mann-Whitney* para diferencias significativas (TB) entre datos sobre errores cometidos por usuarios de diferentes grupos

Test Statistics^a

	BRZ-ARZ-32	BRZ-ARZ-33	BRZ-AOD-11	BRZ-AOD-12	BRZ-AOD-13
Mann-Whitney U	208,000	204,000	221,000	217,000	208,000
Wilcoxon W	559,000	357,000	572,000	370,000	559,000
Z	-1,237	-1,157	,000	-,225	-1,237
Asymp. Sig. (2-tailed)	,216	,247	1,000	,822	,216

Test Statistics^a

	BRZ-AOD-21	BRZ-AOD-22	BRZ-AOD-23	BRZ-AOD-31
Mann-Whitney U	204,000	170,000	204,000	217,000
Wilcoxon W	357,000	323,000	357,000	370,000
Z	-1,157	-2,110	-1,157	-,225
Asymp. Sig. (2-tailed)	,247	,035	,247	,822

Test Statistics^a

	BRZ-AOD-32	BRZ-AOD-33
Mann-Whitney U	208,500	195,500
Wilcoxon W	361,500	348,500
Z	-,617	-1,434
Asymp. Sig. (2-tailed)	,537	,151

a. Grouping Variable: GRP

Fig. 11. Test de *Mann-Whitney* para diferencias significativas (TB) entre datos sobre errores cometidos por usuarios de diferentes grupos

Test Statistics^d

	BRZ11 - BOD11	BRZ12 - BOD12	BRZ13 - BOD13	BRZ21 - BOD21	BRZ22 - BOD22	BRZ23 - BOD23
Z	-1,000 ^a	-,577 ^a	,000 ^b	,000 ^b	-1,000 ^a	-1,000 ^a
Asymp. Sig. (2-tailed)	,317	,564	1,000	1,000	,317	,317

Test Statistics^d

	BRZ31 - BOD31	BRZ32 - BOD32	BRZ33 - BOD33	AOD11 - ARZ11	AOD12 - ARZ12	AOD13 - ARZ13
Z	-,577 ^a	-,816 ^a	-1,414 ^a	-2,000 ^a	-,816 ^a	-1,000 ^a
Asymp. Sig. (2-tailed)	,564	,414	,157	,046	,414	,317

Test Statistics^d

	AOD21 - ARZ21	AOD22 - ARZ22	AOD23 - ARZ23	AOD31 - ARZ31	AOD32 - ARZ32	AOD33 - ARZ33
Z	-1,134 ^a	-,378 ^c	-,378 ^c	-,577 ^c	-1,732 ^c	-,816 ^c
Asymp. Sig. (2-tailed)	,257	,705	,705	,564	,083	,414

- a. Based on positive ranks.
- b. The sum of negative ranks equals the sum of positive ranks.
- c. Based on negative ranks.
- d. Wilcoxon Signed Ranks Test

Fig. 12. Test de *Wilcoxon* para diferencias significativas (TB) entre datos sobre errores cometidos por usuarios del mismo grupo usando diferentes interfaces

3.3 Análisis comparativo de la eficiencia.

Para caracterizar la eficiencia de los interfaces hemos medido el tiempo empleado en realizar cada tarea. Aunque las tasas de error son bastante bajas, han provocado diferencias entre los diferentes grupos, por lo que la eficiencia tendrá en cuenta tanto la eficacia (errores cometidos) como los recursos utilizados (tiempo empleado). Hemos caracterizado la eficiencia con la siguiente fórmula: $tiempo / (4 - errores)$. Cuanto menor sea este valor, mejor será el resultado de la eficiencia.

Análisis GIB

De las seis posibles comparaciones a realizar con el análisis GIB, detectamos tres con diferencias significativas. En la tabla 4 mostramos un resumen de los resultados en términos de eficiencia y tiempo empleado por los usuarios en realizar las tareas.

	Eficiencia (media , desviación típica)	Tiempo (media , desviación típica)
BOD	$M = 27,41 \quad SD = 18,680$	$M = 106,18 \quad SD = 71,148$
ARZ	$M = 24,79 \quad SD = 21,609$	$M = 94,31 \quad SD = 78,241$
BRZ	$M = 17,95 \quad SD = 14,712$	$M = 70,93 \quad SD = 57,087$
AOD	$M = 22,26 \quad SD = 16,012$	$M = 85,39 \quad SD = 60,497$

Table 4. Medida GIB de la eficiencia. Nótese que estas medidas son aproximadas, ya que los datos que se tendrán en cuenta en el estudio comparativo dependen de si los grupos que comparamos tienen alguna tarea anómala o no

En la comparativa entre AOD y BOD, al ignorarse los datos sobre AOD11, AOD12 y BOD22, tuvimos que ignorar también los datos sobre BOD11, BOD12 y AOD22. El resultado de la comparativa se puede ver en la Fig. 13

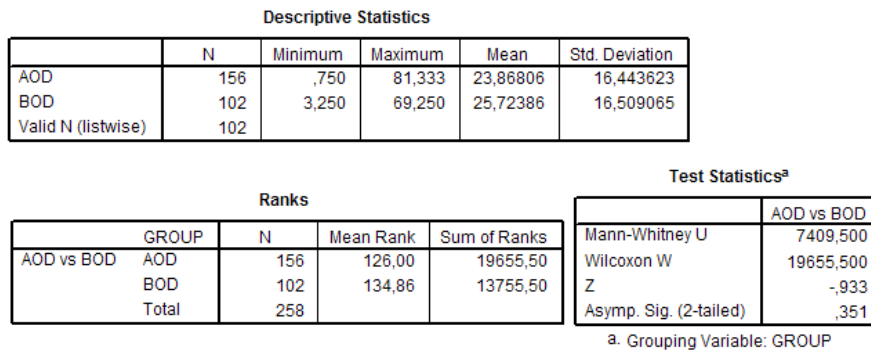


Fig. 13. Test de *Mann-Whitney* para diferencias significativas (GIB) entre AOD y BOD

En la comparativa entre ARZ y AOD, al ignorarse los datos sobre AOD11 y AOD12, tuvimos que ignorar también los datos sobre ARZ11 y ARZ12. En la comparativa entre BOD y BRZ, al ignorarse los datos sobre BOD22, BRZ11 y BRZ12, tuvimos que ignorar también los datos sobre BRZ22, BOD11 y BOD12. El resultado de ambas comparativas se puede ver en la figura 14.

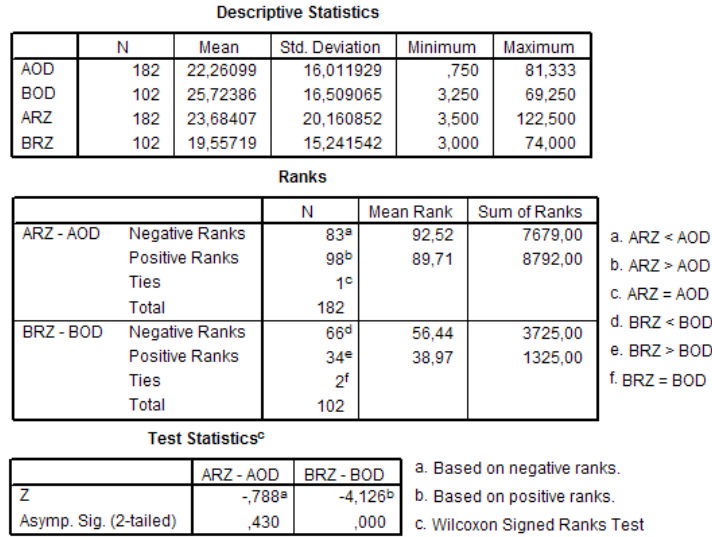


Fig. 14. Test de *Wilcoxon* para diferencias significativas (GIB) entre datos sobre el tiempo empleado por usuarios del mismo grupo usando diferentes interfaces

En la comparativa entre AOD y BRZ, no hace falta ignorar más tareas de las anómalas, ya que se ignoran los datos tanto de AOD11 y AOD12, como de BRZ11 y BRZ12. El resultado de la comparativa se puede ver en la Fig. 15

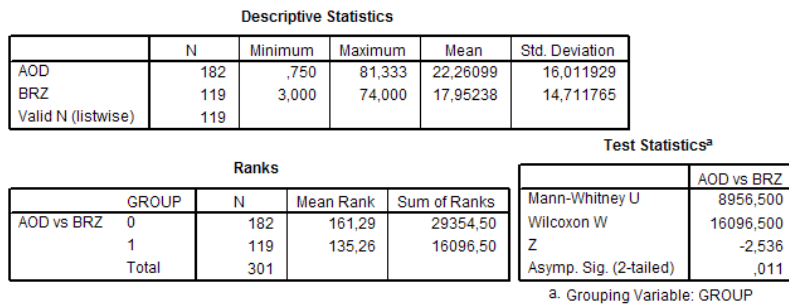


Fig. 15. Test de *Mann-Whitney* para diferencias significativas (GIB) entre entre AOD y BRZ

En la comparativa entre ARZ y BRZ, al ignorarse los datos sobre BRZ11 y BRZ12, tuvimos que ignorar también los datos sobre ARZ11 y ARZ12. El resultado de la comparativa se puede ver en la Fig. 16

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
ARZ	182	3,500	122,500	23,68407	20,160852
BRZ	119	3,000	74,000	17,95238	14,711765
Valid N (listwise)	119				

Ranks					Test Statistics ^a	
GROUP	N	Mean Rank	Sum of Ranks		ARZ vs BRZ	
ARZ vs BRZ	ARZ	182	162,09	29500,50	Mann-Whitney U	8810,500
	BRZ	119	134,04	15950,50	Wilcoxon W	15950,500
	Total	301			Z	-2,734
					Asymp. Sig. (2-tailed)	,006

a. Grouping Variable: GROUP

Fig. 16. Test de *Mann-Whitney* para diferencias significativas (GIB) entre entre ARZ y BRZ

En la comparativa entre ARZ y BOD, al ignorarse los datos sobre BOD22, tuvimos que ignorar también los datos sobre ARZ22. El resultado de la comparativa se puede ver en la Fig. 17

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
BOD	136	3,000	105,667	27,41238	18,679896
ARZ	208	1,500	127,667	26,03165	21,907309
Valid N (listwise)	136				

Ranks					Test Statistics ^a	
GROUP	N	Mean Rank	Sum of Ranks		BOD vs ARZ	
BOD vs ARZ	BOD	136	181,89	24737,00	Mann-Whitney U	12867,000
	ARZ	208	166,36	34603,00	Wilcoxon W	34603,000
	Total	344			Z	-1,416
					Asymp. Sig. (2-tailed)	,157

a. Grouping Variable: GROUP

Fig. 17. Test de *Mann-Whitney* para diferencias significativas (GIB) entre entre ARZ y BOD

Análisis TB

Debido a las anomalías detectadas en las tareas BRZ11 y BRZ12, y que todas las comparaciones son con el grupo BRZ, en general sólo hemos considerado el resto de tareas. Excepto en la comparativa con BOD, que también hemos tenido que ignorar BOD22.

En primer lugar mostramos las estadísticas descriptivas de todas las tareas que hemos tenido en cuenta, véase figura 18.

Hemos aplicado el análisis TB a los resultados del análisis previo, así que sólo hemos comparado las tareas correspondientes a las siguientes parejas: BOD-BRZ, ARZ-BRZ y BRZ-AOD. De las 20 comparaciones realizadas (3 parejas con 9 tareas cada una, menos 7 anómalas), hemos detectado 10 diferencias significativas, siempre en favor de los usuarios experimentados utilizando *R-Zoom*, que corresponden al grupo BRZ.

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
AOD13	26	10,7212	8,76565	1,00	36,50
AOD21	26	29,8173	15,91456	4,75	81,33
AOD22	26	12,6186	8,18285	3,00	30,25
AOD23	26	14,8782	6,61871	4,50	29,33
AOD31	26	45,4199	11,45679	26,50	71,00
AOD32	26	20,7468	14,37636	7,75	55,75
AOD33	26	21,6250	12,86628	,75	53,00
ARZ13	26	7,8494	3,94462	3,50	21,33
ARZ21	26	33,4135	12,56851	16,25	73,00
ARZ22	26	14,8654	16,20967	5,25	91,00
ARZ23	26	14,5321	7,23913	5,50	32,33
ARZ31	26	53,2019	26,49189	19,50	122,50
ARZ32	26	19,5385	8,07053	4,00	34,00
ARZ33	26	22,3878	15,30192	5,50	65,00
BOD13	17	10,7941	5,03677	3,25	21,00
BOD21	17	33,3676	12,34343	12,50	58,75
BOD23	17	16,4706	6,07240	5,00	25,50
BOD31	17	51,6373	8,88775	33,75	69,25
BOD32	17	23,6176	12,98380	7,50	46,50
BOD33	17	18,4559	9,70047	8,00	43,25
BRZ13	17	8,9412	3,53391	3,75	15,00
BRZ21	17	24,0000	6,68019	15,50	38,50
BRZ22	17	8,3235	3,85556	4,50	16,75
BRZ23	17	9,7206	5,98493	5,50	30,75
BRZ31	17	40,1716	20,32722	19,00	74,00
BRZ32	17	14,2745	5,79195	4,00	29,50
BRZ33	17	20,2353	14,19768	3,00	50,50

Fig. 18. Estadísticas descriptivas de las tareas consideradas en el análisis TB de eficiencia

A continuación, en la figura 19 mostramos el análisis ARZ-BRZ, que estudia el aumento de la experiencia en las tareas realizadas con *R-Zoom*. De nuevo, de las 6 comparaciones posibles, 4 tienen diferencias significativas. Mientras que en la figura 20 mostramos el análisis AOD-BRZ, que estudia el cambio de interfaz en usuarios experimentados, donde sólo 2 comparaciones de 7 ofrecen diferencias significativas.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
ARZ_BRZ_21	Equal variances assumed	3,414	,072	2,830	41	,007	9,41346	3,32639	2,69569	16,13123
	Equal variances not assumed			3,191	39,692	,003	9,41346	2,94969	3,45047	15,37645
ARZ_BRZ_23	Equal variances assumed	2,065	,158	2,276	41	,028	4,81146	2,11390	,54236	9,08056
	Equal variances not assumed			2,370	38,629	,023	4,81146	2,03042	,70329	8,91963
ARZ_BRZ_31	Equal variances assumed	,227	,636	1,721	41	,093	13,03035	7,57093	-2,25945	28,32016
	Equal variances not assumed			1,819	39,831	,076	13,03035	7,16231	-1,44713	27,50784
ARZ_BRZ_32	Equal variances assumed	3,956	,053	2,322	41	,025	5,26395	2,26657	,68653	9,84138
	Equal variances not assumed			2,487	40,567	,017	5,26395	2,11624	,98874	9,53917
ARZ_BRZ_33	Equal variances assumed	,154	,697	,464	41	,645	2,15253	4,64139	-7,22095	11,52600
	Equal variances not assumed			,471	36,178	,640	2,15253	4,56761	-7,10943	11,41448

Ranks					Test Statistics ^a	
GROUP	N	Mean Rank	Sum of Ranks			ARZ_BRZ_22
ARZ_BRZ_22	17	15,71	267,00		Mann-Whitney U	114,000
BRZ	26	26,12	679,00		Wilcoxon W	267,000
Total	43				Z	-2,659
					Asymp. Sig. (2-tailed)	,008

a. Grouping Variable: GROUP

Fig. 19. Análisis de eficiencia ARZ-BRZ

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
BRZ_AOD_13	Equal variances assumed	5,470	,024	,793	41	,432	1,77998	2,24323	-2,75032	6,31027
	Equal variances not assumed			,927	35,542	,360	1,77998	1,92090	-2,11754	5,67749
BRZ_AOD_21	Equal variances assumed	6,338	,016	1,423	41	,162	5,81731	4,08881	-2,44022	14,07484
	Equal variances not assumed			1,654	36,184	,107	5,81731	3,51657	-1,31338	12,94799
BRZ_AOD_22	Equal variances assumed	11,183	,002	2,017	41	,050	4,29506	2,12988	-,00632	8,59644
	Equal variances not assumed			2,312	38,012	,026	4,29506	1,85736	,53507	8,05505
BRZ_AOD_23	Equal variances assumed	,995	,324	2,592	41	,013	5,15762	1,98961	1,13952	9,17571
	Equal variances not assumed			2,649	36,772	,012	5,15762	1,94729	1,21121	9,10402
BRZ_AOD_31	Equal variances assumed	11,697	,001	1,083	41	,285	5,24830	4,84491	-4,53620	15,03280
	Equal variances not assumed			,969	22,710	,343	5,24830	5,41793	-5,96748	16,46409
BRZ_AOD_33	Equal variances assumed	,103	,749	,332	41	,741	1,38971	4,18003	-7,05204	9,83145
	Equal variances not assumed			,326	31,908	,747	1,38971	4,26899	-7,30692	10,08633

Ranks					Test Statistics ^a	
GROUP	N	Mean Rank	Sum of Ranks			BRZ_AOD_32
BRZ_AOD_32	17	20,29	345,00		Mann-Whitney U	192,000
AOD	26	23,12	601,00		Wilcoxon W	345,000
Total	43				Z	-,721
					Asymp. Sig. (2-tailed)	,471

a. Grouping Variable: GROUP

Fig. 20. Análisis de eficiencia AOD-BRZ

Con los resultados anteriores podemos ver cómo la mejora media en las tareas relacionadas con la segunda y tercera imagen, 36,13% y 36,25% respectivamente es mayor que la mejora media de la primera imagen, 26,15%. Además la mayoría de las tareas relacionadas con la segunda visualización, 4 tareas de 5 posibles, han dado diferencia significativa. Las razones para estos resultados se pueden justificar con una de las características principales de *R-Zoom*, mantener en la medida de lo posible la localización de los elementos del contexto. Así, después de haber encontrado la primera visualización, el usuario ha tenido la oportunidad de ver los elementos del contexto, donde se encuentran la segunda y tercera visualización, o al menos otras que actúan como pistas o indicadores de dónde se pueden encontrar.

3.4 Análisis comparativo de la satisfacción de los usuarios.

Para medir la satisfacción de los usuarios con ambas interfaces usamos un cuestionario (véase apéndice 4). En primer lugar les preguntamos sobre la necesidad de interfaces con posibilidad de zoom para construir animaciones de programas o algoritmos. El 98% de los usuarios están de acuerdo en mayor o menor medida sobre la necesidad de estas utilidades de zoom, véase la figura 21a. En cuanto a la preferencia de los usuarios por una u otra interfaz, figura 21b, el 86% de los usuarios prefieren *R-Zoom* frente al 12% que prefieren la interfaz *O+D*. A continuación preguntamos a los usuarios sobre la facilidad de uso, figura 21c, el 86% de los usuarios piensan que *R-Zoom* es más o mucho más fácil de usar que *O+D*, mientras que sólo el 8% piensan lo contrario. A continuación, preguntamos a los usuarios por la utilidad de ambas interfaces, figura 21d, el 64% de los usuarios piensan que *R-Zoom* es más o mucho más útil que *O+D* y el 34% piensan que ambos interfaces son igual de útiles.

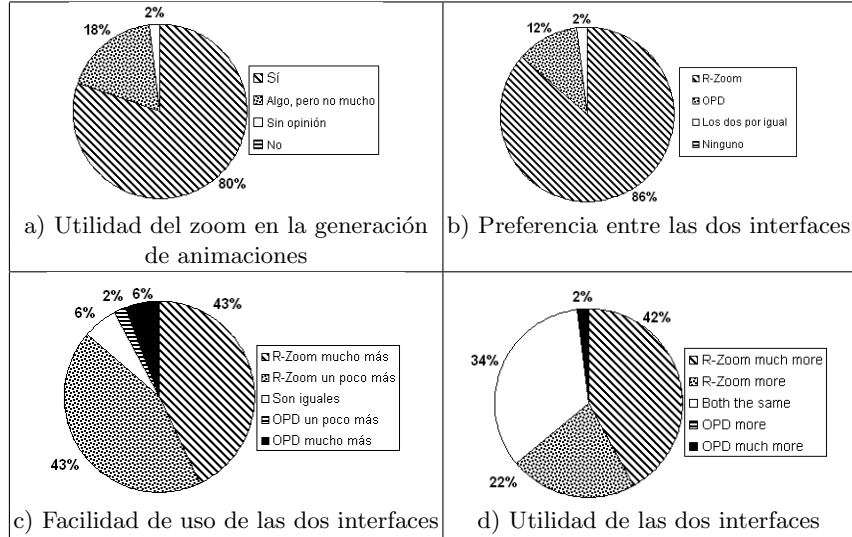


Fig. 21. Opinión de los usuarios

Finalmente, ofrecimos a los estudiantes la posibilidad de comentar aspectos positivos y negativos de ambas interfaces. Los comentarios se centraron principalmente en tres aspectos: la interfaz ofrece una visión global de la animación, separación/integración de las vistas global y detallada de las visualizaciones, y sobre la orientación en la colección de miniaturas. En la tabla 5 enumeramos estos comentarios, proporcionando el porcentaje de usuarios que han hecho ese comentario, si es una ventaja o un inconveniente, y una breve descripción del mismo.

40%	Ventaja	R-Zoom ofrece una visión global de la colección de miniaturas.
24%	Ventaja	R-Zoom muestra las vistas global y detallada en la misma ventana.
20%	Inconveniente	O+D muestra las vistas global y detallada en distintas ventanas.
17,1%	Ventaja	O+D ofrece una visión global de la colección de miniaturas.
17,1%	Ventaja	R-Zoom permite al usuario orientarse en la colección.
14,3%	Ventaja	R-Zoom permite moverse con los cursores.
11,4%	Inconveniente	Con O+D es fácil que el usuario se pierda.
8,6%	Ventaja	Con R-Zoom se identifica fácilmente el foco.
8,6%	Inconveniente	En ocasiones, en R-Zoom el aumento aplicado al foco es algo menor que en O+D.
6,7%	Ventaja	Ambas interfaces muestran una buena visualización de los pasos de una evaluación.
6,7%	Inconveniente	Con O+D es más difícil realizar las búsquedas de visualizaciones.
6,7%	Inconveniente	Con O+D, si el número de visualizaciones es alto la visibilidad es poca.
2,9%	Ventaja	R-Zoom permite manejar muchas visualizaciones a la vez.
2,9%	Inconveniente	Con O+D hay que cambiar el punto de atención constantemente.
2,9%	Inconveniente	Con O+D se solapan las vistas global y detallada.

Table 5. Listado de comentarios libres realizados por los usuarios. Muestra el % de usuarios que ha realizado el comentario, si se identificó como una ventaja o un inconveniente y una breve descripción del comentario

References

1. ISO 9241-11. Ergonomic requirements for office work with visual display terminals (vdts) - part 11: Guidance on usability, 1998.
2. S.K. Card, J.D. Mackinlay, and B. Shneidermann, editors. *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann Publishers, 1999.
3. J. Ángel Velázquez Iturbide, Cristóbal Pareja Flores, and Jaime Urquiza Fuentes. An approach to effortless construction of program animations. *Computers & Education*, In press, 2007.

4. R.A. Rensink. The invariance of visual search to geometric transformation. *Journal of Vision*, 4:178a, 2004.

4 Cuestionario de opinión sobre la evaluación de *R-Zoom*

Para cada pregunta, selecciona la opción que te parezca más oportuna, redondeando la letra correspondiente.

1. ¿Crees que una utilidad de zoom ayuda a la generación de animaciones con WinHIPE?
 - a) Sí.
 - b) Algo, pero no mucho.
 - c) Sin opinión.
 - d) No.
2. De los zoom con los que has trabajado ¿cuál te gusta más?:
 - a) El primero.
 - b) El segundo.
 - c) Los dos por igual.
 - d) Ninguno.
3. Opina sobre la utilidad de los dos zoom que has probado:
 - a) El primero es mucho más útil que el segundo.
 - b) El primero es un poco más útil que el segundo.
 - c) Son iguales.
 - d) El segundo es un poco más útil que el primero.
 - e) El segundo es mucho más útil que el primero.
4. Opina sobre la facilidad de uso de los dos zoom que has probado:
 - a) El primero es mucho más fácil de usar que el segundo.
 - b) El primero es un poco más fácil de usar que el segundo.
 - c) Son iguales.
 - d) El segundo es un poco más fácil de usar que el primero.
 - e) El segundo es mucho más fácil de usar que el primero.
5. ¿Qué características te parecen más positivas de los zooms con los que has trabajado?
6. ¿Qué características te parecen más negativas de los zooms con los que has trabajado?